MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D2.6 Report on enabling technologies for Transport Cloud

23/01/2023

Author(s): Chiara RENSO, Emre BAYAMLIOĞLU, Alberto BLANCO JUSTICIA, Emanuele CARLINI, Thierry CHEVALLIER, Patrizio DAZZI, Didier DE RYCK, Francesco LETTICH, Alexandros PAPACHARALAMPOUS, Raffaele PEREGO

MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

Summary sheet

Deliverable Number	D2.6
Deliverable Name	Report on enabling technologies for Transport Cloud
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Chiara RENSO (CNR)
Contributing Partner(s)	AETHON, AKKA, CNR, HERE, HOVE, KUL, URV
Peer Review	F6S, ICOOR
Contractual Delivery Date	31-07-2021
Actual Delivery Date	30-07-2021
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	78
WP/Task related to the deliverable	WP2/T2.4
WP/Task responsible	AKKA/CNR
Document ID	MobiDataLab-D2.6-ReportEnablingTechnoTranspCloud-v1.0
Abstract	This deliverable is a report on our analysis and evaluation of the available solutions (e.g., frameworks, tools, environments, infrastructures, etc.) that could be leveraged in MobiDataLab to enable transport data sharing in the Cloud. A SWOT analysis is conducted on the most promising. Enabling technologies for transport cloud data processors are also included.

Legal Disclaimer

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

Project partners

Organisation	Country	Abbreviation
AETHON SYMVOULI MICHANIKI MONOPROSOPI IKE	Greece	AETHON
AKKA I&S	France	AKKA
CONSIGLIO NAZIONALE DELLE RICERCHE	Italy	CNR
F6S NETWORK IRELAND LIMITED	Ireland	F6S
HERE GLOBAL B.V.	Netherlands	HERE
CONSORZIO INTERUNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
HOVE	France	HOVE
KATHOLIEKE UNIVERSITEIT LEUVEN	Belgium	KUL
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES	Belgium	POLIS
UNIVERSITAT ROVIRA I VIRGILI	Spain	URV

Funded by the European Union

Document history

Version	Date	Organisation	Main area of changes	Comments
0.1	30/04/2021	CNR	First draft with Table of contents	
0.2	30/05/2021	CNR	Second draft – integrated with first contributions from partners	
0.3	16/06/2021	CNR, AETHON, AKKA, HERE, HOVE, KUL, URV	First complete version integrated with all contributions	
0.4	21/06/2021	CNR	Refined version ready for internal review	Version sent to internal reviewers and Advisory Board
0.5	20/07/2021	CNR, F6S, ICOOR	Revised version with AKKA contribution	Version modified to take into account reviewers' comments (with track changes)
0.6	23/07/2021	CNR	Adding of Mobility data solution and general consistency check	Mobility Data solution added by an advisory board member. No track changes
0.7	23/07/2021	CNR	Added HOVE contribution	
1.0	30/07/2021	AKKA	Quality check	Final version

Executive Summary

This deliverable is a report discussing the results of Task 2.4, which is aimed at performing a 360-degree analysis and evaluation of the available solutions (e.g., frameworks, tools, environments, infrastructures, etc.) that could be leveraged in MobiDataLab to enable transport data sharing in the Cloud. Such analysis takes into account the key features of existing approaches and provides as output an analysis of state-of-the-art solutions from different perspectives, encompassing both technical and non-technical viewpoints. A SWOT analysis is conducted on the most promising solutions to highlight their strengths and weaknesses, as well as the opportunities and threats that could derive from their exploitation in MobiDataLab.

Given the heterogeneity and peculiarities of mobility data, and the various constraints related to their safe and trusted sharing, MobiDataLab envisions the usage of an open federated cloud architecture where complex and often contrasting requirements can be enforced easily and practically. In

principle, a federated cloud could support the sharing of arbitrary resources from arbitrary application domains with arbitrary consumer groups across multiple administrative domains.

This deliverable aims to identify the open frameworks and solutions that can provide effective and efficient ideas, solutions, approaches and services to MobiDataLab stakeholders and facilitate their collaboration, including secure methods to selectively share and access the data subject to any specific combination of privacy/visibility rules and conditions. In preparing this report we established a fruitful collaboration and mutual fertilisation with other tasks belonging to WP2 (Open Knowledge Base) and WP4 (Transport Cloud Prototype). In particular, T2.1 (Legal and Regulatory Requirements) and T2.2 (Data Privacy Requirements) helped us to define a common ground on the available solutions able to match the national and European requirements in terms of the legal and privacy framework. A similar interplay was achieved with T2.3 (Standard Requirements), which aims at defining the requirements in terms of standards that could be supported by the Cloud environment to the data federation.

The deliverable is structured in three main parts, i.e., (1) the solutions for data sharing, including the relevant dimensions, European projects and initiatives, mobility data platforms and technological federated cloud solutions, (2) the SWOT analysis, and finally (3) the solutions for the Transport Cloud data processors.

The deliverable will serve as input for Task 4.1 Architecture/Design of the cloud solution starting at M6. The addition of a section on Transport Cloud data processors makes it an input also for other tasks of WP4 starting at M6, specifically: T4.2 Reference Data Catalogue; T4.3 Data Access Services and Data Channels; T4.4 Data processors. This deliverable is also related to WP3 (New data sharing services and business models), and specifically to T3.2 (Data sharing market technological developments monitoring), where European projects and initiatives are considered from the business and economic point of view.

Table of contents

List of figures

Funded by the European Union

List of tables

Abbreviations and acronyms

Abbreviation	Meaning
AIS	Automatic Identification System
API	Application Programming Interface
B2B	Business-to-Business
DBMS	Database Management System
EDI	Electronic Data Interchange
FAIR	Findable, Accessible, Interoperable, Reusable
FACT	Fair, Accurate, Confidential and Transparent

GBFS	General Bikeshare Feed Specification
GDPR	General Data Protection Regulation
GIS	Geographical Information System
GTFS	General Transit Feed Specification
IaaS	Infrastructure as a service
JSON	JavaScript Object Notation
MaaS	Mobility as a Service
OGC	Open Geospatial Consortium
PaaS	Platform as a service
PoC	Proof of Concept
RDF	Resource Description Framework
REST	Representational state transfer
RDBMS	Relational Database Management System
SaaS	Software as a service
URI	Uniform Resource Identifier
VRE	Virtual Research Environments
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Map Service Interface Standard
WMTS	Web Map Tile Service
WP	Work Package
XACML	eXtensible Access Control Markup Language

Funded by the European Union

1. Introduction

1.1. Project overview

There has been an explosion of mobility services and data sharing in recent years. Building on this, the EU-funded MobiDataLab project aims at fostering the sharing of data amongst transport authorities, operators and other mobility stakeholders operating in Europe. MobiDataLab aims at developing knowledge and a cloud solution with the objective of easing the sharing of data. Specifically, the project is based on a continuous co-development of knowledge and technical solutions, where the advice and recommendations of experts and supporting cities, regions, clusters and associations is constantly collected and analysed. Such actions are assisted by the incremental construction of a cross-thematic knowledge base and a cloud-based service platform, which will improve access and usage of data sharing resources.

1.2. About the enabling technologies for the Transport Data Cloud

This deliverable is a report discussing the results of Task 2.4, which aims to perform a 360-degree analysis and evaluation of available solutions (e.g., frameworks, tools, environments, infrastructures, etc.) that may be leveraged in MobiDataLab to enable transport data sharing in the Cloud. Such analysis takes into account the key features of existing approaches and provides as output an analysis of state-of-the-art solutions encompassing both technical and non-technical dimensions. A SWOT (Strength, Weakness, Opportunity, Threat) analysis is then conducted on the most promising solutions to highlight their strengths and weaknesses, as well as the opportunities and threats that could derive from their exploitation in the context of this project.

Given the heterogeneity and peculiarities of mobility data, and the various constraints related to their safe and trusted sharing, MobiDataLab envisions the usage of an open federated cloud architecture where complex and often contrasting requirements coming from FAIR (Findability, Accessibility, Interoperability, and Reusability) and privacy principles can be enforced easily and practically. In principle, a federated cloud can support the sharing of arbitrary resources from arbitrary application domains, with arbitrary consumer groups across multiple administrative domains (Robert B. Bohn, Craig A. Lee, Martial Michel (2020). A federated architecture would be beneficial for MobiDataLab given its main objective of fostering the sharing of data amongst transport authorities, operators and other mobility stakeholders operating in Europe which in most of the cases want to maintain the governance of their data. The present deliverable aims at identifying the open frameworks and solutions that can provide effective and efficient data-sharing services to MobiDataLab stakeholders, and that facilitate their collaboration by supporting secure methods to selectively share and access the data subject to any specific combination of privacy/visibility/trust rules and conditions.

We observe that the solutions identified and reported in this deliverable are quite heterogeneous, as they range from the experiences gathered by various European projects and initiatives on data sharing, to mobility data platforms and technological solutions that are extensively used in the real world and that support the notion of federated cloud. Accordingly, the deliverable proceeds to

Funded by the European Union

identify the most promising solutions from which MobiDataLab can take inspiration to design the future architecture of the Transport Cloud (which is the goal of D4.1), by means of a SWOT analysis. Such analysis is used to develop a discussion around the relevance of the identified solutions in relation to MobiDataLab's goals, and is performed at a qualitative level due to the aforementioned heterogeneity and the fact that not all dimensions strictly apply to all solutions (e.g., some dimensions are more technical, while others are not).

Finally, at the end of the report we included a discussion on the state-of-the-art related to data processors (Section 4), i.e., the technical solutions that will allow MobiDataLab prototypal services to interact with the transport cloud data sharing (and that will be the objective of Task 4.4). Therefore, these solutions are not included in the SWOT analysis.

1.3. Structure of the document and how it relates with other work packages and deliverables

We identify different levels in the state-of-the-art literature related to data sharing solutions. We list the main approaches with respect to three different categories:

1) European projects and initiatives related to Data Sharing are reported and discussed in Section 2.2;

2) mobility data sharing frameworks are analysed in Section 2.3;

3) we report on cloud-based federated data sharing enabling solutions in Section 2.4.

These three categories are compared via a SWOT analysis in Section 3. Finally, the solutions enabling MobiDataLab processor services are discussed in Section 4.

In preparing this report we established a collaboration and mutual fertilisation with other tasks belonging to WP2 (Open Knowledge Base) and WP4 (Transport Cloud Prototype). In particular, T2.1 (Legal and Regulatory Requirements) and T2.2 (Data Privacy Requirements) helped us to define a common ground on the available solutions able to match the national and European requirements in terms of the legal and privacy framework. A similar interplay was achieved with T2.3 (Standard Requirements), which aims at defining the requirements in terms of standards that could be supported by the Cloud environment to the data federation. The solutions for the access and the geographic/semantic enrichment of the data stored on the MobiDataLab transport cloud will be instead developed in T4.4 (Data Processors) and here in this deliverable we perform an analysis of the state of the art on the data processors as an input to T4.4. This deliverable is also related to WP3 (New data sharing services and business models), and specifically to T3.2 (Data sharing market technological developments monitoring), where European projects and initiatives are considered from the business and economic point of view.

This deliverable serves as input for Task 4.1 Architecture/Design of the cloud solution starting at M6. The addition of a last Section 4 makes it an input also for other tasks of WP4 starting at M6, specifically: T4.2 Reference Data Catalogue; T4.3 Data Access Services and Data Channels; T4.4 Data processors.

This report is due at M6 (31 of July 2021) but it is very likely that some relevant initiative will start or become mature after the delivery of the report, or that it will be realised at a later point that something relevant was missing or just will become available later on, as we also highlight later on. Monitoring the state-of-the-art should be considered a continuous activity to contribute to the Open Knowledge Base, therefore all the partners involved agreed on maintaining knowledge base live by updating its content even after its delivery to the commission. Indeed, the content of the report, and the insights gained from it, will be included in the MobiDataLab knowledge base as soon as it will be created and deployed.

It is also to be noted that the MobiDataLab advisory board members contributed to the deliverable in suggesting improvement and additional contributions and specifically **Stephane Dreher** (Ertico), **Leo Frachet** and **Tu-Tho Thai** (MobilityData.org).

It is likely that each of the solutions considered in this deliverable covers only specific aspects and portions of the MobiDataLab desiderata. Indeed, no single framework is universally adopted by data providers and consumers. This report thus investigates the various dimensions and features relevant for MobiDataLab's goals, with the purpose of identifying a common denominator that will allow us to identify the minimal key ingredients for a federated cloud infrastructure, where such infrastructure should have the capacity of referring and collaborating with tier resources using open standards, or standard-compliant API (Application Programming Interface) façade that interfaces to ad-hoc connectors when necessary.

2. Solutions enabling mobility data sharing

2.1. Relevant dimensions for a suitable MobiDataLab Data Sharing solution

In this deliverable, we report on the spectrum of state-of-the-art for cloud-based data sharing solutions from different perspectives that are relevant for the MobiDataLab's solution encompassing both technical and non-technical viewpoints (dimensions). Inspired by (Robert B. Bohn, Craig A. Lee, Martial Michel (2020)), we collaboratively identified a number of criteria and aspects to characterise the existing solutions based on the MobiDataLab's requirements for the Transport Cloud.

Non-technical perspectives include the administrative, governance and regulatory requirements, and possible mechanisms to enforce the laws, in interaction with Task 2.1. The technical dimensions include Cloud Federation aspects and Data aspects. Note, however, that in this report we do not consider economic/market and business models of mobility data-sharing, since these dimensions are specifically addressed in WP3.

We highlight here that each of the solutions considered may cover only specific aspects and subsets of the MobiDataLab's desiderata. We thus want to analyse the aspects and features relevant for MobiDataLab's goals with the purpose of identifying a common denominator that will allow us to devise the key ingredients for a minimal and extensible federated cloud infrastructure with capacity of referring and collaborating with tier resources using open standards or standard-compliant API façade that interfaces to ad-hoc connectors when necessary.

2.1.1. Governance and regulatory aspects

Non-technical perspectives include the administrative, governance and regulatory requirements, and possible mechanisms to enforce the laws. These aspects are in interaction with Task 2.1 (Legal and Regulatory Data sharing gap analysis) and related Deliverable D2.1 due at M12. Furthermore, it is related to T2.5 (Data Governance Requirements) and related deliverable D2.7 due at M12.

In the current EU policy context, data governance is mainly referred to as an organisational and legal framework that could serve the objective to facilitate access to, and reuse of, data. All users and service providers shall observe all the relevant laws and regulations defined at their over national (EU), national, regional, and local levels. In the case of a federation, the members have to observe the union of the regulations defined therein. The governance body of the federation has to establish, define, and enforce the strategies for identification and access to data, as well as ensuring that services are compliant with the existing regulatory environment. This must be done through the identity and authorisation credentials that are associated with users, and the access policies that are defined for any given resource.

From a legal standpoint, data transactions (as uni-, bi- and multi-lateral exchanges) through which data sharing and reuse are realised make up the kernel of this framework. Hence, we need a more

forward-looking approach to data governance that can be conceptualised, according to what it is going to be developed in T2.1 and T2.5 (D2.1 and D2.7) around the following four pillars:

- **First pillar: substantial rights and major legal frameworks relevant to data and data transactions.** New upcoming regulations are to be expected. There are, however, already many legal frameworks dealing, more or less directly, with data and pursuing various legal and policy objectives – just to name a few, GDPR, intellectual property and trade secrets, data sharing legal obligations, and so on. Thus, how new regulations will interact with existing legal frameworks relevant for data and data transactions in a consistent way is in itself a challenging effort. This pillar is relevant for MobiDataLab since the Transport cloud should be aware of the new regulations and consider how to include them in the current regulatory frameworks.
- **Second pillar: interplay between legal and technical infrastructures.** One key challenge in MobiDataLab is how legal contracts concerning mobility data can be technically enforced and implemented within the context of the transport cloud – for instance, contracts for data transactions may be drafted via rights expression languages, such as smart contracts and algorithmic contracts. Legal frameworks should also favour the natural emergence of ecosystems by promoting interactions and collaborations. This pillar is clearly relevant in considering the possible technical solutions and data access methods.
- **Third pillar: contract - mobility data (re)use and sharing.** (Re)use of data viewed as a public good should be promoted, while preventing unfair practices and enforcing contractual terms. The MobiDataLab transport cloud should be therefore placed on general principles such as proportionality, fairness, and equal treatment. Said principles should then be enforced via contractual terms and technical means.
- **Fourth pillar: regulations relating to organisational aspects.** Current policy developments within the EU show the trend to regulate organisational aspects separately from substantive rights. This policy agenda is closely related to that of “digital sovereignty”. EU currently lacks data infrastructures, which in turn requires regulation efforts combined with other types of actions. To this end we highlight those important efforts in terms of proposed best practices, although limited to (sectorial) cases of specific (more collaborative) data governance mechanisms, e.g., data pools, data spaces, data commons, data trusts, data federations, data altruism, data cooperatives, data marketplace, and so on.

2.1.2. Cloud federation aspects

In this section we start the discussion of more technical perspectives. When we talk about a federation, we refer to an administrative, technological and security domain where users and resources are consistently managed. Such domain is however virtual as it logically comprises multiple parts of different sites or organisations. That virtual domain is not necessarily owned by one organisation, but is an entity where the participants can agree on the purposes, goals, and governance of any federation instance (see Robert B. Bohn, Craig A. Lee, Martial Michel, 2020).

When considering the relevant aspects for the design of a cloud federation enabling transport data sharing, we focus on the access control and authentication methods, the cloud service type, usage and policies, and the technical compliance with relevant standards.

2.1.2.1. Federated access control and authentication methods

This aspect includes the kind of federated authentication, identity and trust provided by the federate cloud solution. It also includes the granularity of access that can be at the level of datasets and/or single records. In MobiDataLab, we specifically need to provide robust and full authentication services to properly ensure identity and trust. For the granularity level, the MobiDataLab transport cloud might support accesses at both dataset and record level, although the record level seems to be at the moment not a strong requirement.

2.1.2.2. Cloud Services type: IaaS, PaaS, SaaS

Cloud federation can involve resources having different levels of abstraction and that can be exploited in different ways. Infrastructure as a service (IaaS), Platform as a service (PaaS) and Software as a service (SaaS) represent the three main categories under which fall most of the computer platforms that are provisioned using a cloud.

- Infrastructure as a Service, or **IaaS**, also known as cloud infrastructure services, is a form of cloud computing in which an IT infrastructure is provided to end users through the internet. IaaS provides high-level APIs used to dereference various low-level details of underlying computing infrastructure, like physical computing resources, location, data partitioning, scaling, security, backup, and so on. A hypervisor (e.g., Xen, Oracle VirtualBox, VMware ESX/ESXi, etc.) runs the virtual machines as guests. Pools of hypervisors within the cloud operational system can support large numbers of virtual machines and the ability to scale services up and down according to customers' varying requirements. With IaaS the user does not have to manage an on-premise datacentre and does not have to worry about physically updating or maintaining these resources themselves.
- Platform as a Service or **PaaS** allows providers to deliver a computing platform and solution stack as a service to customers who have the need to facilitate deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. Customers handle the actual application and data. Primarily for developers and programmers, PaaS gives the user a platform on which to develop, run, and manage their own applications without having to build and maintain the infrastructure usually associated with the process. Cloud PaaS, includes services provided by Alibaba Cloud, Microsoft Azure, Google Cloud, Amazon Web Services (AWS), and IBM Cloud;
- Software as a Service or **SaaS** allows service providers to license software applications to their customers for use as a service on demand. Providers can host these applications on their own servers or upload the applications to consumer devices. SaaS are usually delivered through web applications, which the service provider manages. Customers use the service typically through a web browser or API.

For MobiDataLab the SaaS type is the most relevant, although a PaaS might be considered for the data processors, while IaaS is not relevant. Indeed, the Transport Cloud will provide services in the form of software services (e.g. data access, API, data processors, etc) while it is currently not planned to provide a Platform as a Service neither an Infrastructure as a Service.

2.1.2.3. Technical compliance

Technical compliance represents the act of obeying some particular technical rule, norm, data standard and of acting according to the requirements for technical products. Focusing specifically on federated clouds, this indicates the adherence to standards, best practices, requirements, and specifications by all the entities involved. This is a fundamental quality for MobiDataLab Transport Cloud in order to have fully functioning federations and ease the integration of the Transport Cloud and the hosted shared data in novel or already existing services.

A few practical examples of requirements for technical compliance with standards are W3C/OGC best practices[1] and data providers like journey planners (Navitia, OpenTripPlanner), National Access Points (e.g., Data4PT[2]) or other like Safety Related Traffic Information (SRTI) for road traffic[3] or new mobilities (Vianova platform) or compliance with ITxPT[4].

This aspect is relevant for MobiDataLab since the transport cloud should possibly provide compliance with standards, and we must ensure that the identified technical solutions will support this compliance.

2.1.3. Data Aspects

2.1.3.1. Privacy and Security

These criteria concern the fact that mobility data provided by the MobiDataLab transport cloud should respect strict privacy and security requirements. Mobility data collected from individuals, communities or human activities display in fact high levels of unicity and regularity, that is: i) mobility traces from an entity are quite different from mobility traces from other entities; and ii) entities follow tight mobility patterns (see Deliverable D2.3 entitle “State of the art on Mobility and Transport data protection technologies”). These characteristics render mobility data highly vulnerable to record linkage and attribute linkage attacks, which in turn may lead to identity or attribute disclosure, even when only data aggregates are released.

On the one hand, these data ought to be protected by secure, standardised encryption schemes so that the confidentiality of the data is ensured against external attackers and be subjected to tight authentication and access control mechanisms, with as much granularity over the resources, roles, and permissions as needed.

When collecting mobility data from individuals, the service provider may opt to put mechanisms in place so that the mobility traces from the individuals are distorted or aggregated. These protection

[1] https://www.w3.org/TR/dwbp/ and https://www.w3.org/TR/sdw-bp/
[2] https://data4pt-project.eu/
[3] https://www.dataforroadsafety.eu/
[4] https://itxpt.org/

Funded by the European Union

mechanisms can be collectively named as privacy-preserving location-based services, which typically generalise or obfuscate data at the client level. Note that if some per-user analysis is required, the mechanisms put in place should maintain some level of consistency in the user identifiers so that complete mobility traces can be collected. Additionally, some of these mechanisms require either the assistance of a trusted third party or a peer-to-peer network between the users and the collector. Mechanisms for privacy-preserving location-based services are described in D2.3, Section 5.

As for mobility data sharing approaches we distinguish between the publishing trajectory microdata and aggregate data (for example heat maps). This publishing may lead to privacy breaches as in traditional statistical databases release, and several methods have been proposed, with the most relevant ones described in deliverable D2.3, Sections 6 and 7. The application of these methods requires generic computations to be executed in the cloud.

The MobiDataLab transport cloud must consider these aspects and therefore the technical solution for data sharing and processing should include support to enforce advanced privacy and security solutions.

2.1.3.2. FAIR principles

The FAIR principles[5] emphasise machine-actionability, i.e., the capacity of computational systems to find, access, interoperate, and reuse data, with none or minimal human intervention, because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data (see also real-time data below) ((Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship).

The first step in (re-)using data is to find them. Metadata and data should be easy to find (e.g., via standard formats and languages that support efficient querying, such as RDF with SPARQL) for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services. Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorisation. Also, the data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. The ultimate goal is then to optimise the reuse of data, as discussed later in the data access policies. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. Finally, we observe that in the context of a transport cloud the availability of standard and well-defined APIs would help to implement part or all of the above principles.

[5] https://www.go-fair.org/fair-principles/

Funded by the European Union

The machine actionability of data in the MobiDataLab transport cloud is one of the main objectives of the project given the overall goal of incrementally building a cross-thematic knowledge base and a cloud-based service platform, which will improve access and re-use of shared mobility data resources.

2.1.3.3. Data access policies

Broadly speaking, the notion of *data access* encompasses the set of policies, standards, and technological solutions that *regulate* the *access, flow, use*, and *re-use* of data. While the access aspect represents the traditional access control facet of the dimension presented herein, the other three introduce novel challenges that are definitely relevant for the MobiDataLab transport cloud. The policies for granting flexible and novel models for interorganisational data sharing among the members of the federation are in fact a key requisite for new value creation in the digital economy envisioned in MobiDataLab.

For instance, let us consider the case in which some provider (e.g., an individual, a company, an institution, etc.) grants access to their data to a consumer. The provider may then also want to **control** if or how such data can be *altered, copied*, or further *disseminated*.

The notion of *usage control*[6] deals with the specification and enforcement of restrictions regulating what must (or must not) happen to data. Thus, usage control deals with requirements that pertain to data processing (obligations), rather than data access (provisions). Usage control is relevant in several contexts, e.g., intellectual property protection, compliance with regulations, and digital rights management.

This aspect includes all the policies that the cloud federation provides or enforces for data access and data processing. A policy could include, for example, a threshold in the number of accesses or downloads beyond which a user cannot access directly the data[7] or a threshold in the freemium models, e.g. access to real-time open data as for example the RATP/CityMapper[8]

2.1.3.4. API access

API access considers the fact that the solution will support the access to data by means of APIs (Application Programming Interfaces). This feature is extremely relevant for MobiDataLab (see the above mention to FAIR principles) to be able to support novel services exploiting the shared mobility data sources and also the MobiDataLab processors enriching or transforming the data (see Section

[6] Data Usage and Access Control in Industrial Data Spaces: Implementation Using FIWARE. https://www.mdpi.com/2071-1050/12/9/3885/pdf
[7] https://operations.osmfoundation.org/policies/api/ or https://operations.osmfoundation.org/policies/tiles/
[8] https://citymapper.com/

4.2). The definition of the APIs relates also to the granularity of access, since the API can provide access at the dataset level or at the record level depending on the solution considered.

2.1.3.5. Static vs. real-time data

Static data represents information that is infrequently modified and can be considered likely immutable. **Real-time** (or streaming) data represents information that is continuously generated by some source (e.g. cars, cameras deployed within a city's road network, mobile devices, sensors, etc.), possibly at high speed. Real-time data typically yield huge volumes of information. These two different types of data require different technological solutions to be effectively stored, indexed, analysed, and processed.

In MobiDataLab we mainly focus on static data as datasets made available by date providers which are presenting static information and not real time (e.g. see Deliverable on use cases D2.9). We envision data sources that allow to add fresh, updated information related to some phenomenon, but they commonly append record to existing datasets. Very rarely the raw data records have to be updated. Nevertheless, a support for real time data can be useful for a possible follow-up.

2.1.3.6. Data Trust

Worth of mention is the relatively recent notion of **data trust**[9] . A data trust is "created when someone, or a lot of someones, hand over their data assets or data rights to a trustee. That trustee can be a person or an organisation who will then hold and govern that data on behalf of a group of beneficiaries and will do so for a specific purpose. The beneficiaries could be those who handed the data to the trust, or anyone else (including society at large). Importantly, the trustee has a fiduciary responsibility to look out for the interests of the beneficiary […]. That also means that the trustee is not allowed to have a profit motive or, more generally, a conflicting interest in the data or data rights under its custody.

One important feature of a data trust is that the trustee can decide who has access to the data under the trust's control and who can use it. And, importantly, if that data user fails to comply with the terms and conditions, the trustee can revoke access."

A proper handling of data trust is essential for the MobiDataLab transport cloud; thus, such aspect is extremely relevant. However, the concept of data trust is transversal to many more detailed aspects already discussed in this document. For example, an effective handling of data trust relies on properly enforcing policies for privacy, security, data access, governance and regulatory requirements. We finally report that there exists also trustless solutions based on Distributed Ledger Technologies (DLT) that can be considered.

[9] Data Trusts: Why, What and How. https://algorithmwatch.org/en/data-trusts-why-what-and-how/

2.2. European projects and initiatives on federated data sharing

Data Sharing is of paramount importance for Europe. The European commission recognises how exploiting data allows the private sector to continuously innovate and create new types of thriving businesses, while public entities and institutions can leverage public data to better understand societal dynamics and take informed decisions.

The document A European Strategy for Data[10] provides the European commission's vision on how the EU should take advantage of, and deal with, the vast amounts of data nowadays available, as well as the associated emerging digital technologies, over the next 5-10 years. The vision highlights, and revolves around, the sheer importance of data in nowadays world.

The European commission observes that the above positives come together with grave privacy and security issues that need to be dealt with appropriately. Said issues are clearly recognisable in the approaches to data followed by other major countries. For instance, the US tend to favour corporate interests over public and individual ones, while China's approach is heavily skewed towards government surveillance and strong control of big tech companies.

The European Strategy for Data attempts to establish a well calibrated balance between private, public, and individual interests according to the European values, while allowing the EU digital economy to be on par and extremely competitive with other (currently dominant) global players.

To achieve the above goals the vision provides a high-level outline of the legal framework and technological solutions that should be implemented and used across the European bloc. Among these, we highlight the importance given to federated cloud infrastructures that should bring the common European data space to life, favouring the enforcing of policies regulating the use and re-use of data, data governance, data interoperability, and data quality.

Finally, the vision expressly considers the need of a common European mobility data space, illustrating the trends, needs and issues of various transportation systems within the context of smart cities, and the commission's future actions to attain such goal.

The MyData initiative[11] consists of a joint international effort, backed by the Finnish Ministry of Transport and Communications, that aims to empower individuals by improving their right to self-determination regarding their personal data, a.k.a. “data sovereignty”[12]. The activities behind the MyData initiative are carried on by MyData Global, an international noprofit non-governmental organisation based in Helsinki that encompasses over 100 organisation members and close to 400 individual members from over 40 countries, on six continents.

[10] https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
[11] https://mydata.org
[12] https://mydata.org/wp-content/uploads/sites/5/2020/08/mydata-white-paper-english-2020.pdf

Similar to the “A European Strategy for Data” vision illustrated previously, the MyData vision recognises the enormous societal and economical potential behind individual data. This vision, however, highlights how individuals, while being theoretically protected by existing legislation, e.g., General Data Protection Regulation (GDPR), have in practice very little control on the use and flow of their information due to the lack of effective tools and awareness. This is especially evident in the US, where a few big tech companies have an almost unchallenged control over personal data and enjoy lax regulations. Such companies have established a de facto monopoly over personal data, thus stunting innovation, competition, and potential economic growth. Going back to the MyData initiative, its overarching goal is to establish a fine balance between protection and utilisation of individual data. To reach such goal, MyData strives to trigger three main shifts over time.

The first one is the shift from formal (e.g. GDPR legislation) to actionable rights, where individuals should be able to regulate access to their data, correct their data, enjoy portability, and exercise their right to be forgotten, whenever they want and in the simplest possible way (i.e., “one-click rights”). The second shift aims at making people aware of their rights and instructing them on the opportunities and challenges behind a data-agile economy driven by European values. The third shift aims at moving from closed to open ecosystems, characterised by open and well-established standards, actors having well defined roles, and backed by appropriate legislation. The goal is to impact the current state of affairs in order to break current monopolies and instigate innovation and competition, while empowering individual rights to data. We note that the development and adoption of a European cloud federation where the flow of personal data is controlled by individuals is expressly mentioned as a mean to implement such third shift.

Given this context that shows a strong interest of Europe to create opportunities to data sharing, in the following of this section we go more concrete and we report on specific European projects for federated data sharing (including or not mobility data) that are relevant for MobiDataLab with respect to the dimensions presented above and can offer solutions that can be exploited in the project. For each item we report a short summary of the initiative and we briefly assess the relevance for MobiDataLab. Furthermore, we introduce a synthetic table to summarise the dimensions where we assess if the dimension is somehow covered by the reported initiative/solution. Therefore, **YES** means that, from the available documentation, we could assess that the dimension is considered and covered in the solution, **NO** means that, form the available documents we could assess the dimension is not covered, **NOT ASSESSABLE** means that from the documents we could not assess if the dimension is considered or not in the solution, and **NOT APPLICABLE** means that the specific dimension is not applicable to the considered solution.

2.2.1. GAIA-X

GAIA-X is an industrial-driven EU project, aligned with the "European data strategy" vision[13] , that is managed by French and German institutions and entities. The core tenet behind GAIA-X is that of "digital sovereignty", which is "[...] the power to make decisions about how digital processes, infrastructures, and the movement of data are structured, built, and managed [...] according to EU

[13] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

standards". This in turn requires "[...] to design and implement a data sharing architecture -- including standards for data sharing, best practices, and tools -- and governance mechanisms, as well as an EU federation of cloud infrastructure, related infrastructure, and data services".

From the above we notice a remarkable intersection between GAIA-X's goals and those of MobiDataLab, although the latter is focused on a more limited scope related to the value of sharing mobility data. Indeed, the **GAIA-X technical architecture document**[14] shows how GAIA-X considers many of the dimensions relevant for MobiDataLab. Specifically, in reference to the cited document we see that:

- Gaia-X is a federation of virtual administrative domains. The GAIA-X technical architecture document illustrates the "core architecture elements", i.e., the entities and the types of information involved. Furthermore, it focuses on the enablement and governance of federated ecosystems (which are referred as "GAIA-X Federated Ecosystems" in the technical document).
- Various dimensions related to the Data aspects are prominent in GAIA-X. The "Data access" and "Data sharing policies" are covered in Section 2 of GAIA-X technical architecture document, where the notion of data asset (i.e., dataset) is provided, along with the entities in charge of producing (providers) and consuming (consumers) data assets. Finally, the same section provides the technical notions and means that should be used to enforce policies and sharing mechanisms dictating how the access, control, and flow of data should be regulated within GAIA-X's ecosystems. Section 5.2 of the technical architecture document deals with the problem of future use of data, once access to it has been granted to consumers -- for instance, how to ensure that intellectual property is protected. Finally, we observe that this issue is also relevant to some of the governance aspects introduced in the present document, more specifically to the second and third pillars.
- The "Federated access control and authentication methods" dimension is covered in Section 3 of the GAIA-X technical architecture document, where the authors detail how Federated Identity Management, Decentralized Identifiers, cryptographical Verification of Self-Descriptions, and Accreditation and Certification Processes shall be realised within the context of the project. Finally, the Sections 3.3 and 3.4 provide a high-level description on how trust and identity shall be realised in GAIA-X. Briefly speaking, identity -- which is the key to gain access to the ecosystem – is implemented by means of a unique identifier plus a list of attributes which describe the entity behind said identifier, while trust is established by cryptographically verifying identities using the federated Identity Management component of GAIA-X.
- The "Cloud Services type: IaaS, PaaS, SaaS" dimension is covered in Section 2 of the GAIA-X technical architecture document, where the authors introduce the notion of node, which is defined as some set of computational resources (possibly arranged hierarchically, where a node can contain other nodes) and service, which the authors define as some functionality, or some data asset, that is exposed via a node. The authors then observe how GAIA-X's notions of "node" and "service" can be interpreted to realise IaaS, PaaS, and SaaS cloud services.

14 https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.html

- The "Privacy and Security" dimensions are touched in Section 5 of the GAIA-X technical architecture document, where the authors provide a high-level framework on how these aspects should be realised within the project – to this end we highlight the strong emphasis the authors place on the onboarding procedure (introduced in Section 6), i.e., the act of some provider to register at GAIA-X to integrate their nodes and services – and how well-known and well-established technical standards and legal frameworks (e.g., GDPR) should be employed to implement these aspects effectively. We finally observe how these aspects also touch some of the governance and regulatory aspects introduced in the present document, specifically the second and fourth pillars.
- The “FAIR principles”, although not explicitly mentioned in the GAIA-X technical architecture document, are de facto addressed throughout the whole document. Specifically, we highlight that the FAIR "findability" principle is covered by GAIA-X via the notion of "self-description" (Section 2.4 of the technical document), which represents the set of properties that some asset or participant within the GAIA-X ecosystem possesses. The authors indicate that self-descriptions should be represented by means of some well-established standard (e.g., RDF/OWL, JSON-LD) which allows query support. The notion of "Catalogue" is introduced in Section 2.5 of the cited document to represent the main building block for the publication and discovery of self-descriptions, and "Catalogue Federation" (in line with the federated nature of GAIA-X's ecosystems).
- The "Technical Compliance" dimension is covered in Section 5.3 of the GAIA-X technical architecture document. Here the authors clarify how the onboarding process can be used to ensure that participants (i.e., providers and consumers), services, and nodes comply with inter¬nal requirements, as well as external regulations and policies.

We report that the GAIA-X’s website introduces a use case involving mobility data[15] . Here the project’s members highlight how the work being done in the context of GAIA-X has the potential to address serious urban issues and bring real added value. More specifically, said use case considers scenarios where automotive mobility is increasingly reaching its limits in densely populated urban areas, e.g., parking spaces are becoming scarcer, traffic jams and commuting times are becoming longer, while air pollution and road accidents are on the rise of Intermodal mobility, i.e., the combination of alternative modes of transport, and vehicle pooling all offer potential solutions. However, they require the ability to access and share business and end customer data. Therefore, ensuring and safeguarding data sovereignty is a key requirement. Currently, concerns about the loss of personal data severely prevent or limit such intermodal mobility approaches. For intermodal mobility solutions to work, providers must exchange various data for different process steps both with each other and with customers. In particular, the cross-provider provision of (sensitive, as in part personal) mobility data must be guaranteed via a secure and trustworthy data exchange in which the use of data is controlled and regulated by the recipient.

In this type of scenarios GAIA-X can clearly help to safeguard data sovereignty and create the necessary framework for the implementation of alternative mobility concepts (e.g., intermodal

[15] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Artikel/UseCases/mobility-data-interoperability-and-data-sovereignty.html

mobility that relies on the combination of different means and modes of transport). The framework includes safe data transfer and data sovereignty as the foundation for new business models.

Relevance for MobiDataLab: GAIA-X looks extremely relevant as it addresses many of the dimensions considered in this report. GAIA-X scope seems to be general and very ambitious since it addresses the data sharing in different contexts, while MobiDataLab specifically focuses on a more limited scope since it is related to the value of sharing mobility data through a transport cloud. The aim of GAIA-X governance is to transfer the project into fixed structures. For this purpose, companies and organisations from Germany and France have started to set up an international non-profit organisation that will further develop the technical solution and regulatory framework. The first prototype is supposed to be delivered in 2021. We, however, observe that the roadmap is pre-Covid pandemic, thus delays might be possible. The project appears to be nonetheless very active, with many events already held or planned[16] .

Table 1 : GAIA-X coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NO
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	YES
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES

[16] https://events.talque.com/gaia-x-summit/en/6iq6yI5LPSxaIRA6cmnq?talque=lecture-list&lectureId=aK613tw514yHwDjMOg3G

Data access policies	YES
API access	YES
Static vs. real-time data	NO
Data trust	NO

More info: https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

2.2.2. FEDeRATED

FEDeRATED is an EU funded project started in 2019 which focuses on logistics. Its objective is to provide an infrastructure provision containing a set of agreements and technical applications to enable data already available in existing IT systems (platforms) of companies and public administrations to become available to authorised users through a publish and subscribe approach. The scope of FEDeRATED is a federated network of platforms, including the sharing of both data and services, that covers end-to-end transportation chains, from the consignor to the consignees, thus ensuring enhanced visibility and transparency and enabling value adding service development for third parties. The federated network of platforms concept allows various information hubs and enterprises to connect and serve as an infrastructure distributing high quality data amongst relevant stakeholders. It also enables those enterprises and public authorities to develop services and new business, and to improve their operation.

One useful metaphor that can be used to illustrate FEDeRATED's motivating goals is the one where isolated islands need to be integrated towards standardised, regulated networks designed for delivering predictable, dependable services across the infrastructure network. In transport and logistics, these islands are the information hubs, which comprise different networks that are not connected (yet).

The project plans a number of virtual labs where to experiment the federation. Some of the considered technical and non-technical dimensions include, among others: **Trust, Accessibility, Legislation and legal framework, Data quality and integrity, (Cyber) Security, Operational management, Sustainability objectives, Data sovereignty Accessing data at its source, Investments, Governance**.

A reference technical architecture is proposed[17] and it includes several functionalities that correspond to the above dimensions, i.e., identity and authentication; different techniques for exchanging data (APIs/messages with different views of data such as EDI, JSON and RDF); data

[17] http://www.federatedplatforms.eu/index.php/library/item/detailing-the-reference-architecture-discussion-paper?category_id=18

access control with XACML (eXtensible Access Control Markup Language); search systems offering search possibilities for data; logistics services registers; API Registry – i.e., a registry of all APIs developed and used by participants Log and audit trail – each of the organisations that share data with each other should have a log and audit trail for non-repudiation (e.g. in the case of liability issues); Platform – which is an IT component for data sharing between stakeholders including for example the Logistics Market Place, more precisely, a platform where goals are matched with logistics services stored in a service registry (a Logistics Market Place might also be an algorithm that is implemented by a supply chain stakeholder and/or a platform); Booking – or contracting platform –, that is, a platform supporting contracting between a customer and service provider(s), potentially resulting in framework contracts; Ordering platform – a platform supporting ordering of logistics services.

FEDeRATED includes also data and process transformation functions implemented by a triple store and, finally, the Data Store is implemented with Triple stores to store data in the RDF (or JSON-LD) format. The FEDeRATED Semantic Model defines the structure of the triple store. A Trusted Data Store is a data store that adheres to particular governance rules (like for instance formulated by the eFTI Regulation).

Relevance for MobiDataLab: The objectives of MobiDataLab and FEDeRATED are similar since FEDeRATED has the objective of creating a federated infrastructure to cover both data sharing and services to be exploited in several virtual labs by commercial actors in the logistics domain. However, this architecture is very tailored and instantiated to specific use cases and, at the moment, there is no general-purpose federated infrastructure available, or that can be directly exploited to satisfy MobiDataLab's desiderata. It is also important to highlight that FEDeRATED does not federate data clouds but logistic systems providing interoperable services. As it offers a platform to federate different actors and their services it can be considered as SaaS.

The complexity of FEDeRATED problems and solutions is considerablely higher than MobiDataLab's. Nevertheless, the experience of the FEDeRATED project in realising such federation, and especially the experimentation in the virtual labs, is extremely interesting. For example, the use of Resource Description Framework (RDF) and triple store for data storage and semantic integration is very related. Triple stores can be used both to share linked data sharing and to link data as a basis for data analytics that is quite similar to MobiDataLab's vision.

The technical compliance is discussed broadly in the architecture documentation and it includes the compliance with several logistics actors including various means of transportation, i.e., air, road, rail, and sea transportation actors that join the federation.

We intend to keep monitoring the ongoing development of FEDeRATED to better understand the technological solutions that will be proposed to realise their federation infrastructure, and possibly take advantage of its achievements.

Table 2 : FEDeRATED coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	

Substantial rights and major legal frameworks relevant to data and data transactions	NO
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	NO
Regulations relating to organisational aspects	YES
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	YES
Data trust	YES

More info: http://www.federatedplatforms.eu/index.php

2.2.3. OPENCLOUDMESH

OpenCloudMesh is a joint international initiative under the umbrella of the GÉANT Association that is built on OwnCloud's open Federated Cloud sharing application programming interface (API) taking Universal File Access beyond the borders of individual Clouds and into a globally interconnected mesh of research clouds — all this without sacrificing any of the advantages in terms of privacy, control, and security that an on-premises cloud provides.

OpenCloudMesh provides a common file access layer across an organisation or globally interconnected organisations, whether the data resides on internal servers, on object storage, in applications like SharePoint or Jive, other OwnClouds, or even external cloud systems such as Dropbox and Google (syncing them to desktops or mobile apps, thus making them available offline).

Unfortunately, the web site does not provide any additional information regarding the possible use of the platform outside the universities' cluster, and we report it here as an example of use of the technology of OwnCloud for a federation of universities clouds. OwnCloud technology is presented later in Section 2.4.1.

Relevance to MobiDataLab: OpenCloudMesh is relevant for MobiDataLab since it is an initiative which aims at federating different universities and institutions located in Europe to share their data. Although they do not focus on mobility data specifically, it is interesting to observe that the technological solution they employ to realise the data federation is OwnCloud. It is important to remark from the technical compliance point of view that each federated cloud must be a local instance of OwnCloud or a supported data sharing solution like DropBox, SharePoint or SeaFile.

Table 3 : OPENCLOUDMESH coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	NO
Data trust	YES

More Info: https://oc.owncloud.com/opencloudmesh.html

2.2.4. BASMATI

BASMATI was a joint European and Korean project funded in the context of the H2020 framework programme. The project builds over a key fact: "*mobile devices are becoming more and more powerful and sophisticated, enabling rich-multimedia service provisioning and accessing at the same time various information sources [..] clouds aim at amplifying value by supporting any class [..] of mobile cloud services. Enhanced mobile application provisioning could be achieved through the seamless usage of cloud platforms and mobile devices for overcoming mobile devices limitations and supporting the management of services and data.*" In fact, BASMATI was aimed at delivering an integrated platform that supports the dynamic needs of mobile applications and their users by following an end-to-end approach for the definition and implementation of cloud services: (1) analysis, modelling and runtime-adaptable prediction of user and application requirements, (2) situational knowledge acquisition to understand real-world dynamics, (3) business-aware federation of cloud resources to deal with emerging situations, (4) ultra-scalable brokerage and dynamic offloading to enable optimal service offerings, and (5) functional and non-functional properties consideration across the whole service lifecycle to ensure provisioning of quality of service, security and privacy guarantees.

BASMATI emphasised on enabling runtime adaptation, including user and application prediction models, federation patterns, resources and data management policies, brokerage and offloading decisions. More in detail, the BASMATI Unified Data Management Framework (BUDaMaF) can be considered as a reference for MobiDataLab.

BUDaMaF is an approach aimed at providing the ability to easily and seamlessly relocate and offload data between the different cloud providers that contribute to the federation. This ability will aid the application providers in their efforts to uphold their quality-of-service requirements, especially in tasks that are highly dependent on quick response times. Between these functionalities we can distinguish the ability to dynamically relocate, replicate or offload data between the cloud providers that take part in the BASMATI federation of clouds. BUDaMaF was designed to handle the needs of many different platforms for cloud federation management, covering the needs for data store management and high-level data management. Being a context-agnostic approach, it is not specialised to specific domains or even underlying technologies.

Compatibility with other systems is achieved by means of proper wrappers, that can be developed using any technology, by any person and then attached to the framework using the loosely coupled architecture of RestFul Web Services. We recall that In the REST architecture, data and functionality are considered resources and are accessed using Uniform Resource Identifiers (URIs), typically links on the Web[18] . BUDaMaF design has at its core the scalability. It can be deployed either in a single machine, in a cluster of machines or in a cloud.

[18] https://docs.oracle.com/javaee/6/tutorial/doc/gijqy.html

Relevance to MobiDataLab: the concept behind and the conceptual architecture of BUDaMaF could be considered for the creation of the data plan of the transport Cloud. It allows dynamic management of application data, enabling the assisted migration, movement and replication of mobility data depending on the actual needs of the project. However, the scope of BUDaMaF is very specific and targeted towards the seamlessly relocation and offloading of data among federated cloud providers in order to support dynamic needs of mobile applications. Prototype software is available, but it does not address many of the aspects highlighted in this deliverable e.g., FAIR principles, data access policies at different granularities, trust, technical compliance, privacy, enforcement of legislative and legal frameworks.

Table 4 : BASMATI coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	NO
FAIR principles	NO
Data access policies	NO
API access	NO
Static vs. real-time data	NO
Data trust	NO

More info: https://cordis.europa.eu/project/id/723131

2.2.5. SoBigData++

SoBigData++ is an ongoing H2020 funded project to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics aimed at using social mining and big data. SoBigData++ supports social mining experiments with the FAIR (Findable, Accessible, Interoperable, Reusable) and FACT (Fair, Accurate, Confidential and Transparent) principles so to be repeatable by domain experts that are not data scientists. Furthermore, it considers ethical and legal challenges for concrete tools that operationalise ethics with value-sensitive design, incorporating values and norms for privacy protection, fairness, transparency and pluralism.

This project provides an e-infrastructure integrating new and existing services and quantitatively and qualitatively enhancing them. Activities includes the best practices and policies for the harmonisation of federated resources available at local infrastructure sites and the maintenance of the e-infrastructure for providing Virtual Research Environments (VRE). VREs provide virtual access to the integrated resources, integrates existing and newly collected datasets in the infrastructure and integrate existing tools and methods for mining social data in the infrastructure.

The SoBigData e-intrastructure is based on the D4Science infrastructure and provides access to datasets and analysis methods through the catalogue, a facility to discover, index and search all the datasets accessible through the infrastructure. It is based on CKAN technology[19] . This catalogue is expected to serve both human users - willing to know the offering of the e-Infrastructure in terms of datasets - and services / methods and other services - willing to dynamically discover resources to consume / interact with to deliver their services.

Relevance for MobiDataLab: SoBigData++ is not a cloud federation but it offers a centralised infrastructure where all data and analysis algorithms are stored and possibly used. Moreover, the main difference relies on the fact that SoBigData provides a whole research environment while MobiDataLab Transport Cloud provides a data sharing federation and data processors. Nevertheless, we believe the SoBigData infrastructure could be of inspiration for the MobiDataLab transport cloud regarding the possible analysis environment that can be useful in designing the data processors.

Table 5 : SoBigData++ coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	YES
Interplay between legal and technical infrastructures	YES

[19] http://ckan.org

Funded by the European Union

Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	YES
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	NO
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	NO
Data trust	NO

More info: https://plusplus.sobigdata.eu/

2.2.6. FENIX

The goal of the FENIX project, started in 2019, is to develop the first European federated cloud architecture cantered on the data sharing needs of the European logistics community of shippers, logistics service providers, mobility infrastructure providers, cities, and authorities, thus aiming at offering seamless interoperability among existing and future platforms. The federation is established by means of the following dimensions that will be defined in more detail, and then realised, via the involvement of specific *pilot sites*[20] that will participate to the federation platform:

[20] In this context a pilot site represents some institution, or group of institutions, either public, private, or both, that operates in the logistic sector in a precise geographical area and collaborates to the FENIX project in order to satisfy some specific needs.

1. **Trust**: rules and regulations to establish trust among members that want to exchange data in the federation; the required information is the definition of roles, identity management mechanisms and user certification.

2. **Governance**: rules and regulations for data exchange within the federation; the required information is about data quality, data provenance and data ownership.

3. **Security**: definition of main standards and best practices to be used to achieve a trustworthy communication; the topics covered are authentication and authorisation, usage policies and usage enforcement.

4. **Connectivity**: the federation should offer an interface to enable data exchange with other platforms/federations and the integration of federation services with IT back office systems.

5. **Interoperability**: interoperability is a key requirement for the collaboration between stakeholders; the needed information covers the requirements to be met by the FENIX ecosystem to achieve secure and reliable corporate interoperability. This also involves the usage of existing standards, or the definition of new ones, with a particular focus on logistics.

The proposed federated infrastructure will be compliant with the logistics standards and it will offer cloud services of the type Platform as a Service. It appears that FENIX focuses on offering services for static data, but we report that dynamic data (e.g., AIS maritime data) is mentioned as well.

Relevance for MobiDataLab: FENIX represents a specific instantiation of a federated infrastructure applied to the logistic domain, where different actors can join the federations in different pilot cases. The MobiDataLab project can thus take advantage of the ongoing work and results from FENIX to better design the federated architecture, and to take inspiration on how the different dimensions are instantiated case by case by the different pilot sites. The dimensions that are relevant to FENIX are detailed in the FENIX public deliverable[21] and briefly reported here:

- **Federated access control and authentication methods**: FENIX aims at introducing a federated cloud. This dimension relates to all the FENIX's dimensions illustrated before.
- **Cloud Service Usage: Policies, paradigms and thresholds**: providers participating in the federation appear to be able to decide which services should be visible to whom. This is reported in the Section 4.4 of the FENIX public deliverable cited above.
- **Technical Compliance**: there appears to be a strong emphasis on defining or using a set of well-established standards, as discussed in Section 5 of the FENIX public deliverable.
- **Privacy and security**: as mentioned before, two key dimensions of FENIX are those of Trust and Security, which align with our dimensions.
- **API access** is briefly covered in Section 5 of the FENIX cited public deliverable and is discussed within the context of standards to be used to enable data transfer mechanisms.
- The **data access** dimension is covered via the governance, connectivity and interoperability dimensions mentioned before.

[21] https://fenix-network.eu/wp-content/uploads/2020/07/FENIX-Deliverable-D2.2.2_v2.0_FINAL.pdf

- **Data sharing policies**: FENIX appears to offer a data hub perspective, as access to data is differentiated by the type of user, country of provenance, and so on.

Table 6 : FENIX coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NO
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	YES
Data trust	YES

More info: https://fenix-network.eu/

2.2.7. SUNFISH

Sunfish was an H2020 funded project that started in January 2015 and ended in December 2017. The project aimed to conceive, design, and implement the notion of Federation-as-a-Service (FaaS),

i.e., to offer a service that federates private and public clouds thus allowing interested parties to exchange data and services in a secure and controlled manner, based on a democratic governance model. To achieve these goals the project members did not limit themselves to propose a set of APIs, but appear to have focused -- and extensively exploited -- the capabilities offered by the blockchain technology, for example for distributed governance without trust.

A FaaS solution can be flexibly adapted to the needs of participating clouds by making use of existing identity-management components, allowing public administrations and governments to collaborate and securely share their private cloud resources. The SUNFISH platform therefore focuses on enabling the sharing of data between potentially untrusted entities while protecting the sensitive data of each entity. This is achieved through several components for controlled data sharing between services provided by different private clouds, with said components being invoked when the mechanism they provide is the most efficient among those available.

The cloud federation is intended to enable the management and optimisation of computing resources in a secure and compliant manner across the cloud federation to deliver business services for end users.

The practicability of the project has been tested around some of the real needs of three public administrations that were partners of the project, i.e., the Ministry of the Economy and Finance of Italy, the Maltese Ministry of Finance, and the UK South East Regional Organised Crime Unit.

Relevance for MobiDataLab: In general, Sunfish appears to touch several of the dimensions introduced in this document, such as the “Federated access control and authentication methods” ones. There appears to be also a strong emphasis on “Data sharing policies” as well as on “Data privacy and security”, as Sunfish employs data masking, data anonymisation, and secure multi-party computation techniques to ensure high security of provisioned services and managed data.

SUNFISH Software provides micro data and macro data anonymisation services. Micro data anonymisation means that a data set is released with the k-anonymity guarantee. This process ensures the protection of sensitive information against linkage attacks using other open data sets. Macro data anonymisation means that statistical data are released with differential privacy guarantees and this is interesting for MobiDataLab.

However, the software, although available in a Github repository, does not appear to be maintained anymore.

Table 7 :SUNFISH coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB’s goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NO
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	YES

Regulations relating to organisational aspects	YES
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES (Federation as a Service)
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	NO
Data trust	NO

More info:

http://www.sunfishproject.eu/

https://github.com/sunfish-prj

2.2.8. KRAKEN

KRAKEN (BroKeRage and MArKet platform for pErsoNal data) is an ongoing EU Horizon 2020 funded project that started in December 2019 and that aims to develop a trusted and secure personal data platform with state-of-the-art privacy aware analytics methods (with guarantees on metadata privacy, including query privacy), where it is possible *to share, broker, and trade potentially sensitive personal data*. The end goal is to develop a marketplace for personal data representing a paradigm shift towards a true alternative to organisation-centric data management, which returns management control of their own data to the users (i.e., it aims to realise the paradigm of *self-sovereignty*).

The marketplace being developed in KRAKEN promises to facilitate the sharing of potentially sensitive personal data by ensuring the *preservation of privacy*, enabling *value creation* from personal data. The KRAKEN marketplace will connect data subjects and organisations holding aggregated datasets on large populations with organisations or institutions needing more data for

research and innovation, while respecting the privacy of the data subjects. The data shared may be personal in nature, or it can consist of special categories of data.

The project combines, interoperates, and extends the best results from two existing mature computing platforms developed within two H2020 actions: CREDENTIAL[22] and MyHealthMyData[23]. It also has a special focus on two pilot sectors: healthcare and education.

The project involves integrating various technological components brought to the project by the consortium partners, namely, (1) a marketplace for potentially sensitive personal data developed by TX – Tomorrow Explored using the Streamr technology stack, (2) a Self-Sovereign Identity wallet to support registration and authentication of users, being developed by Atos and InfoCert, (3) a blockchain-backed GDPR-compliant tool for enhanced access control based on Hyperledger Fabric, provided by Lynkeus, and (4) cryptographic tools for users to perform privacy-preserving analytics, such as Secure Multi Party Computation. The initial integrations between these technologies have already begun as a result of work in 2020.

Relevance for MobiDataLab: the deliverables currently available in the project's website[24] cover ethical and legal aspects (i.e., D2.1, D7.1, and D.7.2), while those that should cover technical aspects appear not to be available at the time of writing. Thus, considering the documentation available in the project's website in terms of deliverables and overall content, KRAKEN appears to touch the four governance and regulatory pillars outlined in Section 1.4.1 -- in particular, there appears to be a strong emphasis on privacy and security aspects, and the technology that shall be used to address them, i.e., cryptographic techniques and the use of a distributed blockchain ledger technology. We report that the project does not seem to consider the notion of "federation of clouds" to reach its objectives. Finally, we report that KRAKEN's coverage of the dimensions considered in the present deliverable may change as the project evolves.

Table 8 :KRAKEN coverage of dimensions relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	YES
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	YES
Cloud federation aspects	

[22] https://credential.eu
[23] http://www.myhealthmydata.eu
[24] https://www.krakenh2020.eu/resources/work-package-deliverables

Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	NOT ASSESSABLE
Data access policies	YES
API access	YES
Static vs. real-time data	NO
Data trust	NO

More Info: https://www.krakenh2020.eu/

2.2.9. MOBiNET

MOBiNET was an EU collaborative project, framed within the context of the 7th EU Framework Programme for Research and Technological Development, that was conducted between the 1st November 2012 and the 30th June 2017.

MOBiNET's main goal was to introduce the MOBiNET platform, an e-marketplace and a technical platform enabling the interactions between suppliers, developers and users of mobility related content and service providers. Such platform offers a centralised directory for publishing and editing Business-to-Business (B2B) services, as well as functionality enabling and supporting interoperability between data sources. By showing the need for and delivering this e-marketplace, the development of innovative features and key technical components, the MOBiNET project claimed to pave the way for the creation of true European-wide Mobility e-market place on which the potential Mobility as a Service (MaaS) could run.

Indeed, the project's stakeholders highlight how mobility operators or transport providers are used to negotiate contracts on a one-to-one basis, with no common set of rules on how to engage with other mobility operators on a European level. MOBiNET consequently aimed to deliver the technical components, the business models, and organisational structure as building blocks required to *set up MaaS in Europe*.

Relevance for MobiDataLab: unfortunately, the website that hosted the project's documentation, deliverables, and platform, i.e., http://www.mobinet.eu/, seems to have been abandoned, and thus

no proper assessment was possible. There also appear to be two research papers associated with the project (https://cordis.europa.eu/project/id/318485/results); these papers, unfortunately, have restricted access.

Table 9 : MOBiNET coverage of dimensions relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	YES (MaaS)
Technical compliance	POSSIBLY YES
Data aspects	
Privacy and security	NOT ASSESSABLE
FAIR principles	NOT ASSESSABLE
Data access policies	POSSIBLY YES
API access	POSSIBLY YES
Static vs. real-time data	NOT ASSESSABLE
Data trust	NO

More Info: https://cordis.europa.eu/project/id/318485

2.2.10. National Data Warehouse for Traffic Information

The National Data Warehouse (NDW) was a project started by the Dutch Ministry of Transportation in 2007 which aimed to collect accurate, detailed, and reliable data from a large part of the national road network and to act as data-bank for urban municipalities, information service providers, research teams and, in turn, to the road user. The project thus revolved around the development of a centralised national data warehouse in which traffic data is stored and made accessible to traffic managers and service providers.

Relevance for MobiDataLab: the website that appears to be the project's reference (i.e., https://english.ndw.nu/) has no documentation in terms of deliverables and results achieved since the project was established. There appears to be a research paper associated with the project (Viti, F., Hoogendoorn, S.P., Immers, L.H., Tampère, C.M. and Lanser, S.H., 2008), but it dates back to 2008 and its purpose was to define the requirements behind the development of the aforementioned centralised data warehouse. Overall, it appears that the project did not consider cloud-based solutions, and even less the notion of "federation of clouds".

Table 10 : NDW coverage of dimensions relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	NOT ASSESSABLE
Data aspects	
Privacy and security	NOT ASSESSABLE
FAIR principles	NOT ASSESSABLE
Data access policies	POSSIBLY YES

Funded by the European Union

API access	NOT ASSESSABLE
Static vs. real-time data	NOT ASSESSABLE
Data trust	NO

More Info: https://english.ndw.nu/

2.2.11. Data for Road Safety Initiative

The European Commission-backed Data for Road Safety initiative, which originated from the activities, and is supervised by, the EU Data Task Force established in 2017, aims at the long-term deployment of a Safety-Related Traffic Information (SRTI) ecosystem, a scalable solution in which automobile manufacturers, automotive suppliers, road traffic authorities, EU member states and location technology providers commit to the long-term exchange of safety data in order to make roads safer and warn drivers about dangerous driving conditions.

Many vehicles are nowadays equipped with the latest technologies that detect dangerous road conditions and warn drivers — for example, when roads are slippery. Road operators can detect potential danger areas with their comprehensive infrastructure. However, these warnings are beneficial to other drivers and road operators, and the Data for Road Safety members believe that, regardless of which car brand or navigation application is used, all drivers across Europe should have access to a consistent minimum set of safety information that can help them make better informed decisions, thereby saving many lives.

Between June 2019 and October 2020, members of the Data for Road Safety partnership took the first step towards a harmonised exchange of safety-related data by conducting a proof of concept (PoC) test of a decentralised data collaboration architecture. During the PoC, tens of millions of messages were provided to the ecosystem. Vehicles demonstrated that they are able to produce data for five of the eight SRTI categories of EU Regulation 886/2013 (unprotected accident area, broken down vehicle, temporary slippery road, reduced visibility and exceptional weather conditions).

After the success of said PoC, a multi-party agreement[25] was then signed by the members in November 2020. The agreement defines the technical and organisational framework of how safety data from multiple brands and multiple countries can be made accessible and used within the SRTI eco-system in a fair and trustworthy manner. It also defines the roles and responsibilities along the SRTI value chain. The SRTI ecosystem is based on a reciprocity model – with safety data being offered in return for safety services.

[25] https://www.dataforroadsafety.eu/images/Documenten/Multi_Party_Agreement_-_SRTI_Ecosystem_-_Data_for_Road_Safety_final_bundled_PDF_signed_version.pdf

Relevance for MobiDataLab: this initiative covers some of the dimensions that we deem relevant for MobiDataLab, and considering this is an ongoing effort it is possible that other relevant dimensions will be covered over time – for this reason the evolution of this initiative should be kept under observation.

From the multi-party agreement document previously cited, it appears that the initiative considers some of the governance and regulatory aspects that are relevant for MobiDataLab, in particular those that deal with compliance with GDPR regulations and how to enforce intellectual property.

From the current version of the initiative's technical document available[26] , we see that the members already defined the high-level architecture of the future ecosystem.

For what concerns the *cloud federation aspects*, we observe that the use of a federation of clouds is not considered. The initiative's ecosystem rather relies on a conceptual model which is based on stages that the flow of data should follow. Such flow originates from entities (e.g., car manufacturers, state members) that produce data in its raw form (i.e., the so called "stage 1") and terminates into *end user services* (i.e., stage 4) that make use of processed data to serve end users. Within said model there exist two different types of clouds, namely, the *vehicle clouds*, which duty is to communicate with, and gather data from, a fleet of vehicles, and a *service clouds*, which purpose is to process data or provide services to end users. Finally, there appears to be a strong emphasis on defining or using well-established standards that data and entities within the infrastructure should comply with (to this end the reader may have a look at the DATEX II and SENSORIS standards highlighted in the technical document), which in turn covers the *technical compliance* dimension.

For what concerns the *data aspects*, there seems to be a strong emphasis on *privacy and security*, although said impression can be inferred only from the multi-party agreement as the technical document currently available do not specify how such dimension should be technically covered. *API access* seems to be also covered, as the technical document specifies how entities within the ecosystem should interact via interfaces. Finally, from the technical document we see how raw data produced in stage 1 will be provided in real-time, thus it is safe to assume that the initiative will have to cover the *Static vs. real-time* data dimension to make their ecosystem work appropriately.

Table 11 : Data for Road Safety Initiative coverage of dimensions relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	YES
Interplay between legal and technical infrastructures	NO
Contract - mobility data (re)use and sharing	NOT ASSESSABLE

[26] https://www.dataforroadsafety.eu/images/Documenten/Annex_1_-_Data_For_Road_Safety__Technical_Documentation_Version_1.01.pdf

Regulations relating to organisational aspects	NO
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	POSSIBLY YES
FAIR principles	NO
Data access policies	NOT ASSESSABLE
API access	YES
Static vs. real-time data	YES
Data trust	NO

More Info: https://www.dataforroadsafety.eu/

2.2.12. MobilityData

MobilityData is an international non-profit organisation based in Montréal, Canada, funded in 2019. Directed by a majority (two third) of European in the board of Directors, MobilityData is now strong of a team of 20 employees in Canada, United States, France and Germany, while having a European subsidiary, MobilityData Europe, based in Paris, France.

After only two years of existence, MobilityData has members on 5 continents, with one third (27 among 98) in Europe, including European private sector (e.g. HERE, HOVE, Ito World, Padam, Flucuto), European public sector (e.g. transport.data.gouv.fr in France, Entur in Norway, Samtrafiken in Sweden, STIB/MIVB in Belgium, AMB & EMT in Spain) and European non-profit like La Fabrique des Mobilités France.

The organisation addresses shared industry needs by systemic solutions, such as infrastructure, softwares, documentation and data models for public transit passenger information and shared mobility passenger information. The core work of MobilityData is to address challenges that require industry consensus and leave any ad-hoc implementation or needs to the private sector. It currently offers:

1. **Open resource centres**, for public transit data in GTFS (gtfs.mobilitydata.org) and shared mobility data in GBFS (gbfs.mobilitydata.org).

2. **Training**, through the MobilityData Academy, about the challenges and solutions for high quality public transit and shared mobility data. In the first half of 2021, the training has been provided in English, French, German & Spanish. Portuguese and Japanese are the next languages in development for the Academy.

3. **A world-wide repository of public transit datasets**, called The Mobility Database suite, already containing GTFS Schedule datasets, and being under expansion to add GTFS Realtime datasets, GBFS datasets and NeTEx datasets. More about that project in the subsequent section.

4. **Canonical validators** for GTFS Schedule datasets and GBFS datasets. To address the issue that the diversity of validators creates a diversity of interpretation of the data format. Open coded and hosted on GitHub, the GTFS Schedule validator is already being implemented by some major stakeholders of the industry.

5. **Canonical grading scheme** for GTFS Schedule, to go further than automated validation, and because a valid dataset is not always a high quality dataset. This Grading Scheme is already used by the US state of California on their public transit datasets. The process is open and hosted on GitHub, receiving contributions from various stakeholders of the industry.

6. **An open governance process**, to improve open data formats like GTFS Schedule, GTFS Realtime & GBFS.

Relevance for MobiDataLab: The MobilityData approach is an answer to two major challenges that the industry of data for mobility is facing:

- Open source projects supported by only one or two for-profits are often not widely adopted because their competitors see as a threat to rely on a competitor's technology (e.g. Mapzen).
- Independent open source projects often die after a few years by lack of funding (e.g. transitfeeds.com, gtfs-data-exchange.com).

MobilityData approach is to create consensus amongst a broad variety of stakeholders: public or private entities, from many continents and cultures. Those stakeholders are funding and co-building a suite of tools that can be shared and reused by each of them, regardless if they are members or not. As a non-profit, the organisation cannot and does not intend to do the work alone, which is why the solution is to host the projects and collaborate with the stakeholders to co-build them. When a for-profit is financially supporting a project, they can improve or expand a tool hosted by MobilityData; when a public entity wants to improve a tool, they can maximise the use of public funds by improving an existing and reliable product.

Table 12 : MobilityData coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	

Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	NOT ASSESSABLE
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	NOT ASSESSABLE
Static vs. real-time data	NOT ASSESSABLE
Data trust	YES

More Info: https://mobilitydata.org/

2.3. Mobility Data Sharing platforms

2.3.1. Navitia.io

Navitia.io is a portal administrated and maintained by HOVE, one of the partners of MobiDataLab, giving access to these two main services:

- **Open data portal**: enriched public transport data sets in over 30+ countries. These open data sets are retrieved by HOVE from many diverse locations. They are enhanced and enriched by HOVE, for example stop points are merged into stop areas, connections are added. Quality checks are also performed. Once processed, these data sets are distributed through Navitia.io platform. Anyone can download them from this portal, without registration.

- **Open Navitia API**: Navitia is the open source trip planner and traveller information system developed and maintained by HOVE. The Navitia instance that is accessible through Navitia.io integrates all the open data sets that are made available by HOVE though Navitia.io open data portal. In addition, some coverages also integrate real time data that are not available for download through the open data portal (since only static data can be retrieved from the open data portal). Access to this Navitia instance requires user to be registered. It is completely free, up to 20000 requests/month. Main features: journey planning (multi modal transport), next departures and arrivals, timetables, places nearby, isochrones.

Navitia.io is also the main portal for developers and community around Navitia:

- Access to documentation and source code (available on Github)
- Access to Navitia Playground: this tool allows developers to test their queries and to visualise the answer in a easy and straightforward way
- Access to community group discussion

Navitia.io open data portal is based on OpenDataSoft, while Navitia.io open API is entirely hosted and managed by HOVE.

Relevance for MobiDataLab: Navitia.io provides a relevant example of data aggregating platform (restricted to open public transport data sets at this stage. It is expected that open data sets available through Navitia.io will also be available through MobiDataLab. A Navitia instance will also be made available so it is possible to use Navitia API to exploit these data sets for journey planning and passenger information purposes.

Table 13 : Navitia coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT APPLICABLE
Interplay between legal and technical infrastructures	NOT APPLICABLE
Contract - mobility data (re)use and sharing	NOT APPLICABLE
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	YES
Data aspects	

Privacy and security	YES
FAIR principles	NOT APPLICABLE
Data access policies	YES
API access	YES
Static vs. real-time data	YES
Data trust	YES

More Info: https://www.navitia.io

2.3.2. SNCF (French national railway)

SNCF is the French national railway operator. They provide two services that allow them to share mobility data:

- **Open data portal**: https://ressources.data.sncf.com/pages/accueil/
- **Open API**: https://www.digital.sncf.com/startup/api

The SNCF data portal gives access to 224 datasets. Some of them are retrieved by HOVE and integrated with other open datasets available through Navitia.io, i.e.:

- TER and Intercités (local trains)
- TGV (national lines)

The SNCF data portal is based on OpenDataSoft. The SNCF open API service is actually based on Navitia. SNCF hosts a Navitia instance with all SNCF data integrated. This includes local and national lines, but also international lines (Lyria, Eurostar, Deutsche Banh) and other transports (night buses). Users of the SNCF API can send requests to calculate trips, retrieve timetables and so on. Calls to the SNCF API are limited to 5000 requests per month.

Relevance for MobiDataLab: SNCF data portal and SNCF open API provide examples of mobility data sharing platforms. However, no direct integration of SNCF data portal or open API is planned, since SNCF data is already available through Navitia.io and SNCF API is actually based on Navitia API.

Table 14 : SNCF coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	

Substantial rights and major legal frameworks relevant to data and data transactions	NOT APPLICABLE
Interplay between legal and technical infrastructures	NOT APPLICABLE
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	NO
Data aspects	
Privacy and security	YES
FAIR principles	NOT APPLICABLE
Data access policies	YES
API access	YES
Static vs. real-time data	YES
Data trust	YES

More info: https://ressources.data.sncf.com/pages/accueil/

2.3.3. PAN (French government)

PAN stands for “Point d’Accès National” (national access point). This is the platform where every city and region in France is supposed to upload their transport datasets and make them available to everyone. This is mandatory since 2019 (mobility orientation law). Most large cities in France (Paris, Lyon …) upload their public data sets on this portal. However, some smaller cities are not uploading their data yet: most of them because they have no budget and no incentive to do so (although upload is supposed to be mandatory, there is no enforcement), some of them because they do not want to do so. In any case, the number of cities that are sharing these data is still growing.

It shall be noted that HOVE is actually retrieving all these datasets from transport.data.gouv.fr and process them for HOVE customers using HOVE internal tools (quality enhancements). As required by the open license used for these datasets, HOVE redistribute them through Navitia.io (see above

paragraph). It is therefore usually better for people who want to get access to open transport datasets in France to download them from Navitia.io rather than from transport.data.gouv.fr. HOVE is not aware of such other initiative in France (i.e. another data portal where anyone could download enhanced and enriched open datasets).

Relevance for MobiDataLab: PAN provides an example of transport data sharing platform at the national level, backed by a government. However, no direct integration of this portal is planned, since data available through this portal is already available through Navitia.io data portal.

Table 15 : PAN coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	YES
Interplay between legal and technical infrastructures	YES
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	NO
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	NO
Static vs. real-time data	NO
Data trust	YES

More info: https://transport.data.gouv.fr/

2.3.4. OpenDataSoft

The OpenDataSoft platform is a cloud-based data publishing solution that lets users share their data easily. Cities, transportation providers, government administrations, and other private-sector companies use the platform to publish, visualise, and share data in order to break down silos and facilitate data reuse. Data become easily understandable thanks to charts, tables, search, filtering, and mapping features. In turning all datasets into APIs with aggregation, faceting, query, and real-time capabilities, data can be reused in applications and new services that create new or improved services employable, for instance, in a Smart City environment.

It is important to note that OpenDataSoft by itself is not restricted to mobility data, nor does it host any data -- it is a tool used by other companies or providers to host data.

Relevance for MobiDataLab: In the context of the MobiDataLab project, it is used for example:

- By HOVE, to redistribute public transport open data sets, through Navitia.io
- By SNCF, to let people access train data

Table 16 : OpenDataSoft coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT APPLICABLE
Interplay between legal and technical infrastructures	NOT APPLICABLE
Contract - mobility data (re)use and sharing	NOT APPLICABLE
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	NOT APPLICABLE
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	NOT APPLICABLE

Data access policies	NOT APPLICABLE
API access	YES
Static vs. real-time data	NOT APPLICABLE
Data trust	NOT APPLICABLE

More info: https://www.opendatasoft.com

2.3.5. CKAN

CKAN is a tool designed to make open data websites – something akin to a content management system (e.g., WordPress) targeting data rather than pages and blog posts. CKAN helps in managing and publishing collections of data. It is used by national and local governments, research institutions, and other organisations that collect a lot of data.

Once data is published, users can use its faceted search features to browse and find the data they need, and preview it using maps, graphs and tables – whether they are developers, journalists, researchers, NGOs, citizens, and so on.

CKAN is an open source software, with an active community of contributors who develop and maintain its core technology. CKAN is modified and extended by an even larger community of developers who contribute to a growing library of CKAN extensions.

Relevance to MobiDataLab: similar to OpenDataSoft, CKAN itself is not specifically related to mobility data and does not host any data. It is a tool used by other companies or providers to publish their data though a web portal. This might be relevant for the Transport Cloud to offer data sharing via web portal (thus not only via an API).

Table 17 :CKAN coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT APPLICABLE
Interplay between legal and technical infrastructures	NOT APPLICABLE
Contract - mobility data (re)use and sharing	NOT APPLICABLE
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	

Federated access control and authentication methods	NOT APPLICABLE
Cloud Services type: IaaS, PaaS, SaaS	NOT APPLICABLE
Technical compliance	YES
Data aspects	
Privacy and security	NOT APPLICABLE
FAIR principles	NOT APPLICABLE
Data access policies	NOT APPLICABLE
API access	NOT APPLICABLE
Static vs. real-time data	NOT APPLICABLE
Data trust	NOT APPLICABLE

More info: https://ckan.org/

2.3.6. Chordant CONVEX

Chordant Convex proposes a.platform to share, transform, and integrate dynamic mobility data between organisations and systems. The platform proposes a catalogue of UK based mobility datasets to be shared and searched. Convex provides organisations a self-managed automated framework to arrange agreements through licensing, pricing, and sharing levels.

In fact, there is a data monetisation approach through which datasets can be shared according to some price and licence of use. The real-time nature of Convex means that data is available to be consumed very quickly after it is provided, with Convex accepting and providing data at a rate of one sample per second (1Hz). Data in Convex can be consumed either automatically (through APIs), or manually via the Convex Service Portal.

For Data Providers the platform provides:

- Access Control (which Organisations have access to the Dataset)
- Pricing (how much the Data Provider charges for consuming data from the Dataset)
- Licensing (the terms under which Users of Organisations can use the data)

Data Consumers can access through a Data Catalogue enabling:

- Dataset descriptions (detailed information explaining the type of data)
- Licensing (the terms under which data can be used by consumers)

- Pricing (the pricing charged for consuming the data)

For both Data Providers and Data Consumers:

- Applications provide authorised Users of Organisations access to Convex APIs
- Insights into data consumed
- Insights into the consumption of data provided
- User administration

The Convex APIs provide several methods for providing and consuming real-time data.

- Streaming API – To continuously stream data in real-time from Data Providers to Data Consumers
- oneM2MTM API – Provides a machine to machine IoT standard method of sharing real-time data www.onem2m.org1.
- Files API – Provides access to provide or consume file-based data
- API Proxy – Provides a secure, Licenced, and metered access to data stored externally

Relevance for MobiDataLab: this platform is very relevant for MobiDataLab. Although it has been initially developed for exchanging data of autonomous vehicles, now it provides more general mobility datasets located in UK. They provide access control, real-time access to data, monetisation data access, and API access as well. However, this platform is not a federation, and how to deal with the privacy aspect is not clearly described.

Table 18 : CONVEX coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NO
Cloud Services type: IaaS, PaaS, SaaS	NO
Technical compliance	NOT ASSESSABLE
Data aspects	

Privacy and security	NOT ASSESSABLE
FAIR principles	NOT ASSESSABLE
Data access policies	NOT ASSESSABLE
API access	YES
Static vs. real-time data	YES
Data trust	YES

More info:

https://convexglobal.io

https://convexglobal.io/guides

2.3.7. The Mobility Database suite

The Mobility Database suite is a set of few products administered and maintained by the international non-profit organisation MobilityData (see previous section 2.2.12), which counts as their members organisation like HERE & HOVE, who are partners of MobiDataLab. The Mobility Database suite is offering multiple services, such as:

- Open index of datasets, The Mobility Database (mobilitydatabase.org): base on the open software of semantical database WikiBase developed by the WikiMedia Foundation (the organisation is also in charge of Wikipedia), The Mobility Database contains reference to thousands of public transit dataset worldwide, and automatically extract metadata about their location, size, and temporal scope. Linked with the GTFS Schedule Canonical Validator, it provides the canonical errors and warnings of the datasets. All the APIs of WikiBase are the same as the ones on WikiData.org, the WikiMedia project using WikiBase. Launched in June 2021 with GTFS Schedule dataset, it is going to be extended to GBFS datasets, GTFS Realtime datasets and NeTEx datasets.
- Open registry of datasets, The Mobility Archives (mobilityarchives.org): record of the mobility datasets published through time, to keep a historical archiving of them, and to give the choice to reusers to use the version hosted by the provider or by the Mobility Archives. It can be used by mobility information providers, but also by researchers.
- OpenMobilityData: User friendly interface for the data in the Mobility Database. By separating the database and its interface, MobilityData intends to offer the possibility to other stakeholders to provide other services and other views of the data. For example, a personalised portal for a specific state or area.

Relevance for MobiDataLab: To be successful, data aggregation platforms have to be perceived as:

- **Neutral**: not dependant of a specific for profit entity which could decide to limit features unilaterally to its competitor;
- **As exhaustive as possible**: not limited to a specific country or territory, like the national access points or the regional open data portals;
- **Resilient**: it should survive the departure of any of its financial supporters, by being supported by a large diversity of stakeholders, public like privates, from different countries.

Only at those conditions will organisations trust a data aggregation platform, and replace their existing structure with the one presented above, The Mobility Database suite MobilityData has been created to address those challenges and enable such transformation in the industry.

Table 19 : MobilityDataBase Suite coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NOT APPLICABLE
Cloud Services type: IaaS, PaaS, SaaS	NOT APPLICABLE
Technical compliance	NOT APPLICABLE
Data aspects	
Privacy and security	NOT ASSESSABLE
FAIR principles	NOT ASSESSABLE
Data access policies	YES
API access	NOT ASSESSABLE
Static vs. real-time data	NO
Data trust	NOT ASSESSABLE

More Info: https://mobilitydata.org

2.4. Federated Cloud-based Data Sharing Technologies

A federated cloud data sharing is a software platform that ideally allows to seamlessly share data volumes among members of the same (virtual) organisation or with other people. Data sharing is a fairly common scenario for private users and large organisations whose data sources are geographically sparse. Here we report on the most promising technological, commercial solutions that can support the federated cloud data sharing of MobiDataLab. Recent cloud storage services for end-users have much commercial success due to their ease of integration with consumer devices, desktop environments, and web clients. Such services, such as Dropbox or Google Drive, rely on a backend usually deployed in public or private highly-available data centres.

However, private and public commercial and research organisations do not always desire to adopt public cloud computing solutions. Such organisations usually place great importance on privacy and data ownership. In public clouds, data can reside in many different geographical areas (i.e., Amazon availability zones) under different data protection laws, which increases the confusion in the end-users. Therefore, on-premise cloud solutions get much attention from research organisations as a realistic alternative. They offered the potential decentralisation of the data access model, in which organisations have access to a data federation and also maintain a part of the data backend infrastructure.

Many products that encourage the use of on-premise cloud computing have been initially designed for private use, and then they evolved to solutions that large organisations can use. It is the case for one of the most prominent solutions, OpenCloudMesh already described in Section 2.2.3

2.4.1. OwnCloud

OwnCloud is an open-source enterprise-grade file sync, share and content collaboration software focused on scale and security. It is widely used by several industries and institutions including the European Commission. OwnCloud includes the *Federated Cloud Sharing*[27] feature, which enables users to quickly and easily share files with users belonging to other OwnCloud server instances. Users can simply enter a Federated Cloud ID, rather than a local username, to share a file or folder with other parties, who can then have access to and edit the file. The federated cloud sharing feature is particularly useful to promote collaboration across different organisations (e.g., between companies, universities, public institutions, and so on).

We report that the price listing available in the OwnCloud's web site includes free of charge and various enterprise versions. These prices are essentially different kinds of support subscriptions and it seems it is not required to purchase the software. OwnCloud Standard Edition is licensed under

[27] https://owncloud.com/features/federated-cloud-sharing/

the GNU AGPLv3 and the core code is covered by the OwnCloud Contributor License Agreement (CLA). OwnCloud Enterprise Edition is not open source and is licensed under the OwnCloud Commercial License[28].

In OwnCloud file access control is available for enterprise version. It provides API and external storage integration.

Relevance for MobiDataLab: This software is very relevant for MobiDataLab since it offers a solid and maintained technological platform to share data, with said platform supporting also the notion of federated cloud. Furthermore, it is widely used in both industry and academia, and it is highly configurable in terms of privacy, security, legal aspects, as well as data access control mechanisms and compatibility with other platforms. It also offers APIs to access shared data from external applications that is an important aspect for MobiDataLab.

Table 20 : OwnCloud coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	YES
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	NOT APPLICABLE
Cloud Services type: IaaS, PaaS, SaaS	NOT APPLICABLE
Technical compliance	NOT APPLICABLE
Data aspects	
Privacy and security	NOT ASSESSABLE
FAIR principles	NOT ASSESSABLE
Data access policies	YES

[28] https://civihosting.com/blog/nextcloud-vs-owncloud/

API access	NOT ASSESSABLE
Static vs. real-time data	NO
Data trust	NOT ASSESSABLE

More info: https://owncloud.com/federated-cloud-sharing/

2.4.2. NextCloud

Nextcloud is a 2016 fork of the old OwnCloud code base that still uses PHP and seems focused on implementing lots of features for small businesses and home users.

NextCloud provides a common file access layer keeping data where it is and retaining the management and control mechanisms currently in place to manage risk by leveraging existing management, security and governance tools and processes. Nextcloud brings data from cloud storage, Windows network drive and legacy data storage to users empowering them to access, sync and share files.

Nextcloud gives control over data access by enabling the definition of strict rules requests: when users in certain groups or geographic regions should not be given access to certain file types, or if data with a specific tag should not be shared outside the company, administrators can make sure their Nextcloud instance enforces these rules.

Nextcloud includes the Federation feature that enables transparent file sharing between users on different server instances. To this end users can use a 'Federated Sharing ID', a unique identifier consisting of the user name and server address of a Nextcloud server. When a file is shared between two Nextcloud servers, the server can optionally exchange address books, enabling auto-completion of user names on the other server. The system administrator can enable or disable this functionality and manually add or remove trusted servers.

Communication between Nextcloud servers takes place through a REST API, standardised in the Open Cloud Mesh initiative and also implemented by some other cloud vendors. File exchange is based on WebDAV and uses standard TLS based security.

Files remain on the server they were shared from, ensuring that the user who owns the file and the server admin remain in full control of the data.

This solution has a pricing plan depending on the number of users and the required support.

The Nextcloud has invested a lot in improving security for their product, such as better password handling, brute force protection, rate limiting, integration with various enterprise authentication methods, credential service provider (CSP) and more.

In NextCloud file access control is available for base version.

Relevance for MobiDataLab: Similar to OwnCloud this solution is relevant for MobiDataLab since it provides a file sharing software that can be personalised and with API support. To evaluate as alternative to OwnCloud. Interesting comparison between NextCloud and OwnCloud is at the link https://civihosting.com/blog/nextcloud-vs-owncloud/

Table 21 : NextCloud coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT APPLICABLE
Interplay between legal and technical infrastructures	NOT APPLICABLE
Contract - mobility data (re)use and sharing	NOT APPLICABLE
Regulations relating to organisational aspects	NOT APPLICABLE
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES
Data access policies	YES
API access	YES
Static vs. real-time data	YES
Data trust	YES

More info: https://nextcloud.com

2.4.3. CORTEZA

Corteza is an open-source platform based on Low Code developed entirely in the public domain for building applications. Corteza Low Code is a self-managed cloud solution that includes identity management infrastructure and federated identity, application menus and user interfaces, and a Low Code development environment. Recently the founder company, Crust Technology, has received funding by the European Commission to deliver the Corteza platform as a federated cloud solution for record sharing. Crust will enable the platform to share selected components of data layer or entire data models, on a one-to-many or many-to-many basis. The new features will allow business data, but also open datasets created by public sector organisations and academia, to be shared. Data sharing can be done at the record level.

Relevance for MobiDataLab: The Corteza platform is currently being developed, and it might be interesting to monitor the federated extension to see the options available, as the data access level can be limited to single records. We also report, however, that the dashboard seems to target business applications rather than proper data sharing scenarios.

Table 22 : CORTEZA coverage of dimensions that are relevant for MobiDataLab

Dimensions	Coverage of dimension is relevant w.r.t. MOBIDATALAB's goals?
Governance and regulatory aspects	
Substantial rights and major legal frameworks relevant to data and data transactions	NOT ASSESSABLE
Interplay between legal and technical infrastructures	NOT ASSESSABLE
Contract - mobility data (re)use and sharing	NOT ASSESSABLE
Regulations relating to organisational aspects	NOT ASSESSABLE
Cloud federation aspects	
Federated access control and authentication methods	YES
Cloud Services type: IaaS, PaaS, SaaS	YES
Technical compliance	YES
Data aspects	
Privacy and security	YES
FAIR principles	YES

Data access policies	YES
API access	NOT ASSESSABLE
Static vs. real-time data	NOT ASSESSABLE
Data trust	YES

More info:

https://cortezaproject.org/

https://www.crust.tech/european-commission-funding-for-open-source-federated-cloud-platform/

https://cortezaproject.org/corteza-federation-proof-of-concept/

2.4.4. Other solutions

Other file sharing solutions are SeaFile (https://www.seafile.com), Pydio (https://pydio.com) and PowerFolder (https://www.powerfolder.com). We report that these solutions do not provide federation and rely on Owncloud (to which they can be connected). Therefore, they are not specifically relevant for MobiDataLab as technical solutions.

Other research solutions are present in the context of cloud synchronisation and sharing services (CS3) for Science, Education, and Research (see Mościcki, Jakub T., and Luca Mascetti, 2018). CS3 aims to provide new ways of accessing, sharing, and interacting with data repositories and storage services. Recent CS3 services are usually described in terms of research endeavours and include: CERNBox (Labrador, H.G., Alexandropoulos, G., Bocchi, E., Castro, D., Chan, B., Contescu, C., Lamanna, M., Presti, G.L., Mascetti, L., Moscicki, J. and Musset, P. 2019), CloudSTOR (Mason, W.G., Aben, G., Meijer, J.J., Richter, C., Bonnington, C.P., Moresi, L. and Betts, P.G., 2010), SCIEBo (Vogl, R., Rudolph, D., Angenent, H., Thoring, A., Schild, C., Stieglitz, S. and Meske, C., 2015). Such services are typically designed with the particular requirements of some scientific community in mind, thus enabling new forms of collaboration between researchers that allow to achieve results faster and more effectively.

3. SWOT analysis

For the SWOT analysis, we identified the most promising solutions with highest potential to be adopted or exploited in MOBIDATALAB. This identification is based on an analysis of the dimensions as reported in the previous chapters, the relevance to MobiDataLab and the results of internal discussions and feedbacks with consortium partners.

In this SWOT analysis we highlight their strengths and weaknesses as well as the opportunities and threats that could derive from their exploitation in MobiDataLab. It is to be noted that, unlike common SWOT analyses, here we refer as *Strengths* how good some solution is for MOBIDATALAB, *Weakenesses* how inadequate it is for MobiDataLab, *Opportunities* are the opportunities in exploiting some solution for MobiDataLab, while *Threats* represents the potential drawbacks that may arise from adopting the solution.

Table 23 : SWOT analysis of enabling solutions for Transport Cloud

Solution	Strengths	Weaknesses	Opportunities	Threats
Gaia-X	GAIA-X is extremely relevant, its scope very general and might cover many of the MobiDataLab's needs.	No major weaknesses.	MobiDataLab can be supported by a European federated cloud infrastructure.	The technology is not ready yet, a first release was planned by early 2021 but due to the pandemic there are delays.
FEDeRATED	FEDeRATED is extremely relevant, it focuses on a federated solution for data and services sharing in logistics, and might include the MobiDataLab needs.	No major weaknesses.	MobiDataLab benefit from a federated infrastructure to share data and services among different data producers.	The technology is not ready yet, the project started in 2018 and it is ongoing, no software is available, it will be developed to be used in 15 virtual labs in the logistic domain.
SoBigData++	This project is partially relevant since it proposes a research infrastructure for sharing data and analysis algorithms.	It is not a federated approach since it is centralised.	The research environments might be of inspiration for the data processors of MobiDataLab.	As a centralised approach this is not a suitable solution for the MobiDataLab's Transport Cloud.
BASMATI	The conceptual architecture of BUDaMaF could be considered for the creation of the data plan of the transport Cloud.	The scope of BUDaMaF is the seamlessly relocation and offloading of data among federated cloud providers for mobile applications.	It allows dynamic management of application data, enabling the assisted migration, movement and replication of mobility data depending on the	The software does not provide support for federated data sharing and it is not addressing many dimensions. Thus, it does not provide support for many of the MobiDataLab 's needs.

		Prototype software is available, but it is not maintained, and it does not address many of the aspects highlighted in this deliverable e.g., FAIR principles, data access policies at different granularities, trust, technical compliance, privacy, enforcement of legislative and legal frameworks.	actual needs of the project.	Furthermore, it is not maintained anymore, therefore it is risky to adopt this solution.
FENIX	Fenix is relevant for MobiDataLab since it represents a specific instantiation of a federated infrastructure applied to the logistic domain. The MobiDataLab project can thus take advantage of the ongoing work and results from FENIX to better design the federated architecture, and to take inspiration on how the different dimensions are instantiated case by case by the different pilot sites.	The objectives of FENIX are different from MobiDataLab, therefore we cannot apply their solution or technologies directly to our case.	The study of the use cases can be of interest for MobiDataLab specifically for the data sharing solutions.	The project is ongoing since just started in 2019, therefore at the moment they do not provide any solution that can be adopted by MobiDataLab.
SUNFISH	Sunfish is very relevant for MobiDataLab since it provides Federation-as-a-Service (FaaS), to offer a service that federates private and public clouds for public administration.	The software, although available in a github repository, does not appear to be maintained anymore and this can be a risk for MobiDataLab.	A strong emphasis on data sharing policies and Data privacy and security, as Sunfish employs data masking, data anonymisation, and secure multi-party computation including micro data and macro data	Adopting a non-maintained software is very risky for MobiDataLab.

			anonymisation services including k-anonymity.	
CKAN	It is a CMS for data sharing and it is very suitable in case MobiDataLab wants make data available with a web portal besides an API.	It does not provide cloud services neither federation therefore it is not a solution for Transport Cloud federation.	To be used if Transport Cloud will provide datasets access via web.	--
CONVEX	It is very relevant for MobiDataLab since it shares general mobility datasets located in UK. They provide access control, real-time access to data, the monetisation data access and the API access.	This platform is not a federation and how to deal with the privacy aspect is not clearly described.	This platform implements real-time access and data monetisation and this can be of interest for MobiDataLab.	Since it does not provide a technical solution for a federated data sharing this cannot be directly used in MobiDataLab transport cloud implementation.
OwnCloud / NextCloud	It offers a solid and maintained technological platform to share data, the notion of federated cloud. Furthermore, it is highly configurable in terms of privacy, security, legal aspects, as well as data access control mechanisms and compatibility with other platforms. It also offers APIs to access shared data from external applications	It is not specific for mobility data and some features are for Enterprise version only in OwnCloud (e.g. no rate limit)	It allows different OwnCloud / NextCloud instantiation to easily share data in a federated way,	The federation should be formed by OwnCloud /NextCloud installations, e.g. each interested data producer might make available datasets with OwnCloud/NextCloud.
Corteza	It as an ongoing development of the platform, and it might be interesting to monitor the federated extension to see the options available since here the data access level can be one at the record level.	The dashboard is more targeted for business applications then for a proper data sharing. Furthermore, the software is not available yet.	The data access level at the record granularity is an interesting feature to consider.	Not fully supporting the MobiDataLab needs and software not available therefore not possible to adopt this solution.

4. Solutions enabling the processors

The MobiDataLab transport cloud aims to facilitate access to mobility data on the web through a federation of cloud services, but it will also include an open data catalogue referencing transport datasets and corresponding metadata, and provide data enrichment techniques through so-called “data processors” (i.e. semantic enrichment, geospatial combination, complementary presentations, etc).

The strategy of the consortium for achieving this purpose is to develop open tools and/or contribute to already existing ones. The aim of this section is to explore the solutions on which these components of the MobiDataLab Transport cloud can be favourably built. The selection of which of the following frameworks the project will use, and which it will not, will be made during the implementation phase, namely the WP4. Nevertheless, we propose here a preliminary evaluation of these solutions with regard to their suitability for MobiDatalab, as in the previous section. The dimensions for assessing this suitability are however specific, depending on the kind of processing and/or data enrichment they make possible and therefore these solutions are not included in the SWOT analysis.

4.1. Solutions for cataloguing mobility data

4.1.1. Open Data Catalogues

Open Source catalogues are nominally “free,” in that they may be acquired via download for no cost, and may be modified or customised without restriction or licensing fees. These products can be hosted on the owners’ own dedicated servers or on cloud-based infrastructure. In contrast, Software as a Service (SaaS) products are available from various vendors for a monthly or annual fee, and vendors assume responsibility for IT management, security and software updates. Commonly used open data platforms are OpenDataSoft and CKAN (see section 3 above on Mobility data sharing platforms).

Related standards: DCAT, DCAT-AP

Relevance for MobiDataLab: As stated in the previous section, OpenDataSoft and CKAN solutions will be interesting solutions to be integrated in the MobiDataLab transport cloud, as they provide support to DCAT standards. Also, they allow to provide access to mobility datasets via a web portal, and thus not only via an API. These solutions are already widespread in the transport sector and used by several stakeholders.

4.1.2. Geospatial data catalogues

To manage easily geospatial data, solutions exist and may be added or plugged to Open Data platforms via APIs or extensions. The most popular open source catalogues with geographical

dimension using the OGC Catalogue Service for the Web (CSW) standard CSW and the recommendations from the INSPIRE Directive:

- **GeoNetwork**[29] : is a catalogue application to manage spatially referenced resources. GeoNetwork provides powerful metadata editing and search functions as well as an interactive web map viewer.
- **GeoNode**[30] : is an open source platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures. It is designed to be extended and modified and can be integrated into existing platforms.
- **Esri Geoportal Server**[31] : is a free, open source product that enables discovery and use of geospatial resources including datasets, rasters, and Web services.

Related standards: CSW, INSPIRE

Relevance for MobiDataLab: Geospatial catalogues help to manage spatially referenced resources providing powerful metadata editing and search functions as well as an interactive web map viewer. These solutions, although very rich and functional, are rather reserved for experienced users, familiar with the particular standards of GIS, such as geomatics experts and spatial analysts. They will not necessarily be integrated into MobiDataLab, at least not at first.

4.2. Solutions for accessing mobility data (APIs)

Having all the data is not enough, we could create value if we were able to easily combine them. In fact, we can do much less with data alone than with data which is cross-referenced with other data. The MobiDataLab data enrichment components aim to build richer and more powerful datasets in a synergic and complementary way. The foundation for this combination is APIs. APIs enable real-time data usage, filtering on request, and the ability to work with the data at an atomic level.

4.2.1. Open APIs for accessing mobility data

In order to complement the open mobility data catalogue, the MobiDataLab transport cloud will also provide a library of open APIs for accessing mobility data. The Transport cloud will therefore act as a **gateway towards relevant APIs**. Many APIs allow accessing mobility data, provided by different organisations, associations, public actors or private companies - starting with consortium partners HERE and HOVE. The following list is not intended to be exhaustive but most of them will be referenced in the MobiDataLab data library.

[29] https://geonetwork-opensource.org/
[30] http://geonode.org/
[31] https://www.esri.com/en-us/arcgis/products/geoportal-server/overview

Navitia API[32] : this open API developed by HOVE provides multi-modal journeys computation, line schedules, next departures, exploration of public transport data, search & autocomplete on places, isochrones, etc. Navitia is already reported in Section 2.3.1

HERE REST APIs[33] : HERE REST APIs provide developers with maps, weather information, batch geocoding, comprehensive routing, etc. In particular the public transit API, the Intermodal routing API and the Traffic API are dedicated to mobility data.

OpenMobilityData[34] : this API allows accessing programmatically information about transit feeds to ease integration into existing systems.

OpenTripPlanner APIs[35] : OpenTripPlanner is an open source multi-modal trip planner, focusing on travel by scheduled public transportation in combination with bicycling, walking, and mobility services including bike share and ride hailing. OTP exposes REST and GraphQL APIs that can be accessed by various clients including open source Javascript components and native mobile applications.

TOMP-API [36] implementations: The Transport Operator and MaaS Provider API is not an API per se but rather a standard that was created in the context of MaaS to fill the need for an industry standard booking API allowing to book taxis, ride share, trains, coaches, ferries, scooters, city bikes, parking spots, and even electric vehicle charge spots, etc. The MaaS Alliance endorsed the standardisation work, it was proceeding in the Netherlands and now the MaaS industry in Europe is coalescing around this standard. Some TOMP-API implementations are already made available by several organisations as a complement of their APIs like Skedgo API, CityWay API, WeMobile sharing, Mobian, inTaxi solutions, etc.

Overpass API[37] : this API serves up custom selected parts of the OpenStreetMap data. It acts as a database over the web: the client sends a query to the API and gets back the dataset that corresponds to the query. The Overpass API is optimised for data consumers that need a few elements within a glimpse or up to roughly 10 million elements in some minutes, both selected by search criteria like e.g. location, type of objects, tag properties, proximity, or combinations of them.

Other APIs: like Citybikes API , EnviroCar API , etc.

Related standards: JSON/GeoJSON, NeTEx, GTFS, GBFS, DATEX II, OSM

Relevance for MobiDataLab: These APIs are fully relevant for MobiDataLab, as they are an essential source of mobility data, following an open approach and the FAIR principles (access through standardised APIs and reusable data). In addition, they comply with mobility standards facilitating the interoperability (GTFS, NeTEx, etc.).

---

[32] https://www.navitia.io/
[33] https://developer.here.com/develop/rest-apis
[34] https://openmobilitydata.org/api/swagger/
[35] http://docs.opentripplanner.org/en/latest/
[36] https://tomp-wg.org/
[37] https://overpass-turbo.eu/

4.2.2. Frameworks for building API Gateways

In order to facilitate mobility data enrichment, the MobiDataLab Transport cloud will in turn expose an API gateway consisting of an “API mashup” aiming to combine the above-mentioned sources of data upon specific criteria like e.g. common geometries. The most well-known frameworks for building APIs are:

- Spring Boot (Java)
- Django (Python)
- Express.js (JavaScript)
- Ruby on Rails (Ruby)

Related standards: REST, HTTP/HTTPS

Relevance for MobiDataLab: These frameworks are very relevant for delivering the MobiDataLab transport cloud functionalities. Indeed, the development of an API gateway is needed to be able to combine different APIs (in a “plug and play” mode) and expose a well-documented API to be used by MobiDataLab hackathon participants. Elicitation of one of these frameworks will depend on the consortium's skills in the said frameworks and/or corresponding programming languages and will be decided during WP4.

4.3. Solutions for the geospatial enrichment of mobility data

The geographical dimension is a mainstay of mobility data and thanks to Geographic Information System (GIS) technologies. All information referenced in a database or an Open Data catalogue can be displayed on a map, as long as it contains a spatial reference (namely coordinates).

4.3.1. Geodata storage

4.3.1.1. Relational databases

PostgreSQL/PostGIS: PostgreSQL[38] : is an open source Relational Database Management System (RDBMS) developed by a worldwide team of volunteers. PostgreSQL is not controlled by any corporation or other private entity and the source code is available free of charge. PostgreSQL runs on all major operating systems, has been ACID-compliant since 2001, and has powerful add-ons such as the popular PostGIS geospatial database extender

[38] https://www.postgresql.org/

Funded by the European Union

Related standards: GIS Vector model, GIS Raster model, INSPIRE

Relevance for MobiDataLab: It might be necessary for the MobiDataLab transport cloud, in some specific cases, to make it possible to store datasets in a structured way. For example, to allow potential data providers to experiment with anonymisation and enrichment features with a static geo-referenced dataset, before publishing it on a data portal. PostgreSQL/PostGIS is a good solution in this case, as it is free and open source, and also provides support to geospatial data.

4.3.1.2. Non-relational databases

A non-relational database, also called NoSQL databases, is a database that does not follow the tabular schema with rows and columns. The most popular NoSQL databases providing support for geo-referenced data are:

MongoDB[39] : used for high volume data storage. Instead of using tables and rows as in the traditional relational databases, MongoDB makes use of collections and documents. Documents consist of key-value pairs which are the basic unit of data in MongoDB. Collections contain sets of documents and function which is the equivalent of relational database tables. MongoDB supports query operations on geospatial data as GeoJSON objects.

ElasticSearch[40] : is a document-oriented search engine, designed to store, retrieve and manage document-oriented, structured, unstructured, and semi-structured data, textual, numerical and geospatial data (geo_point and geo_shape).

Cassandra[41] : a highly scalable and available distributed database that facilitates storing and managing high-velocity structured data across multiple commodity servers without a single point of failure.

Relevance for MobiDataLab: Relational databases are a common type of database, but they are not suitable for storing large amounts of unstructured data. NoSQL databases are useful for dealing with large datasets, which are heterogeneous in format or change frequently. This is the case for real-time data, for example data under the GTFS-RT format, which can be analysed using solutions such as MongoDB. If we consider the static vs. realtime dimension, NoSQL solutions would be the dynamic counterpart of relational databases and need to be integrated into the transport cloud.

4.3.2. Geodata servers

Representing mobility data on a map is possible thanks to specific platforms that manage GIS data and facilitate interoperability between several systems. With the democratisation of the web, the

[39] https://www.mongodb.com/
[40] https://www.elastic.co/fr/
[41] https://cassandra.apache.org/

Funded by the European Union

implementation of geographic solutions has become easier and as with the database, geographic data can be stored on a server and can be called up by standard web services or standard formats issued by the OGC (WMS, WFS, WMTS). To provide geographic access to data, two philosophies coexist with open source solutions (e.g. Free and Open Source Geospatial Software) and proprietary solutions (e.g. Esri solutions).

Related standards: WMS, WMTS, WFS, OGC API Features, Vector tiles

4.3.2.1. Open Source GIS servers

Open-source web GIS has evolved as a cheaper and easier way of disseminating geospatial data. It does not require users to invest in a resource-intensive machine. Combined to a Web Server like Apache and a database, the geospatial server can distribute spatial data and feed applications who may display this information through a map on various devices (mobile, web browser, GIS desktop application). The most popular GIS server Open Source using by a large community of developers, enterprise and institutions are:

Geoserver[42] : is a server for sharing geospatial data. It is designed for interoperability and excels at publishing any major spatial data source using open standards. With suitable preparation of data, it excels at handling very large datasets, both raster and vector. It produces high quality rendering of maps and can handle hundreds to thousands of map layers easily. Data is published via standards-based interfaces, such as WMS, WFS, WCS, WPS, Tile Caching and more. Geoserver comes with a browser-based management interface and connects to multiple data sources at the back end.

MapServer[43] : is a platform for publishing spatial data and interactive mapping applications to the web. It is supported by a diverse group of organisations that fund enhancements and maintenance and is administered within OSGeo by the MapServer Project Steering Committee. In its most basic form, MapServer is a CGI program that sits inactive on your Web server. When a request is sent to MapServer, it uses information passed in the request URL and the Mapfile to create an image of the requested map. The request may also return images for legends, scale bars, reference maps, and values passed as CGI variables.

OpenMapTiles[44] : is an extensible and open tile schema based on the OpenStreetMap. This project is used to generate vector tiles for online zoomable maps. OpenMapTiles is about creating a beautiful basemaps with general layers containing topographic information. The final map can be displayed in the browser using most popular JavaScript libraries like Mapbox GL JS, OpenLayers, Leaflet, in mobile applications using Mapbox GL Native (for both Android and iOS), using WMTS and WMS services, or as a static map for print and export. The big advantage of vector tiles is the separation of data and styles. Therefore, it is possible to use the same data with more styles, which can be adjusted on-the-fly.

[42] https://docs.geoserver.org/latest/en/user/
[43] https://mapserver.org/fr/index.html
[44] https://openmaptiles.org/

QGIS Server[45] : provides a web map service (WMS and WFS) using the same libraries as the GIS desktop application (QGIS). Maps and print templates created in QGIS can be published as web maps simply by copying the QGIS project file to the server directory. The resulting web maps look exactly like those in the office tool. QGIS Server usually runs as a CGI/FastCGI module in the Apache web server.

Relevance for MobiDataLab: GIS servers are needed for making data interoperable according to OGC geospatial standards (e.g. WMS, WFS, etc.). However lighter standards like GeoJSON can be used for displaying data on a map. That being said, MobiDataLab could benefit from GIS servers features for easing spatial analysis or for processing data in specific environmental use cases like Emission Reporting and/or Sustainable transport. In these cases, QGIS and/or GeoServer will be used.

4.3.2.2. Proprietary GIS servers

ArcGIS Enterprise: is a full-featured mapping and analytics platform that includes a powerful GIS web services server plus dedicated Web GIS infrastructure. It corresponds on a list of software who can be installed on classic servers (Windows or Linux) or more and more on the Cloud.

ArcGIS Online: is an online, collaborative Web GIS that allows using, creating, and sharing maps, scenes, apps, layers, analytics, and data. At the difference of ArcGIS Enterprise, the infrastructure (mentioned for ArcGIS Enterprise) is hosted by Esri server Cloud and the GIS services are accessible in SaaS mode through web browsers, mobile devices, and desktop map viewers.

Relevance for MobiDataLab: As the integration of GIS servers in MobiDataLab will be optional (see above), we will favour open source solutions, and therefore proprietary solutions will probably not be used.

4.3.3. Geodata visualisation (web mapping)

The representation of mobility data on a map is essential for locating, visualising and querying information. Several open source mapping libraries are available:

OpenLayers: a completely free, open source JavaScript library released under the BSD 2-clause license (also known as FreeBSD). OpenLayers makes it easy to set up a dynamic map in any web page. It can display map tiles, vector data and markers loaded from any source. OpenLayers was developed to support the use of geographic information of all kinds.

Leaflet: a JavaScript library for browser-based interactive mobile maps. It is lightweight, but has all the features that most developers will need for online maps. Leaflet has been designed with simplicity, performance and usability in mind. It works efficiently on all major desktop and mobile

[45] https://www.qgis.org/

platforms, and can be extended with many plugins, has a nice, easy to use and well documented API and simple, readable source code that is a pleasure to contribute.

Mapbox GL: a JavaScript library that uses WebGL to render interactive maps from vector tiles and Mapbox styles. It is part of the Mapbox GL ecosystem, which includes Mapbox Mobile, a compatible renderer written in C++ with bindings for desktop and mobile platforms, a compatible renderer written in C++ with bindings for desktop and mobile platforms.

Related standards: Raster tiles, Vector tiles

Relevance for MobiDataLab: Web mapping libraries are necessary to display geo-referenced data on a map. Mobility data very often contain such a geographical dimension, and therefore these libraries are very relevant for the MobiDataLab user interface.

4.4. Solutions for the semantic enrichment of mobility data

4.4.1. Linked Open Data

Another method of combining datasets is to use Linked Open Data principles and standards. Linked Data is a way to create a network of standards-based, machine-readable data across Web sites. It allows an application to start at one piece of Linked Data, and follow embedded links to other pieces of Linked Data that are hosted on different sites across the Web. The Linked Open Data Cloud (www.lod-cloud.net) lists datasets that have been published in the Linked Open Data format.

Related standards: RDF, SPARQL

4.4.1.1. Generic Linked Open Datasets

Wikidata: Wikidata is an open and structured knowledge repository which serves as a central storage repository maintained by the Wikimedia Foundation. For this reason, data can be stored only once in Wikidata but are then reusable across the entire Wikimedia universe. Also, whenever the content data on Wikidata is updated, the linked wikis are updated automatically.

The key component of Wikidata is the items. Each item is assigned a unique identifier, a label, a description and aliases. All the data concerning an item are semantically expressed by means of properties and values, and are edited and maintained in the form of statements on the item page. For example the item “Berlin-Kaulsdorf railway station” has the property “instance of” and the value “S-Bahn station”, which, as an item itself, is a subclass of “railway station”, which is a “subclass of” “station” which, in turn, is a subclass of “transport stop”. As each property is also assigned an identifier, a label, a description and aliases that can be added in multiple languages and has its own page with statements, it allows items to be linked to others and thus all the data on Wikidata can be interconnected.

Wikidata offers a Query Service to query the data in SPARQL[46] .

DBPedia: DBpedia.org is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia and to link other datasets on the Web to Wikipedia data. The English version of the DBpedia knowledge base currently describes 6.0M entities.

GeoNames Semantic Web: The GeoNames Ontology makes it possible to add geospatial semantic information corresponding to 6.2 million geonames toponyms which now have a unique URL with a corresponding RDF web service. Other services describe the relation between toponyms. SPARQL endpoints for this dataset are made available by third parties (e.g. FactForge)

Relevance for MobiDataLab: Generic SPARQL endpoints are very common data sources when exploring the Linked Open Data Cloud. As a matter of fact, they are very relevant for the MobiDataLab processors aiming to enrich mobility data through semantic web technologies. For example, Wikidata is a structured information graph of about 100 million items.

4.4.1.2. Transport-related Linked Open Datasets

Semantic web standards make it possible to build a Web of transportation data (exactly like there is a World Wide Web of pages). This means that copying data from one system to the other and converting them from one format to the other is not needed anymore.

The Shift2Rail interoperability framework[47] (IF) aims to facilitate multimodal travel in a highly diverse environment comprising many transport modes. It is a set of web services aiming at masking the complexity of interoperability in the transport sector.

[46] https://query.wikidata.org/sparql
[47] https://shift2rail.org/research-development/ip4/

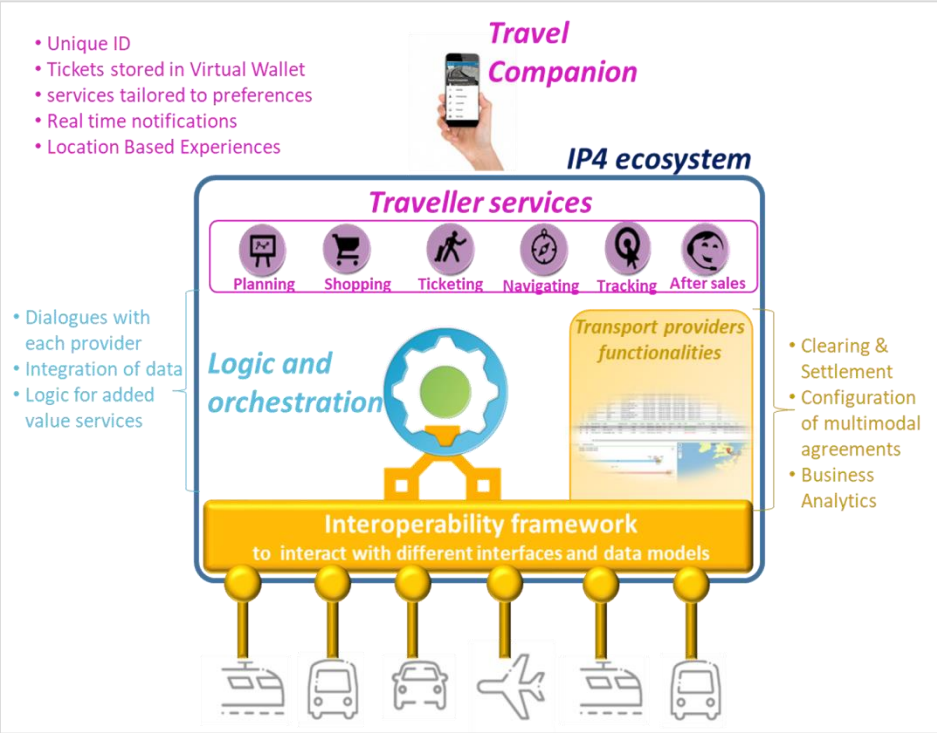

Figure 1 : The Interoperability Framework to facilitate multimodal travel (Source: Shift2Rail)

The semantic part of the interoperability framework defines formal models of the transportation domain to be exchanged automatically by computers (ontologies), automate the generation of ontology-based mappings and annotations, and enables the conversion of heterogeneous messages exchanged by different systems. In this context, the SPRINT[48] project (Semantics for PerfoRmant and scalable INteroperability of multimodal Transport) aimed at defining techniques facilitating activities that are central to the concept of the Shift2Rail Interoperability Framework, such as the collaborative creation and management of ontologies and of semantic-based mappings between heterogeneous data representations.

Relevance for MobiDataLab: Transport-specific SPARQL endpoints are very relevant for the MobiDataLab semantic enrichment processors. Even though the number of records is less important than in generic endpoints, it is important to reference them in the platform – at least for research purposes. Also, it is important that MobiDataLab contributes to these specific ontologies and vocabularies.

4.4.2. Semantic repositories (graph databases)

A semantic repository is a database management system (DBMS) which allows storing, querying, and managing structured data. Semantic repositories are often referred to by synonyms such as

[48] http://www.sprint-transport.eu/

semantic graph database, reasoner, ontology server, semantic store, metadata store, RDF database, RDF triplestore and more. These different wordings emphasise the particular usages rather than a difference in the implementation and performance.

When talking about graph databases we refer to a type of NoSQL database that uses graph theory to represent and store data. It also uses graph structures to semantically query relationships between data. In the graph theory, graph is referred to as a set of vertices or nodes and of edges that connect the vertices. Different features of the graph theory are transposed in the graph database. For example, nodes (vertices) are the data entities in the graph and have several attributes called properties; relationships (edges) are the semantically relevant connections between two node entities (a start node and an end node) and have properties as well.

The major benefit of semantic repositories, compared to traditional databases such as relational databases (RDBMSs), is the usage of semantic data schema paradigm (or ontology) which is stored and managed independently from the data. In graph databases, relationships between data are therefore at least as important as the data itself. The nodes and relationships in a graph database are designed to be scalable and offer flexibility and reliability for real-time data. A graph database allows to automatically discover new facts and build new data based on semantic rules (data inference or reasoning), and to seamlessly integrate data from distributed datasets and data sources (data federation).

The most popular graph databases are GraphDB, Virtuoso and Neo4j

Related standards: OWL, RDF

Relevance for MobiDataLab: Graph databases are the underlying technologies behind Linked Open datasets. It is expected that in the course of the project we will go beyond mere data reuse and open a MobiDataLab SPARQL endpoint to be used by Living and Virtual Labs participants. In this case we will use these technologies.

5. Conclusions

In this deliverable we discussed the results of Task 2.4 aimed at performing an analysis and evaluation of the available solutions (e.g., frameworks, tools, environments, infrastructures, etc.) that could be leveraged in MobiDataLab to enable transport data sharing in the Cloud. Such analysis took into account the key features of existing approaches and provided the relevant state-of-the-art solutions from different perspectives, encompassing both technical and non-technical viewpoints. We listed a number of non-technical dimensions related to the data governance and technical dimensions related to the cloud federation and the data aspects. We have presented the solutions for data sharing and their relevant dimensions, including European projects and initiatives, mobility data platforms and technological federated cloud solutions. On the most promising solutions we performed a SWOT analysis to highlight their strengths and weaknesses as well as the opportunities and threats that could derive from their exploitation in MobiDataLab.

The aim of this report is to identify the open frameworks and solutions that can provide effective and efficient data-sharing services to MobiDataLab stakeholders and facilitate their collaboration by supporting secure methods to selectively share and access the data subject to any specific combination of privacy/visibility rules and conditions.

In the last part of the deliverable, for completeness in relation to the Transport Cloud components, we also presented possible solutions for the data processors (cataloguing mobility data, geospatial enrichment of the data, semantic enrichment of the data). However, since these solutions are not specifically related to the cloud federation, they do not take part on the SWOT analysis.

Funded by the European Union

6. References

Robert B. Bohn, Craig A. Lee, Martial Michel (2020). The NIST Cloud Federation Reference Architecture. NIST Special Publication pp. 500-332, https://doi.org/10.6028/NIST.SP.500-332

Viti, F., Hoogendoorn, S.P., Immers, L.H., Tampère, C.M. and Lanser, S.H., (2008). National data warehouse: how the Netherlands is creating a reliable, widespread, accessible data bank for traffic information, monitoring, and road network control. Transportation research record, 2049(1), pp.176-185.

Labrador, H.G., Alexandropoulos, G., Bocchi, E., Castro, D., Chan, B., Contescu, C., Lamanna, M., Presti, G.L., Mascetti, L., Moscicki, J. and Musset, P. (2019). CERNBox: the CERN cloud storage hub. In EPJ Web of Conferences (Vol. 214, p. 04038). EDP Sciences.

Mościcki, Jakub T., and Luca Mascetti (2018). Cloud storage services for file synchronisation and sharing in science, education, and research. Future Generation Computer Systems, Volume 78, Part 3,
Pages 1052-1054

Mason, W.G., Aben, G., Meijer, J.J., Richter, C., Bonnington, C.P., Moresi, L. and Betts, P.G., (2010), December. Facilitating Research Collaboration in the Australian Geoscience Community Using CloudStor. In 2010 IEEE Sixth International Conference on e-Science (pp. 106-112). IEEE.

Vogl, R., Rudolph, D., Angenent, H., Thoring, A., Schild, C., Stieglitz, S. and Meske, C., (2015). sciebo–the Campuscloud: a Sync and Share Cloud Storage Service for the Academic and Research Community in North Rhine-Westphalia. PIK-Praxis der Informationsverarbeitung und Kommunikation, 38(3-4), pp.105-112.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), pp.1-9.

MobiDataLab consortium

The consortium of MobiDataLab consists of 10 partners with multidisciplinary and complementary competencies. This includes leading universities, networks and industry sector specialists.

@MobiDataLab
#MobiDataLab

https://www.linkedin.com/company/mobidatalab

For further information please visit **www.mobidatalab.eu**

MobiDataLab is co-funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).