MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D4.1 Transport Cloud Architecture Dossier V1

10/02/2023

Author(s): Salvatore TRANI (CNR), Alberto BLANCO JUSTICIA (URV), Thierry CHEVALLIER (AKKA), Didier de RYCK (HOVE), Johannes LAUER (HERE), Francesco LETTICH (CNR), Sorel SIGHOKO (AKKA)

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

Summary sheet

Deliverable Number	D4.1
Deliverable Name	Transport Cloud Architecture dossier V1
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Salvatore TRANI (CNR)
Contributing Partner(s)	Alberto BLANCO JUSTICIA (URV) Thierry CHEVALLIER (AKKA) Didier de RYCK (HOVE) Johannes LAUER (HERE) Francesco LETTICH (CNR) Sorel SIGHOKO (AKKA)
Peer Review	ICOOR / AETHON / AKKA
Contractual Delivery Date	31-01-2022
Actual Delivery Date	28-01-2022
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	39
WP/Task related to the deliverable	WP4 / T4.1
WP/Task responsible	AKKA / CNR
Document ID	MobiDataLab-D4.1-TransportCloudArchitectureDossierV1
Abstract	This deliverable is a report aimed at providing an overview of the Task 4.1, which consists of the conception, definition and design of the cloud solution supporting the data federation to enable the European-wide data sharing and trans-national access to the information provided by the federated repositories.

Funded by the European Union

The Deliverable D4.1 is the first submitted iteration of a living document that will describe the components and architecture of the MobiDataLab Transport Cloud. Within the project time, it is expected that, as the use cases are refined and executed, the overall specification and the corresponding architecture will evolve towards the new requirements that will arise. This document will be reviewed accordingly, and learnings will flow into the second version of the architecture, which will be part of D4.2.

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

Document history

Version	Date	Organisation	Main area of changes	Comments
0.6	17/12/2021	CNR, AKKA, URV, HOVE, HERE	all	Collected contributions from the partners
0.7	04/01/2022	CNR	all	Document consolidated
0.8	17/01/2022	AETHON, ICOOR, AKKA	all	Review
0.9	23/01/2022	CNR	all	Rework including internal review comments
1.0	28/01/2022	AKKA	all	Quality check and submission

Executive Summary

Deliverable D4.1 is the first submitted iteration of a living document that will describe the components and architecture of the MobiDataLab Transport Cloud, providing design principles and technical guidelines that allow the platform's functionalities to be aligned with the requirements of the stakeholders represented in the project (i.e., authorities, operators, MaaS companies and developers).

The current version of the architecture specification is based on the preliminary analysis of the use cases provided in the deliverable "D2.9: Use cases definition (v1)" as well as on the "D2.6 : Report on enabling technologies for Transport Cloud", covering the basic functional and non-functional expectations from the platform in order for the report to conform to the requirements of those use cases.

It is expected that, as the use cases are refined and clarified, the overall specification and the corresponding architecture will evolve towards the new requirements that will arise. To this end, the specifications document is treated as a living document, with regular updates concerning significant changes in design and functionality.

Funded by the European Union

Table of contents

Funded by the European Union

List of figures

List of tables

Abbreviations and acronyms

Abbreviation	Meaning
ACID	Atomicity, Consistency, Isolation, Durability
API	Application Programming Interface
CAS	Central Authentication Service
CKAN	Comprehensive Knowledge Archive Network
CLI	Command Line Interface
CSV	Comma-Separated Values
CSW	Catalogue Service for the Web
DCAT	Data Catalog Vocabulary
DOS	Denial of Service
ETA	Estimated Time of Arrival
GIS	Geographic Information System

Funded by the European Union

GUI	Graphical User Interface
HDFS	Hadoop Distributed File System
HTTP	Hypertext Transfer Protocol
IAAS	Infrastructure as a Service
IAM	Identity and Access Management
INSPIRE	Infrastructure for Spatial Information in the European Community
JS	JavaScript
JSON	JavaScript Object Notation
LDAP	Lightweight Directory Access Protocol
MaaS	Mobility as a Service
OAuth	Open Authorisation
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
PDF	Portable Document Format
POI	Point of Interest
RDF	Resource Description Framework
REST	Representational State Transfer
RPM	Red Hat Package Manager
SAML	Security Assertion Markup Language
SCIM	System for Cross-domain Identity Management
SPARQL	SPARQL Protocol and RDF Query Language
SPML	Service Provisioning Markup language
SSO	Single Sign-On
UMA	User-Managed Access

XACML	eXtensible Access Control Markup Language
XML	eXtensible Markup Language

Funded by the
European Union

1. Introduction

The purpose of this document is to define and design the MobiDataLab Transport Cloud, a cloud-based prototype platform for sharing transport data using a federated approach. It will balance long-term strategic issues with immediate requirements of the project's current phase of technical development.

This first version of this deliverable, which aims to define the MobiDataLab Transport Cloud architecture, is based (1) on the preliminary analysis of the use cases and datasets identified in the deliverable D2.9 and (2) on the detailed analysis and evaluation of the relevant dimensions and state-of-the-art technologies and solutions (see deliverable D2.6) that can be potentially leveraged to realise the Transport Cloud. Indeed, these two deliverables provide the foundations for the definition of the functional and non-functional requirements targeted in this document.

While technical details are not thoroughly analysed at the present stage, the architecture outlined in this first iteration of the deliverable clearly identifies candidate technologies, tools, and frameworks to be used for the development of the Transport Cloud. The selection is based on the proven performance and wide adoption of the technologies, ensuring transnational access in a secure, effective and seamless way.

The report is organised as follows: Section 2 presents the main user roles to be supported by the platform along with their main operations within the platform. Section 3 describes the non-functional requirements from the software stack, while Section 4 deals with the functional requirements. Section 5 analyses the main information flows that will be executed over the platform, proceeds to the presentation of the architectural design of MobiDataLab, placing the components participating in the flows to distinct, clearly defined modules and concludes with a summary of the technologies to be used for implementing the core components of the software stack. Finally, Section 6 presents the candidate technologies to be exploited for the implementation of the MobiDataLab components.

2. MobiDataLab Platform Overview and Main Actors

MobiDataLab targets technology challenges aimed at proposing to mobility stakeholders a replicable methodology and sustainable tools that foster the development of a data sharing culture in Europe and beyond. Facing such challenges requires the use of techniques that deal with data fusion, privacy and anonymisation, geographical and semantic enrichment, harmonisation/standardisation, and data processing in general. These will be indeed required to improve the quality, accessibility and usability of mobility data and to allow each mobility data provider to share securely and safely their data.

The Transport Cloud will demonstrate a cloud-based prototype platform for sharing transport data, accessible to interested mobility actors. It is technically designed according to federated cloud principles. The MobiDataLab platform will showcase how to facilitate access to mobility data, in an open, interoperable, and privacy preserving way, developing open tools and making them available. The Transport Cloud is primarily designed to demonstrate and offer solutions to reduce and, in some cases, remove current technical limitations identified as barriers to data sharing and reuse.

At this stage, we identify four (4) main actors: Administrator, Developer, Data Consumer, and Data/Service Provider. The core activities for each actor are summarised as follows:

- **Administrator**: User management; Access management; Applications and Components configuration;
- **Developer**: Transport Cloud component deployment integration;
- **Data Consumer**: any entity that is interested to use the data and services available within the Transport Cloud. Some examples that can be observed or inferred from the uses cases in the deliverable 2.9 are:
    - Data scientist, researcher, domain expert;
    - Transport customers;
    - Services external to the Transport Cloud and that use its data and services.
- **Data/Service Provider**: any entity that provides, either passively or actively, data or services to the Transport Cloud. Some examples that can be observed or inferred from the uses cases in the deliverable 2.9 are:
    - Trip planners (e.g., Navitia, HERE);
    - MobiDataLab stakeholders, e.g., transport operators or public institutions that actively share their data and services for the good of the MobiDataLab project;
    - Open data/services providers (e.g., OpenStreetMap).

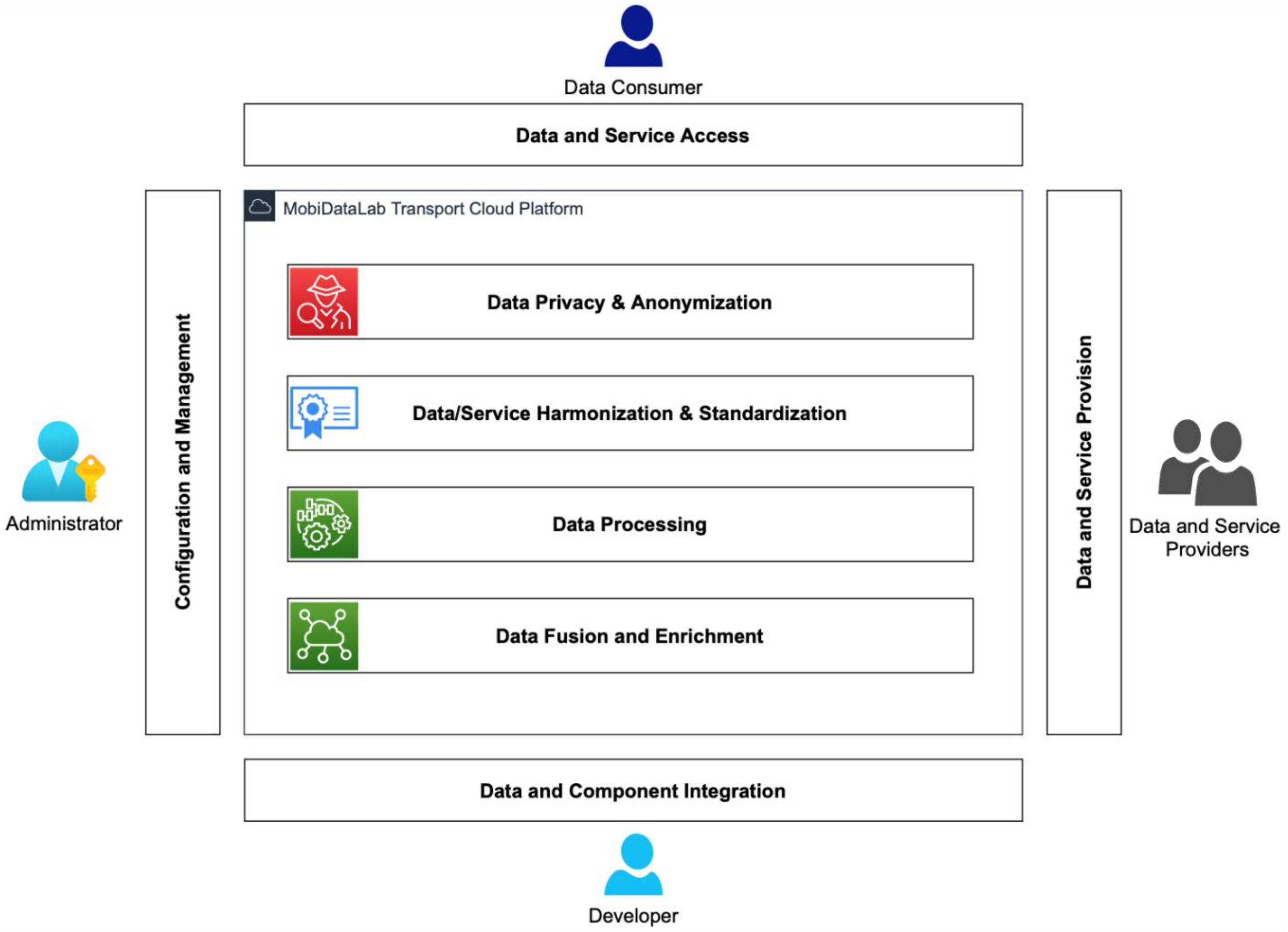
Data Consumer
Data and Service Access
MobiDataLab Transport Cloud Platform
Data Privacy & Anonymization
Data/Service Harmonization & Standardization
Data Processing
Data Fusion and Enrichment
Configuration and Management
Administrator
Data and Service Provision
Data and Service Providers
Data and Component Integration
Developer

Figure 1: Logical Organisation of the MobiDataLab Transport Cloud platform

Figure 1 illustrates the logical organisation of the Transport Cloud platform. The platform comprises components that implement the set of functionalities required by the use cases and that will be presented in more detail in Section 4.

The components depicted in the figure will be exposed via clearly defined and thoroughly documented APIs and deployed on, as well as integrated in, the Transport Cloud platform, which in turn will serve the MobiDataLab's use cases.

Funded by the
European Union

3. Non-Functional requirements

This section summarises the non-functional requirements for the MobiDataLab Transport Cloud platform.

3.1. Governance and Regulation aspects

The storage and processing of data by the MobiDataLab Transport Cloud must satisfy strict key requirements concerning the privacy and trust of European citizens and businesses.

The non-functional requirements presented below address these needs.

NFR-GOV.01	**Data Sovereignty**
Description	Data stored and processed by the MobiDataLab Transport Cloud shall be kept under the European Union authority, thus ensuring data sovereignty.
Objectives	• Data stored and processed by the MobiDataLab Transport Cloud shall be located within premises or datacentres falling in the European Union territory.

NFR-GOV.02	**Data Anonymisation**
Description	The MobiDataLab Transport Cloud must ensure that personal or sensitive data accessed or processed by the platform is not made available to audiences different than those for which the data owners gave their consent.
Objectives	• The MobiDataLab Transport Cloud must implement anonymisation mechanisms that are employed when the platform needs to access or process data to execute some task requiring some form of anonymisation. Examples of types of data that are relevant to the MobiDataLab project are GPS locations, personal information such as e-mail, names, addresses, and so on; • Build confidence with end users over privacy; • Fulfil existing privacy regulations.

Funded by the European Union

3.2. Cloud Federation Aspects

The MobiDataLab Transport Cloud will strive to represent a concrete example of viable cloud federation by satisfying the non-functional requirements introduced below.

NFR-CFA.01	**Transport Cloud agnosticism**
Description	A major requirement which needs to be satisfied to validate the MobiDataLab project calls on the Transport Cloud to be agnostic with respect to any suitable cloud provider that may be used to support its implementation.
Objectives	• The MobiDataLab transport cloud's design must ensure that the Transport Cloud does not require nor depend on any specific cloud editor. As a consequence, the Transport Cloud can be deployed on any suitable cloud technology vendor infrastructure.

NFR-CFA.02	**Use of open source, standard technologies**
Description	Serialisation and standardisation are key tenets that shall underpin the MobiDataLab transport cloud's design. These have clear benefits in terms of costs, wide availability, and open documentation, thus paving the way to technology and knowledge transfer.
Objectives	• Select and leverage open source well-established tools, technologies, and processes that align with the aforementioned tenets; • Ensure compatibility with the external world.

NFR-CFA.03	**Cloud implementation strategy**
Description	Cloud providers differ from one another in the number and types of services they provide, as well as in the way said services are set up within each cloud platform. The MobiDataLab project aims to select a set of services common to all main cloud providers. Such selection shall comprise standard information technology tools required to build the Transport Cloud infrastructure, regardless of whether the infrastructure will be hosted on premise, virtualised on some cloud platform, or some hybrid between these two approaches.
Objectives	• Determine the set of services that are common among the major cloud providers and that can be used to implement the functionalities the Transport Cloud shall provide; • Guarantee that the Transport Cloud's functionalities can be implemented regardless of the approach chosen to implement the underlying infrastructure.

3.3. Data Management Aspects

NFR-DM.01	Data distribution
Description	Data distribution is a fundamental principle upon which the MobiDataLab project is built. The MobiDataLab project intends to offer a single data entry point for a multitude of variegated data sources, either public or privately owned.
Objectives	• Design the MobiDataLab Transport Cloud architecture so that it enables the vision highlighted in the description above.

NFR-DM.02	Data sources identification
Description	The MobiDataLab project has to identify the data sources providing the data needed by the Transport Cloud to satisfy the use cases presented in the deliverable D2.9.
Objectives	• Identify suitable data sources: open datasets to be imported, metadata to be imported, creating a data catalogue within the Transport Cloud; • Create rich content from multiple data sources.

NFR-DM.03	Data ownership
Description	Using privately owned data, and possibly combining it with public open data, poses new challenges to the MobiDataLab Transport Cloud in terms of data ownership that need to be settled among private data owners, the MobiDataLab project stakeholders, and data consumers willing to use the Transport Cloud for the data and services it provides.
Objectives	• Fulfil existing intellectual property regulations, striving to strike a good balance between the need to preserve intellectual property and the need to serve public institutions and the general public at large.

Funded by the European Union

4. Functional Requirements

In this section we introduce the main functional requirements the MobiDataLab Transport Cloud is supposed to satisfy to cover the needs of the use cases introduced in the deliverable D2.9.

4.1. Use Case A: Optimisation of Transport flow and ETA

FR-A.01	Transport flow monitoring
Description	In order to optimise, monitor, and manage commercial transport flow, it is crucial to have updates (either periodical or in real-time) from various data sources -- e.g., fleet status, weather, traffic, planned events, etc. -- to provide an overall picture of the commercial transport system and to trigger specific actions whenever needed, i.e., delays, arrival time sharing, and tour plan update.
Objectives	• The Transport Cloud must provide a **monitor processor** which interacts with the fleet IT system, from which the monitor receives a stream of information; • the ETA estimator, a service used by the Transport Cloud to update/optimise tour planning; • Other data sources such as weather, traffic information, planned events, and so on, which may further help to improve service quality provided by the fleet owner; • The Transport Cloud must provide **notification/planning functionalities** within the monitor that allow the Transport Cloud to achieve the goals specified in the description of this requirement.

FR-A.02	Tour reporting and planning
Description	Planned tours have to follow rest and break time regulations. The system should therefore be able to take these aspects into account when planning tours. Moreover, explicit reporting of already completed tours can be requested by the fleet owner to compare planned, initially estimated, and actual arrival times, thus allowing to identify potential problems. Differently from the activities described in requirement FR-A.01, the activities in this requirement are requested on demand and take advantage of information gathered by the monitor.
Objectives	• The Transport Cloud must provide a **REST-API component** allowing an external user (e.g., the fleet owner) to request on demand specific services to the Transport Cloud and receiving back the result;

Funded by the European Union

	• The Transport Cloud must provide an **advanced dispatcher processor** that uses an ETA estimator and take into account rest and break time regulations, to plan tours that are compliant with existing EU legislation.

FR-A.03	**Data distribution**
Description	Data distribution is a fundamental principle upon which the MobiDataLab project is built. The MobiDataLab project intends to offer a single data entry point for a multitude of variegated data sources, either public or privately owned.
Objectives	• Design the MobiDataLab Transport Cloud architecture so that it enables the vision highlighted in the description above.

4.2. Use Case B: Emission Reporting

FR-B.01	**Emission estimation and reporting**
Description	Reducing environmental impact is highly relevant for any form of mobility and transport. Concrete action for reducing the environmental footprint can only be taken in a systematic way if the impact is reported in a clear and transparent manner. Therefore, the need for predicting and reporting emissions is of great significance.
Objectives	• The Transport Cloud must provide an **emission processor** that is able to predict the emission footprint of a planned/already completed tour, based on the tour plan, telematics data collected by the involved vehicles (if available), and information concerning direct and indirect emissions; • The Transport Cloud must provide an exhaustive **emission report processor** that is able to take into account the emission estimation of a trip possibly composed of several tours (like it happens for the transportation of goods where several transport assets are used).

4.3. Use Case C: Analytics & Learning

FR-C.01	**Data Access, Analytics and Learning**
Description	Being able to provide access, visualise and finally analyse data is a crucial component in every data platform. To this end, this use case is providing a horizontal connection to all use cases, since analysis and learning methods can contribute to most of the presented use cases.

Objectives	• The Transport Cloud must provide a **loader component** that must be able to load spatial data and could be able to load other types of non-spatial data (e.g., tabular, etc.) into the Transport Cloud; • The **metadata catalogue** available in the Transport Cloud must provide functionalities that allow to find, browse, and explore datasets that are of interest to the data consumer; • The **data API** must provide access to data sources containing datasets of interest to the data consumer and that do not require any preliminary data pre-processing (e.g., anonymisation) before being accessed; • The **service API** must provide access to data sources containing datasets of interest to the data consumer and that do require some preliminary data pre-processing (via a suitable processor) before being accessed.

4.4. Use Case D: Re-use of transport data for journey planners / digital services

FR-D.01	**Multi-modal Journey Planning**
Description	Multi-modal journey planning capability is a feature of interest for many digital service providers. However, dealing with raw transport datasets, using different formats and combining them could be particularly difficult and complicated. To this end, a common solution for journey planning would therefore foster the use of transport data and simplify its usage.
Objectives	• The Transport Cloud must provide a **loader component** that allows to load data into the Transport Cloud from public transport operators/authorities. The loader should also allow to load data concerning POIs or from other transport modes; • The Transport Cloud must provide a data converter processor that converts the aforementioned data into a standard data format. This processor must also integrate different data sources. Finally, the processor should be able to process and enrich datasets. The processor must finally provide the end result of its activities available to other entities; • The Transport Cloud must provide a **journey planning processor** using integrated datasets. The planner might be able to use third party external calculators. Finally, the planner should be able to return journey statistics and provide journeys through a standardised API.

4.5. Use Case E: Mobility as a Service (MaaS)

FR-E.01	**Mobility As A Service**
Description	Mobility is a continuously evolving service, including a variety of heterogeneous transport solutions (bus, train, car sharing, bikes, etc) provided by different

Funded by the European Union

	providers (public and private). Thus, it is crucial to provide users with an end-to-end solution encompassing journey planning and ticketing. Within the MobiDataLab project, we aim to provide access to raw datasets on the one hand, and on the other hand to provide a multi-modal journey planning service.
Objectives	• The Transport Cloud must provide a **loader component** for loading data into the Transport Cloud from raw transport, journey planning, and MaaS datasets. This functionality shows the need of dedicated storage solutions to be deployed within the Transport Cloud; • The Transport Cloud must provide access to datasets imported from MaaS operators -- this can be achieved by means of the **metadata catalogue** and the **data/service APIs**; • The Transport Cloud should provide a **data transformer processor** that is able to integrate booking and payment data from the MaaS operators, and could integrate journey data as well; • The Transport Cloud must provide a **journey planning processor**, eventually using third party external calculators, implementing the service and providing the response in a standard format through a MaaS paradigm.

4.6. Use Case F: Geodata Sharing applied to Transport: OpenStreetMap for Inclusive transport

FR-F.01	**Dataset discovery**
Description	The data consumer uses the Transport Cloud to find, browse, and explore transport and accessibility datasets that can suit their needs.
Objectives	• The metadata catalogue available in the Transport Cloud must provide functionalities that allow to find, browse, and explore datasets that are of interest to the data consumer; • The **data API** must provide access to data sources containing datasets of interest to the data consumer and that do not require any preliminary data pre-processing (e.g., anonymisation) before being accessed; • The **service API** must provide access to data sources containing datasets of interest to the data consumer and that do require some preliminary data pre-processing (via a suitable processor) before being accessed.

FR-F.02	**Dataset combination and enrichment**
Description	The data consumer wants to use the Transport Cloud to combine and enrich (either geospatially or semantically) the datasets identified via the functionalities described in FR-F.01. The result of these operations consists in one or more enriched and consolidated datasets.

Funded by the European Union

Objectives	- The Transport Cloud must provide a **dataset joiner** whose purpose is to perform the fusion of the datasets that are of interest to the data consumer. This joiner should be able to combine mobility data (e.g. provided by public transport authorities) with OpenStreetMap accessibility data following the OSM common exchange format including "tags". Such a combination should rely on the identification of a common location (so-called "node" in OSM terminology); - The Transport Cloud must provide a **geospatial enrichment processor** whose purpose is to perform the geospatial enrichment of datasets; - The Transport Cloud must provide **storage mechanisms** to store the enriched and consolidated datasets, so that the data consumer can retrieve them later.

FR-F.03	**Data analysis**
Description	The data consumer wants to retrieve and then analyse with their preferred analysis tool the enriched and consolidated dataset(s) produced via the functionalities described in FR-F.02.
Objectives	- The Transport Cloud must provide **storage mechanisms** to store the enriched and consolidated datasets, so that the data consumer can retrieve them later via the **metadata catalogue** and the **data/service APIs**.

4.7. Use Case G: Geodata Sharing applied to Transport: Environmental Data for Sustainable Transport

FR-G.01	**Dataset discovery**
Description	The data consumer uses the Transport Cloud to find, browse, and explore transport and environmental datasets that can suit their needs.
Objectives	- The **metadata catalogue** available in the Transport Cloud must provide functionalities that allow to find, browse, and explore datasets that are of interest to the data consumer; - The **data API** must provide access to data sources containing datasets of interest to the data consumer and that do not require any preliminary data pre-processing (e.g., anonymisation) before being accessed; - The **service API** must provide access to data sources containing datasets of interest to the data consumer and that do require some preliminary data pre-processing (via a suitable processor) before being accessed.

Funded by the European Union

FR-G.02	Dataset combination and enrichment
Description	The data consumer wants to use the Transport Cloud to combine and enrich, either geospatially or semantically, the datasets identified via the functionalities described in FR-G.01. The result of these operations consists in one or more enriched and consolidated datasets.
Objectives	• The Transport Cloud must provide a **dataset joiner** whose purpose is to perform the fusion of the datasets that are of interest to the data consumer. This joiner should be able to combine mobility data (e.g. provided by public transport authorities) with local environmental data following the Geographical Information Systems formats and exchange standards (OGC, INSPIRE, etc). Such a combination should rely on the identification of a common location (so-called "geometry" in GIS terminology); • The Transport Cloud must provide a **geospatial enrichment processor** whose purpose is to perform the geospatial enrichment of datasets; • The Transport Cloud must provide **storage mechanisms** to store the enriched and consolidated datasets, so that the data consumer can retrieve them at a later time.

FR-G.03	Data analysis
Description	The data consumer wants to retrieve and then analyse with their preferred analysis tool the enriched and consolidated dataset(s) produced via the functionalities described in FR-G.02.
Objectives	• The Transport Cloud must provide **storage mechanisms** to store the enriched and consolidated datasets, so that the data consumer can retrieve them at a later time via the **metadata catalogue** and the **data/service APIs**.

4.8. Use Case H: Transport Data Sharing within the Linked Open Data vision

FR-H.01	Dataset provision and discovery
Description	Some of the actors involved in this use case may need to either provide datasets to the transport cloud, or find, browse, and explore datasets or data sources that can be used to enrich or complete their own datasets.
Objectives	• The Transport Cloud must provide a **loader component** that allow data providers to load the datasets they need to provide to the Transport Cloud; • The Transport Cloud must provide **storage mechanisms** that allow data providers to store the datasets they need to provide on the Transport Cloud;

Funded by the
European Union

	• The Transport Cloud must be able to query data gathered in semantic databases via e.g. SPARQL queries.

FR-H.02	Data combination and enrichment
Description	The tourism service provider wants to combine and enrich geospatially or semantically, the datasets identified via the functionalities described in FR-H.01. The result of these operations consists in one or more enriched and consolidated datasets.
Objectives	• The Transport Cloud must provide a **semantic combination processor** whose purpose is to perform the fusion of the RDF datasets that are of interest to the data consumer. This processor should be able to combine data published according to Linked Open data principles. Such a combination should rely on a common vocabulary (so-called ontology); • The Transport Cloud must provide a **semantic enrichment processor** whose purpose is to perform the semantic enrichment of datasets.

FR-H.03	Tourism analytics
Description	The tourism service provider wants to perform tourism analytics on the dataset(s) produced via the functionalities described in FR-H.02. Such analytics must be performed within the Transport Cloud.
Objectives	• The Transport Cloud shall provide a **processor** whose purpose is to perform **tourism analytics** according to specifications provided by the tourism service provider; • If required, the Transport Cloud may need to provide **storage mechanisms** to store the results within the Transport Cloud, thus allowing the tourism service provider to retrieve them later.

4.9. Generic Functional Requirements

FR-GENERIC.01	Storage capability
Description	Actors interacting with the Transport Cloud, or components within the Transport Cloud, may require storing data within the platform to support their activities.
Objectives	• The Transport Cloud must provide **persistent storage mechanisms** to store the enriched and consolidated datasets, so that the data consumer can retrieve them at a later time. Storage should be **scalable** and **fault tolerant** to the data access;

Funded by the European Union

	• These storage mechanisms may vary according to the actors' or components' needs and the type of data being considered. To this end, we refer the reader to Sections 6.1 and 6.2.

FR-GENERIC.02	**Loader capability**
Description	Data consumers and data providers may want or need to actively upload their datasets within the Transport Cloud -- indeed, in some scenarios they may not be able to provide their data through a service they can expose.
Objectives	• The Transport Cloud must provide a **loader component** that allows data consumers and data providers to load the datasets they want or need to provide to the Transport Cloud.

FR-GENERIC.03	**Metadata catalogue**
Description	Actors interacting with the Transport Cloud may want to find, browse, and explore datasets. Such datasets can be available either within the Transport Cloud or from third party data providers.
Objectives	• The **metadata catalogue** available in the Transport Cloud must provide functionalities that allow to find, browse, and explore datasets that are of interest to the data consumer.

FR-GENERIC.04	**Service catalogue**
Description	Actors interacting with the Transport Cloud may want to find, browse, and explore the services provided by the platform or by third party service providers associated with the Transport Cloud.
Objectives	• The Transport Cloud must provide a **service catalogue** component.

FR-GENERIC.05	**Basic Data Access API**
Description	Actors interacting with the Transport Cloud want to retrieve data from it, or from third party data providers that provide data to the Transport Cloud.
Objectives	• The **data API** must provide access to datasets that are of interest to the actors, either within the Transport Cloud or available from third party data sources, that **do not require any preliminary data processing** (e.g., anonymisation, enrichment, etc.) before being accessed.

FR-GENERIC.06	Advanced Data and Service Access API
Description	Actors interacting with the Transport Cloud want to retrieve data, either from the Transport Cloud or from third party data providers that provide data to it, with the additional requirement that such data require appropriate processing (e.g., enrichment, anonymisation, etc.) before being accessed.
Objectives	• The **service API** must provide access to data sources containing datasets of interest to the data consumer and that **do require some preliminary data processing** (via a suitable processor) before being accessed.

FR-GENERIC.07	Converter API service
Description	Combining several third-party mobility data sources and services is among the main goals of the MobiDataLab platform. However, data and services provided by third parties are likely to follow different standards or conventions. The Transport Cloud must therefore provide a specific converter component aimed at standardising and reconciling different representation formats.
Objectives	• The data and service APIs must provide access to data represented with a homogeneous and standard data format. To this end the platform shall rely on a **converter component**, thus ensuring the data will be interoperable and standardised.

FR-GENERIC.08	Identity management
Description	Actors who want to interact with the Transport Cloud must have the appropriate credentials to access and interact with the platform.
Objectives	• The Transport Cloud must provide an **identity manager** component being able to authenticate the actors.

Funded by the European Union

5. Transport Cloud Architecture

Based on the necessary requirements mentioned in the previous sections (3 & 4), a more detailed Transport Cloud architecture is depicted in Figure 2. The Figure provides a graphical representation of the main components and the connections between them.

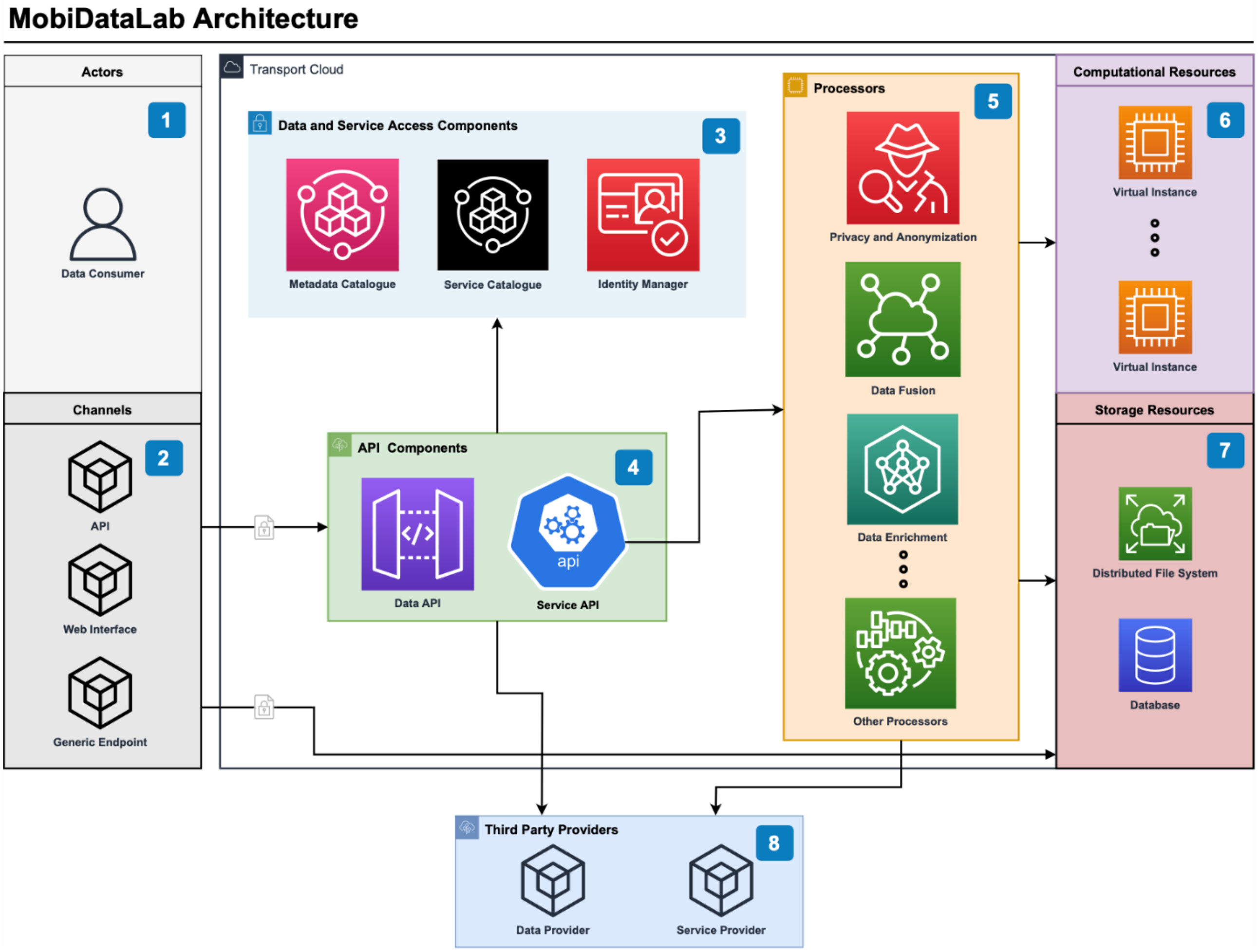

Figure 2: MobiDataLab Transport Cloud architectural design

The actors interacting with the platform that were introduced in Section 2 are present in Figure 2 under the general umbrella of “Data Consumer” (box 1) and “Third-party Providers” (box 8).

The Data Consumer can interact with the platform through several Transport Cloud channels (see also box 2 in Figure 2), i.e., (1) **API endpoints** (mainly dedicated to REST API services), (2) **web interface endpoints**, i.e., dedicated to services which need interaction with the end user (for example scenarios involving data analysis and visualisation tasks), and (3) **generic endpoints** - for instance, a SPARQL endpoint may enable a Data Consumer to access some knowledge base via RDF queries.

The various internal components associated with **Data and Service Access** (box 3), **APIs** (box 4), and **Data Processors** (box 5) lean on the usage of **Computational** (box 6) and **Storage** (box 7) **Resources**. By computational resources we refer to resources providing computation capabilities, i.e., virtual instances provided by the chosen cloud provider and intended to support the execution of the services deployed and running within the Transport Cloud platform. By storage resources, on the other hand, we refer to resources that enable the storage of information within the Transport Cloud such as distributed file system solutions and generic database solutions (e.g., PostGreSQL with PostGIS extension, SPARQL engines, and so on).

The following subsections will introduce the purpose and scope of each component of the architecture. More specifically, Section 5.2.1 introduces the Data Service and Access Components (box 3 in Figure 2). Section 5.2.2 introduces the API Components (box 4 in Figure 2). Finally, Section 5.2.3 introduces the notion of processors (box 5 in Figure 2), along with some specific processor specialisations that are fundamental to the goals of the MobiDataLab project.

5.1. Information Flow

The actors interacting with the Transport Cloud (previously introduced in section 2) will exchange information in ways defined by specific workflows, as derived by the general use case requirements.

The platform’s information entry point is always represented by the third-party providers, who are responsible for the provision of information in the form of datasets and services. The access mechanism to these information sources should be identified and implemented according to the operations to be performed and on the types of data that need to be accessed.

Information retrieved from third-party providers can either be imported within the Transport Cloud, thus requiring appropriate storage solutions (e.g., relational databases, spatial databases, knowledge graph databases), or directly accessed by the Transport Cloud through the use of specific data and service endpoints being exposed by the providers. We report that the latter type of access is conducted by the Transport Cloud by means of the data and service APIs components. The Transport Cloud may also employ proper caching mechanisms to improve the efficiency of such type of accesses.

Data accessed through the Transport Cloud may need some form of processing – a few relevant examples are fusion, enrichment, anonymisation, and format translation and standardisation. To this end, specific processors are identified and implemented, providing such functionalities to the Transport Cloud.

The Transport Cloud finally provides access to the several data sources and services the data consumers may need by exposing several types of channels implemented through specific API endpoints. More precisely, the data consumer interacts with the Transport Cloud first by authenticating themselves via the Identity manager.

Once authenticated the data consumer can then proceed to submit their requests to the Transport Cloud, which then processes them by querying the metadata and service catalogues to find the appropriate data sources and services to satisfy the data consumer's needs.

5.2. Architectural Components

In this section we introduce the main architectural components onto which the Transport Cloud will be built. These components are essential to satisfy or realise the functional and non-functional requirements presented in Sections 3 and 4.

5.2.1. Data and Service Access Components

This section presents the components (box 3 in Figure 2) that deal with identity management, and the discovery, exploration, and research of data and services available within the Transport Cloud.

Serving mobility data faces challenges that could be described as follows:

- Data volume: Consumption of mobility data (e.g. through mobile devices) incurs permanent activity between the application and the information system and that accumulation brings to a high volume of queries to be processed;
- Data accuracy: real time activity such as availability (bikes, scooters, car park, etc.) must be trustworthy for the user to be sure to find on site the promoted asset when necessary;
- Data ownership: mobility data is a data stack valorisation layer on top of layers from different sources whose access may not be totally free of charge or free to be publicly spread to anyone.

5.2.1.1. Metadata Catalogue

A metadata catalogue listing the content made available by the MobiDataLab project as datasets or metadata will be presented to the end-user. This catalogue will be presented regardless of data volumes and ownership and MobiDataLab client applications will take advantage of the data catalogue to bring to the user the expected content.

The metadata catalogue allows to provide access to the mobility datasets via a web portal, and thus not only via an API. It will build on widespread solutions in the transport sector and used by several stakeholders (see Section 6 on candidate technologies). Management of spatially referenced resources (i.e. geospatial metadata) may be needed, as mobility data is often location-based. This support may be added or plugged to the standard catalogue via APIs or extensions.

Funded by the European Union

5.2.1.2. Data Access and Services, Service Catalogue

Access to data and services is a critical part of the whole infrastructure of the Transport Cloud. For the data access, many data providers already build their data offers using standard formats and interfaces to make their data available. Within the context of INSPIRE, a European initiative aimed to create a spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment, especially OGC standardised interfaces for data and services are being used.

However, there also exist proprietary data formats, for reasons of feature richness, interaction with internal workflows or for historical reasons. Examples for that are data provided by platforms such as the HERE platform (https://platform.here.com/, https://developer.here.com/documentation) or ESRI ArcGIS hub (https://hub.arcgis.com/). Further examples for data and services are provided in D2.6 Report on enabling technologies for Transport Cloud.

The services ecosystem is even more heterogeneous, since the services appear later in the processing chain to the end-users, there was less motivation to provide standardised access to them. However, in a world, that is moving towards platform and service-oriented architectures, there is further need to make the services findable, accessible, interoperable and with this reusable for many use cases. With the availability of interface descriptions (e.g., OpenAPI), services are becoming more transparent and a potential to provide metadata for them has grown. Nevertheless, the standardisation of the interfaces needs to be addressed within this part to show the possibilities of interoperability and potentials for adjustments and evolution of the services.

These challenges will be analysed and addressed to extend availability and interoperability of data and services.

5.2.1.3. Identity Management

The MobiDataLab Transport Cloud will manage Identity and Access Management (IAM) which includes identity provisioning, authentication, authorisation and, possibly, Single Sign-On (SSO). In the context of mobility data sharing, these mechanisms can be favourably integrated into an API gateway solution (see Section 6 on candidate technologies).

For clarification purposes, these different concepts are defined below:

- Identity Provisioning is the process of creating, updating, retrieving and deleting user's digital identities. Digital identity is carrying information such as username, password, email, address, phone number, roles, groups, etc. about registered users;
- Authentication is the process of identifying an individual, entity or website based, for example, on a username and password, a client id and client secret or biometric data (fingerprint, iris scan, face recognition, etc.). In security systems, authentication is distinct from authorisation, which is the process of giving individuals access to system objects based on their identity. Authentication merely ensures that the individual is who he or she claims to be, but says nothing about the access rights of the individual;

- Authorisation is the process of granting or denying access to a resource. Most computer security systems are based on a two-step process. The first stage is authentication, which ensures that a user is who he or she claims to be. The second stage is authorisation, which allows the user access to various resources based on the user's identity;
- SSO (Single Sign-On) is an authentication process in a client/server relationship where the user, or client, can authenticate once (with only one user/password pair for example) and have access to more than one application or access to a number of resources within an enterprise or the whole web. Single Sign-On takes away the need for the user to enter further authentications when switching from one application to another.

5.2.2. API Components

The API components consists of two parts: Data API and Service API. With the catalogues described in Sections 5.2.1.1 and 5.2.1.2, there is a list of available data and services including their metadata. To access the services, a direct connection is needed. The API components therefore take the metadata from the catalogues including the references connecting to the services and accessing the data.

The first step will be to make available data and services following a standardised and open interfaces approach. In a second step, the challenges of integrating non-standardised interfaces and data will be analysed. The different options for integrating proprietary data and service APIs will be explored based on the data and services provided by the project's stakeholders (i.e. the MobiDataLab reference group).

5.2.3. Processors

In general terms we define a processor within the Transport Cloud as a component that models a function, which operates on some input data according to some specific logic in order to produce a final output. Such a definition can be used to instantiate the notion of processor in several different ways.

In the context of mobility data sharing, data processors may be needed to perform e.g. semantic enrichment based on common vocabularies, geographical enrichment based on common geometries, data format translation, data fusion, data anonymisation, injection of license specification, and any other data processing task that is relevant to the goals of the project.

5.2.3.1. Data Enrichment Processors

Positioned at the centre of a maze of information flows, the MobiDataLab project intends to fulfil two missions:

- To be a standard and universal entry point to various actors in the mobility industry;
- To provide users with mobility data enriched with on-demand elements from third-party partners.

Enhancing mobility data means adding optional information to the API responses including e.g. points of interest around the end-user, their accessibility to wheelchairs and strollers, the availability of bike sharing services, battery chargers, the weather conditions, etc.

This on-demand enrichment is achieved thanks to:

- APIs entry point: the user will specify here the geographical focus point for which he/she wants to obtain mobility-related information;
- The processing back-office unit: it will consult the content of the data supplied by the data providers, assemble and render the set of features that matches the location; then it will produce the response in the requested format and allow the API to respond to the user.

5.2.3.2. Privacy, anonymisation

The privacy and anonymisation components in the MobiDataLab Transport Cloud are responsible for assessing the privacy risks in the information flow through the computation of privacy metrics on transport data (such as the unicity of trajectories, see D2.3: "State of the art on Mobility and Transport data protection technologies" for more details), and the execution of anonymisation procedures, both for real-time data and for data stored in the Transport Cloud.

Data anonymisation irreversibly transforms data in a privacy preserving way. The outcome is still clear data that can be of use for the users, but with a lower accuracy (and, thus, a lower disclosure risk) than the original data.

Data anonymisation is performed once at the storage stage; after that, any query on the data (search, retrieval, calculations) are transparent, even though they may result in approximate results.

Anonymisation can also occur in real-time, but this comes at a price: not knowing the distribution of data to protect will make it more difficult to achieve acceptable levels of privacy and utility.

Two modes of anonymisation mechanisms are proposed:

- Real-time data anonymisation often relies on the distortion or generalisation of user positions, or on the partition of real-time user trajectories;
- Historical or aggregated data anonymisation typically involves the transformation of data to achieve a privacy guarantee dictated by some privacy model, such as k-anonymity and its extensions or differential privacy.

Funded by the European Union

6. Candidate Technologies

The MobiDataLab project aims to centralise access to data from various data sources, thus allowing users to browse and access data through a single entry-point. Data may be stored either in the Transport Cloud or within third-party infrastructures. Moreover, the Transport Cloud shall provide access to a plethora of different data formats that are relevant to the MobiDataLab project, for instance, structured data from relational databases (rows and columns), semi-structured data (CSV, logs, XML, JSON), unstructured data (documents, PDFs), and binary data (images, video).

Strong candidate technologies to be exploited for the implementation of the MobiDataLab components are presented in the following subsections and summarised in Figure 3.

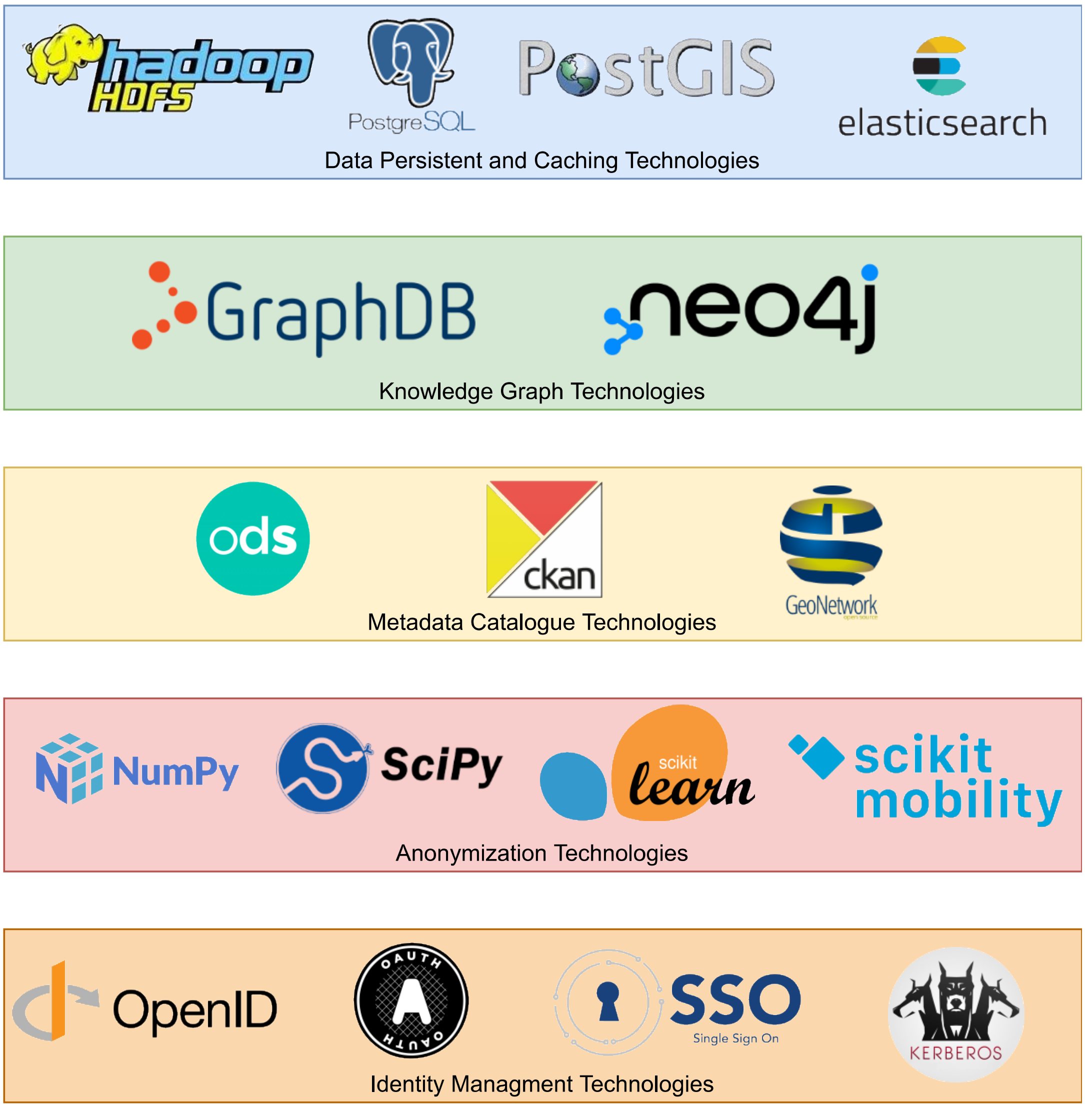
hadoop HDFS
PostgreSQL
PostGIS
elasticsearch
Data Persistent and Caching Technologies
GraphDB
neo4j
Knowledge Graph Technologies
ods
ckan
GeoNetwork
Metadata Catalogue Technologies
NumPy
SciPy
scikit learn
scikit mobility
Anonymization Technologies
OpenID
OAUTH
SSO
Single Sign On
KERBEROS
Identity Managment Technologies

Figure 3: Candidate Technologies and Frameworks

Funded by the
European Union

6.1. API Gateways

Acting as an entry point to the information system, an API gateway is primarily a switching entity between the client and the information backend. Requests will come from the users in different formats in accordance with the user expectation: a user may expect timetables while another will be looking for baby cart compatibility, elevator for wheelchairs, etc. Each format will be analysed by the API gateway before being forwarded to the processing unit hosted by the MobiDataLab platform. That processor will respond to the API gateway with the response to the initial query (electric car charger availability around the user GPS position, etc.). The API gateway will reply to the client in a second phase.

The API gateway is at the frontmost point of the infrastructure and has the duty to welcome users, but also to face potential threats inherent to internet activity such as Denial of Service (DOS) by flooding the gateway with high volumes of traffic or software attacks by combining set of data in order to reveal potential vulnerability and take partially advantage of the platform.

A trade-off must be found between programming such interface that will satisfy the previously described requirements and, alternatively, the time spent in configuring a fully functional API gateway broadly used.

Many API Gateways are actively available on the market, as shown in Table 1 listing a selection of API gateway solutions:

Table 1: API Gateway Solutions

Framework	Pros	Cons
Kong	• Includes key authentication • Includes traffic rate limitation • Free use licence	• Poor Graphical User Interface (GUI) • Configured with REST commands
Tyk.io	• GUI interface • JavaScript Object Notation (JSON) like config language	• Fully licenced • Cloud Infrastructure as a Service (IAAS) and hosted by editor
Apigee	• Edge version from Google • Google account required	• Poorly supported • No software packaging • Poorly documented
Express Gateway	• Node/Express JavaScript (JS) friendly • Multiple modules needed → high cost • Red Hat Package Manager • (RPM) package ready • Microservices oriented • Key authorisation/rate limiter • Commercial support available	• Dedicated shell (still easy configuration) • No GUI

Funded by the
European Union

Goku	• Design for large scale • Design for large volumes • Nice GUI • Identity management module • Load balancing customisation	• Curl configuration interface
Apache APISIX	• Design for large scale • Design for large volumes • Nice GUI • Identity management module • Load balancing customisation	• Curl configuration interface
Gloo	• Nice GUI • External authentication	• Kubernetes oriented • Dedicated Command Line Interface (CLI) • Partly under licence
Krakend	• Active development • Cloud and serverless oriented	• Commercial Licencing • IAAS hosted by the company → captive usage

6.2. Data Persistence and Caching Technologies

This section considers the technologies needed to enable the storage and caching of data within the Transport Cloud.

The following paragraphs therefore focus on solutions aimed at storing possibly large and heterogeneous data efficiently, thus including caching technologies aimed at providing a fast and latency-reduced access to data exhibiting strong spatial or temporal locality.

6.2.1. Distributed File System Technologies

Hadoop Distributed File System (HDFS), the commonly known file system of Hadoop and HBase (Hadoop’s database), is the most topical and advanced data storage and management systems available in the market. Both HDFS and HBase are capable of processing structured, semi-structured as well as unstructured data.

HDFS is fault-tolerant by design and supports rapid data transfer between nodes even during system failures. It provides also redundancy and supports high availability. HDFS is most suitable for performing batch analytics.

6.2.2. Database Technologies

PostgreSQL is a powerful, open-source object-relational database system with over 30 years of active development that has earned a strong reputation for reliability, feature robustness, and performance.

Elasticsearch is a distributed, RESTful search and analytics engine capable of solving a growing number of use cases. It is designed to provide excellent performance in different deployment settings, including an official Hadoop connector that will be integrated in the MobiDataLab Transport Cloud platform.

6.2.3. Caching technologies

Redis is an in-memory data structure store, used as a distributed, in-memory key-value database, cache and message broker, with optional durability. Redis supports different kinds of abstract data structures, such as strings, lists, maps, sets, sorted sets, HyperLogLogs, bitmaps, streams, and spatial indices.

Memcached is a general-purpose distributed memory-caching system. It is often used to speed up dynamic database-driven websites by caching data and objects in Random-access memory (RAM) to reduce the number of times an external data source must be read. Memcached is free and open-source software.

6.3. Knowledge Graph Technologies

In the context of the MobiDataLab project there is the need to store ontological and semantic information. To this end we will consider the industry-leading RDF triple store GraphDB from Ontotext1. GraphDB is a semantic graph database fully compliant with the relevant W3C standards (i.e., RDF, OWL, SPARQL). GraphDB is designed to be highly efficient and robust at large-scale use cases. Moreover, it is one of the few triple store solutions that provide semantic inferencing, enabling its users to derive additional implied facts from the facts already extant in the store.

Another interesting technology we will consider is Neo4j[1], which is an open-source property graph store. This store supports atomicity, consistency, isolation, durability (ACID) transactions and has high-availability clustering for enterprise deployments. The store is accessible from most programming languages using its built-in REST web API interface, and a proprietary Bolt protocol with official drivers.

[1] https://neo4j.com/

Funded by the European Union

6.4. Metadata Catalogue Technologies

A metadata is a descriptive text file into which the data service presents a description of the available datasets to the users; by many means, the metadata can be compared to a header or a presentation cover page that describes a whole set of serialised data. Basically, metadata can be seen as the data describing the dataset itself. Once the user/client is informed about the content of the metadata, that client is therefore instructed about the location of the dataset, but also on the type of data as well as the way to fetch these data from the entry point to query.

The concept of metadata catalogue is a standard interface between a service and a user whose two steps approach is (1) query the metadata service in order to get the list of available metadata and (2) query the service and target precisely one metadata file to get the expected metadata file itself.

Several metadata catalogues exist for the discovery of datasets and the sharing of the corresponding metadata, in particular open source solutions like OpenDataSoft[2], Comprehensive Knowledge Archive Network (CKAN[3]) and GeoNetwork[4]. Complying with metadata exchange standards, the MobiDataLab Transport Cloud is able to interoperate with each of them:

OpenDataSoft (https://www.opendatasoft.com/) is a cloud-based solution allowing users to publish, visualise, and share data.

CKAN (https://ckan.org/) is an open source tool used by data providers to publish their data though a web portal. CKAN is a widely used solution with an active community; Unlike OpenDataSoft, hosting must be provided independently of the platform. CKAN itself is not specifically related to mobility data.

GeoNetwork (https://geonetwork-opensource.org/) is a catalogue application for managing spatially referenced resources. It offers powerful metadata editing and searching features, an integrated interactive web map viewer and is based on open standards.

Harvesting is an important feature of metadata catalogues. Harvesting is the process of ingesting metadata from remote sources and storing it locally in the catalogue for fast searching. Harvesters provide a way for administrators to easily create and update an important number of datasets by importing them from an external source such as a Catalogue Service for the Web (CSW). The Data Catalog Vocabulary (DCAT) is an important standard supported by these catalogues. (DCAT) is "an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web"[5].

[2] https://www.opendatasoft.com/
[3] https://ckan.org/
[4] https://geonetwork-opensource.org/
[5] https://www.w3.org/TR/vocab-dcat-2/

Funded by the European Union

The MobiDataLab project aims to provide user with data from different stakeholders (e.g. public transport authorities, cities, regions, etc.), that may have different approaches to open data; as a result, some data sources will pass content through the MobiDataLab processing units while other data sources will be self-explanatory. The MobiDataLab project will adapt to this paradigm by integrating metadata management within its data description strategy.

6.5. Anonymisation Technologies

Privacy risk analysis builds upon statistical analyses on mobility data, including the similarity of positions and trajectories, their spatiotemporal distance, their unicity, and their autocorrelations through time.

On the other hand, the protection of positions and trajectories, whether following some privacy model such as k-anonymity[6] or differential privacy, entails their generalisation, their distortion by adding random noise, often from specific distributions, and/or their clustering based on their spatiotemporal distances.

Both statistical analyses and transformations can be supported by specialised numerical scientific analysis, as well as machine learning Python packages, such as Numpy[7], Scipy[8], and scikit-learn[9]. We further mention scikit-mobility[10], which includes utilities for synthetic mobility data generation, privacy risk analysis for mobility data, and other general statistical analyses specifically targeting mobility data.

6.6. Identity Management Technologies

In this section we focus on the technologies needed to enable the Identity and Access Management (IAM) needs of the Transport Cloud.

The analysis will be performed identifying main solutions for identity provisioning, authentication, authorisation and Single Sign-On.

Identity Provisioning. The most used standards for identity provisioning are the Lightweight Directory Access Protocol (LDAP), the System for Cross-domain Identity Management (SCIM). and the Service Provisioning Markup Language (SPML).

[6] https://en.wikipedia.org/wiki/K-anonymity
[7] https://numpy.org
[8] https://scipy.org
[9] https://scikit-learn.org
[10] https://github.com/scikit-mobility/scikit-mobility/

Funded by the European Union

Authentication. The most widely used standards for Authentication are:

- **Kerberos**, a computer network authentication protocol which works on the basis of 'tickets' to allow nodes communicating over a non-secure network to prove their identity to one another in a secure manner;
- **OpenID connect**, a simple identity layer on top of the OAuth 2.0 protocol. It allows clients to verify the identity of the end-user based on the authentication performed by an Authorisation Server, as well as to obtain basic profile information about the end-user in an interoperable and REST-like manner;
- **Security Assertion Markup Language (SAML)**, an XML-based, open-standard data format for exchanging authentication and authorisation data between parties, in particular, between an identity provider and a service provider. SAML is a product of the OASIS Security Services Technical Committee.

Authorisation. The most widely used standards for Authorisation are:

- **XACML (eXtensible Access Control Markup Language)**," an OASIS standard that describes both a policy language and an access control decision request/response language (both written in XML). The policy language is used to describe general access control requirements, and has standard extension points for defining new functions, data types, combining logic, etc. The request/response language lets you form a query to ask whether or not a given action should be allowed and interpret the result. The response always includes an answer about whether the request should be allowed"[11];
- **OAuth**, an open standard for authorisation that provides client applications a 'secure delegated access' to server resources on behalf of a resource owner. It specifies a process for resource owners to authorise third-party access to their server resources without sharing their credentials. Designed specifically to work with Hypertext Transfer Protocol (HTTP), OAuth essentially allows access tokens to be issued to third-party clients by an authorisation server, with the approval of the resource owner. The client then uses the access token to access the protected resources hosted by the resource server. OAuth is commonly used as a way for Internet users to log into third party websites using their Microsoft, Google, Facebook or Twitter accounts without exposing their password. Current version is OAuth v2;
- **UMA (User-Managed Access)**, a profile of OAuth 2.0 defining how resource owners can control protected-resource access by clients operated by arbitrary requesting parties, where the resources reside on any number of resource servers, and where a centralised authorisation server governs access based on resource owner policies. Resource owners configure authorisation servers with access policies that serve as asynchronous authorisation grants.

SSO (Single Sign-On). SAML, OpenID Connect and Kerberos could be used to establish an SSO infrastructure as they define Identity Providers or Authentication Servers. Central Authentication Service (CAS) is another authentication protocol specifically targeted to SSO.

[11] https://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.html

Funded by the European Union

Although the choice of one or more of these solutions for the implementation of the Transport Cloud is not part of this version of the document, we can nevertheless make some assumptions. For instance, in an "API gateway scenario", OpenID Connect could be used to perform the initial authentication of the end user signing into the front-end login interface. This initial authentication includes an OAuth2 access token. When the request arrives at the API gateway, the access token must be extracted from the request, validated and "replaced with an access token whose scope matches the API provider's scope". (Broeckelmann, 2017)

6.7. Other Promising Initiatives

As already highlighted in the deliverable D2.6, the GAIA-X project is the most relevant European initiative to the MobiDataLab project, as the former considers many aspects that are key to the realisation of the latter's Transport Cloud.

We also observe that the GAIA-X project did not produce any prototype so far, thus no technical solution can be presently deployed or evaluated within the context of the MobiDataLab project. Considering GAIA-X relevance, however, we will keep track of its evolution and adopt the necessary actions should important developments arise.

7. Summary and Conclusions

The present document reports on the functional and non-functional requirements posed by the initial specification of the MobiDataLab use cases, as the latter are detailed in the deliverables of the relevant work packages. Furthermore, an initial set of non-functional requirements has been reported, based on the observations of the project's partners.

Based on the elicited requirements, the first specification for design of the Transport Cloud platform's architecture has been produced. The specification aims to cover the requirements by using established standards and components, ensuring that the resulting platform will be maintainable and extensible in the long term.

While the present version of the architecture is complete in the sense that it covers all major operations expected by the platform, it will be naturally refined and extended as work on the project progresses and further details and intricacies of each use case become apparent.

To this end, the specification report will be treated as a live document, with updates applied whenever a major shift on the requirements is identified.

In the forthcoming period, the report will serve as the guide for development and integration work on the project's technical work packages. Initial versions of the tools and components foreseen in the architecture will be provided by the partners involved and will be used in the first deployment of the Transport Cloud platform to be evaluated by the end-users.

Feedback will be subsequently used to update the architectural specification and consequently the entailed components in the second and final version of this deliverable.

8. Bibliography and References

Broeckelmann, R. (2017). *Identity Propagation in an API Gateway Architecture*. Retrieved from https://cloud.google.com/blog/products/api-management/identity-propagation-in-an-api-gateway-architecture

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).

The content of this document reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein. The MobiDataLab consortium members shall have no liability for damages of any kind that may result from the use of these materials.