MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D2.4 State-of-the-art on Mobility Data sharing standards

23/01/2023

Author(s): Thierry CHEVALLIER, Eva ANDRZEJEWSKI, Alberto BLANCO JUSTICIA, Patrizio DAZZI, Didier DE-RYCK, Frédéric LATRUBESSE, Johannes LAUER, Chiara RENSO, Sebastian VAN DE HOEF, Chris WONG

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

Summary sheet

Deliverable Number	D2.4
Deliverable Name	State-of-the-art on Mobility Data sharing standards
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	CHEVALLIER Thierry (AKKA)
Contributing Partner(s)	AKKA, CNR, HERE, HOVE, URV
Peer Review	AETHON, KUL
Contractual Delivery Date	31-07-2021
Actual Delivery Date	05-10-2021
Status	Final
Dissemination level	Public
Version	V2.0
No. of Pages	59
WP/Task related to the deliverable	WP2/T2.3
WP/Task responsible	AKKA/AKKA
Document ID	MobiDataLab-D2.4-StateOfTheArtOnMobilityDataSharingStandards-v2.0
Abstract	This document provides an updated state-of-the-art review on existing data sharing standards.

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

Funded by the European Union

Document history

Version	Date	Organisation	Main area of changes	Comments
0.1	17/06/2021	AKKA, CNR, HERE, HOVE, URV	all	Draft for Review
0.2	02/07/2021	AETHON, KUL	all	Review
0.3	15/07/2021	Advisory Board	all	Review
0.3	23/07/2021	AKKA	all	Rework
0.4	29/07/2021	AKKA	all	Quality Check
1.0	29/07/2021	AKKA	all	Final version
2.0	05/10/2021	AKKA	Table of content	Dead link removed

Executive Summary

Availability of mobility data is subject to compatibility, interoperability and standards issues. For solving the interoperability issues, the MobiDataLab aims to conduct a theoretical study on state-of-the-art data sharing standards. This is the topic of the Task 2.3 of the project (Standard Requirements) and of the following deliverable. This study covers public transport Data formats (Transmodel, NeTEx, SIRI, GTFS, etc.), data for micro-mobility and shared mobility, road data, pricing and ticketing data. Cross-domain standards are also considered, for the federation of cloud services, data catalogues (e.g. with DCAT-AP), and for the sharing of geospatial data and Linked open data. More generally this document provides an updated state-of-the-art review on existing standards and provides guidance for the developments in the project and identifies gaps to solve.

Table of contents

Funded by the European Union

List of figures

List of tables

Abbreviations and acronyms

Abbreviation	Meaning
API	Application Programming Interface
AWS	Amazon Web Services
CEN	Comité européen de normalisation (European Committee for Standardisation)
CNA	Calypso Network Association
CSV	Comma-Separated Values
CSW	Catalogue Service for the Web
DCAT	Data Catalog Vocabulary
GBFS	General Bikeshare Feed Specification
GML	Geographic Markup Language
GTFS	General Transit Feed Specification
IAM	Identity and Access Management
IdP	Identity Provider
ISO	International Standardisation Organisation
ITS	Intelligent Transport Systems
KVP	Key-Value pairs
LOD	Linked Open Data
JSON	JavaScript Object Notation
MaaS	Mobility as a Service
MDS	Mobility Data Specification
MMTIS	Multi-Modal Travel Information Services
NAP	National Access Point

NeTEx	Network timetable Exchange
NFC	Near Field Communication
NFV	Network Function Virtualisation
OGC	Open Geospatial Consortium
OMF	Open Mobility Foundation
OJP	Open Journey Planning
OpRa	Operating Raw Data
OSM	OpenStreetMap
OSPT	Open Standard for Public Transport
OWL	Web Ontology Language
RDF	Resource Description Framework
REST	Representational State Transfer
S3	Simple Storage Service
SAML	Security Access Markup Language
SDN	Software Defined Network
SIRI	Service Interface for Real-time Information
SOAP	Simple Object Access Protocol
SP	Service Provider
SPARQL	SPARQL Protocol and RDF Query Language
SSO	Single Sign-On
TEN-T	Trans-European Transport Network
TOSCA	Topology and Orchestration Specification for Cloud Applications
UML	Unified Modelling Language
WFS	Web Feature Service

Funded by the European Union

WMS	Web Map Service
WMTS	Web Map Tile Service
W3C	World Wide Web Consortium
WP	Work Package
XMI	XML Metadata Interchange
XML	eXtensible Markup Language

Content

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 works to foster the sharing of data amongst transport authorities, operators and other mobility stakeholders in Europe. MobiDataLab develops knowledge as well as a cloud solution aimed at easing the sharing of data. Specifically, the project is based on a continuous co-development of knowledge and technical solutions. It collects and analyses the advice and recommendations of experts and supporting cities, regions, clusters and associations. These 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.

We thank **Pedro Barradas from ARMIS, Leo Frachet, and Tu-Tho Thai from MobilityData** for their insightful comments during the preparation of this document.

1.2. Purpose of the deliverable

Interoperability issues are one of the main barriers which prevent mobility data providers to share their data. The lack of interoperability not only decreases connectivity across transport systems, and therefore, reduces the overall efficiency of transport systems, but it also increases the cost of opening data because the complexity of opening data as compared to the value derived from the cost of the operation is very high.

For solving this interoperability issue, MobiDataLab proposed to conduct theoretical studies and to consolidate an Open Knowledge Base of standards as part of the Work Packages (WP) 2 of the project. More specifically the aim of MobiDataLab Task 2.3 (Standard Requirements) is to define a standardisation roadmap for both integrating existing data sharing standards (GTFS, NeTEx, SIRI, MDS, OGC features, CSW, DCAT-AP, etc.) and for promoting standardisation results. In the end this task will result in the mapping of technical requirements to current standards and any necessary extensions for development across all relevant WPs.

To achieve this objective, two deliverables are planned: a first one consisting of a state of the art of the different standards existing today in the field of data sharing and their applicability to MobiDataLab (D2.4), and a second one proposing suggestions about future standards and norms to be adopted for improved data sharing (D2.5).

The present D2.4 makes an inventory of standards for sharing mobility data, whether from public transport, road traffic, new forms of mobility, infrastructure, etc. Data sharing standards for private mobility data in the context e.g. of Mobility as a Service (MaaS) and ticketing data are also considered. Not only vertical (or sector-specific) standards and standardisation bodies are considered, but also horizontal or multi-sector ones, like the geospatial data standards, metadata standards and semantic web standards.

For every standardised domain identified, the MobiDataLab partners contributing to this document have tried to answer the following questions:

- *Why is this standard of interest for MobiDataLab?*

In fact, the purpose of this document is not to make a repetition of other standardisation works like Data4PT, Mobility-Data, MaaS Alliance, OMF, OGC, W3C, etc. especially as some members of these projects are part of the Advisory Board. The standard requirements for MobiDataLab are rather built on the objective to ease the combination of data from different domains, in order to enrich them, experiment with them and innovate during the Living Labs. The following study of standards is therefore quite broad in scope and at the same time focused on what will be useful in the MobiDataLab context.

- *How will MobiDataLab use this standard?*

MobiDataLab partners need to make sure that the following standards will either be integrated in the solution implementation (the MobiDataLab Transport Cloud), or be integrated in solutions that will be used in this context (e.g. GeoServer with WFS, HERE Real Time Traffic API with DATEX II, GTFS with Navitia, etc.)

- *How will MobiDataLab follow the evolutions of this standard?*

D2.4 is a state-of-the art, which means that we are interested in the standards in their current state, i.e. in 2021. On the other hand, the consortium needs to keep updated of the latest evolutions of these standards, and include new standards that may appear in the course of the project. Standards are dynamic, they evolve and they are probably best shown on a timeline (and a standard timeline can be put in parallel with a regulation timeline for instance, this is an approach to be taken for the v2). The best way to follow the evolution of standards is to be part of the working groups for the different standards, because in the end standards are always people who agree on a common way to address a specific problem. The MobiDataLab consortium will keep informed thanks to the advisory board, the reference group of stakeholders, and through conferences and events.

1.3. Intended audience & Review process

The dissemination level of this D2.4 deliverable is ‘public’ (PU). AKKA as WP2/Task 2.3 leader is responsible for it with the contribution of CNR, HERE, HOVE and URV. Appointed peer reviewers are AETHON and KUL. An external review is also conducted by members of the Advisory Board.

It should be noted that the content of this WP2 deliverable will be included in the MobiDataLab Open Knowledge Base and therefore it should be considered as a living document. It will be followed by another deliverable (D2.5) at the 30th month of the project (M30, July 2023).

1.4. Structure of the deliverable and its relation with other work packages/deliverables

This document will serve as an input for most of the WP4 tasks (Transport Cloud prototype). More specifically the Mobility data sharing standards will be supported by the Reference Data catalogue (Task 4.2) while the standards for data exchange in the cloud will be considered for the Architecture and design of the cloud solution (Task 4.1) and Data Privacy (Task 4.5). Also, the best practices for sharing data on the web are very important recommendations for the Data Access Services and Data Channels (Task 4.3), and the Geodata sharing standards and Semantic Interoperability standards will be covered by the Data processors (Task 4.4).

Not only the WP4 is related to this deliverable but also the WP2 itself. De jure standards, i.e. standards according to law such as the Transmodel standards (NeTEx, SIRI, etc.) or the INSPIRE data models, are by definition related to the regulations (Task 2.1). Also data sharing standards are important for defining the use cases (Task 2.6) as most of the standards propose typical use cases. This means that including a standard in the MobiDataLab solution means thinking about its application in relevant use cases. And on the other hand, proposing a use case means implementing the most appropriate standard(s) for this use case.

Standard requirements are also related to WP1. In particular the Data Management plan (Task 1.4) will rely on standardised data models (e.g. Transmodel which proposes data categories for mobility) for defining types of mobility data. Also the relationships of the MobiDataLab consortium with the experts from the Advisory Board (Task 1.5) are extremely important to follow the evolution of the different standardisation efforts, through working groups in different sectors such as public transport, new forms of mobility, Mobility as a Service, geolocation data, etc.

2. Mobility data sharing standards

The field of mobility data is highly standardised, with standards for static or real-time data in e.g. public transport, road traffic management and micro-mobility services. Journey planning applications and Mobility-as-a-Service also depend on standards, e.g. for exchanging ticketing information. However, challenges still exist in the progress towards standards and common protocols. The purpose of this section is to provide an overview about the standards themselves, the standardisation bodies behind them, and their current adoption.

2.1. Public transport data

Public transport standards are related to the European Directive 2010/40 on Intelligent Transport Systems (ITS Directive) that aims to:

- Establish a framework for coordinated and effective deployment and use of ITS;
- Set common priorities;
- Develop specifications and standards[1].

On the other hand, public transport standards are also related to the relevant Delegated Regulations concerning the several aspects of transport, i.e. Delegated Regulations on:

- Interoperable EU-wide emergency call or eCall (305/2013);
- Road safety-related minimum universal traffic information free of charge to users (886/2013);
- EU-wide real-time traffic information services (2015/962);
- EU-wide multimodal travel information service (2017/1926).

The idea of the last three Delegated Regulations is to develop specifications to be followed in order to ensure the compatibility and interoperability among different systems and to provide all the information so that they are accurate and viable across borders. They indicate how to ensure a correct formal communication among member states and how to develop National Access Points (NAP).

NAPs are a mechanism for accessing, exchanging and reusing transport related data under Delegated Acts of the ITS Directive (2010/40/EU): they constitute *"a single point of access for users to at least the static travel and traffic data and historic traffic data of different transport modes, including data updates, provided by the transport authorities, transport operators, infrastructure*

[1] For more information about what the ITS directive mandates on standards see Article 8 and Annex II https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040

managers or transport on demand service providers within the territory of a given Member State'[2]. Every country has its own NAP (for example in Germany, the data is provided via *Mobilitäts Daten Marktplatz MDM - MDM Portal*[3] and in France via *Transport Data Gouv*[4]).

Indeed, common specifications implementation allows reusers to develop a specific solution to each supplier and therefore allows everyone to reuse data on extended perimeters. Without the implementation of standards, only the main stakeholder such as Google would have the capacity to exploit shared data. Furthermore, the reuse of data would remain local. The Delegated Regulation on multimodal travel information services recommends using NeTEx and SIRI European standards to exchange static information and dynamic (real-time) information respectively and to implement and use the Open Journey Planning API (Application Programming Interface) for the implementation of the so-called distributed journey planner, so to guarantee the correct formal data exchange among journey planners in order to implement the distributed journey planner.[5]

In the following document, we are going to describe different standards of data of the public transport sector. Before detailing them, it is important to make a difference between *de jure* and *de facto* standards.

De jure standards are definitions, descriptions or procedures defined by a group of experts and validated by a commission that can be national (e.g. AFNOR), European (CEN) or international (ISO). NeTEx and SIRI are public transport data exchange standards validated at European level by the CEN[6].

De facto standards are definitions, descriptions, or procedures which have been imposed by usage, most often by a commercial will and force. GTFS Schedule, GTFS Realtime and GBFS are widely used *de facto* standards worldwide.

The establishment of both de jure and de facto standards is crucial to:

- develop open data;
- provide data that can be reused by all and on a wide perimeter.

2.1.1. Transmodel

In this section we provide an introduction of Transmodel standard for public transport data exchange. The Delegated Regulation on EU-wide multimodal travel information services establishes a

---

[2] Delegated Regulation on EU-wide multimodal travel information services (https://eur-lex.europa.eu/eli/reg_del/2017/1926/oj), Article 3 – National Access Points
[3] https://www.mdm-portal.de/
[4] https://transport.data.gouv.fr/
[5] Delegated Regulation on EU-wide multimodal travel information services (https://eur-lex.europa.eu/eli/reg_del/2017/1926/oj), Article 4 and Article 5
[6] https://www.cen.eu/

Funded by the European Union

necessary specification to ensure EU multimodal travel information services and this includes the use of a standard conceptual data reference model: Transmodel.

Transmodel is the European Reference Data Model for Public Transport[7]; it provides an abstract model of common public transport concepts and structures that can be used to compare and reconcile information architectures, databases and data exchange interfaces. This model also aims to build many kinds of public transport information systems, including for timetabling, fares, operational management, real time data, etc.

It can be compared to a dictionary which describes the public transport wording and, above all, the links between all these terms.

To be more precise, Transmodel carefully separate 8 concepts to model Public Transport information in IT systems. The detail of these concepts and their format are available in the following scheme (Figure 1):

- P1: Common concepts: concept shared by the different functional domains covered by the model;
- P2: Network description: stop, line, infrastructure, stakeholder;
- P3: Timing Info and vehicle scheduling: Runtimes, vehicle journeys and day type-related vehicle schedules;
- P4: Operational monitoring and control: Operating day-related data, vehicle follow-up and control actions;
- P5: Fare management: Fare structure definition, sales, validation and control of access rights
- P6: Passenger Information: Relevant information on the planned and real-time service;
- P7: Driver management: Definition of day-type schedules, order of driver duties and recording of driver performance;
- P8: Management information and statistics: Additional descriptions in the sub-models, including data dedicated to service performance indicators

[7] http://www.transmodel-cen.eu/transmodel-at-a-glance/

Funded by the European Union

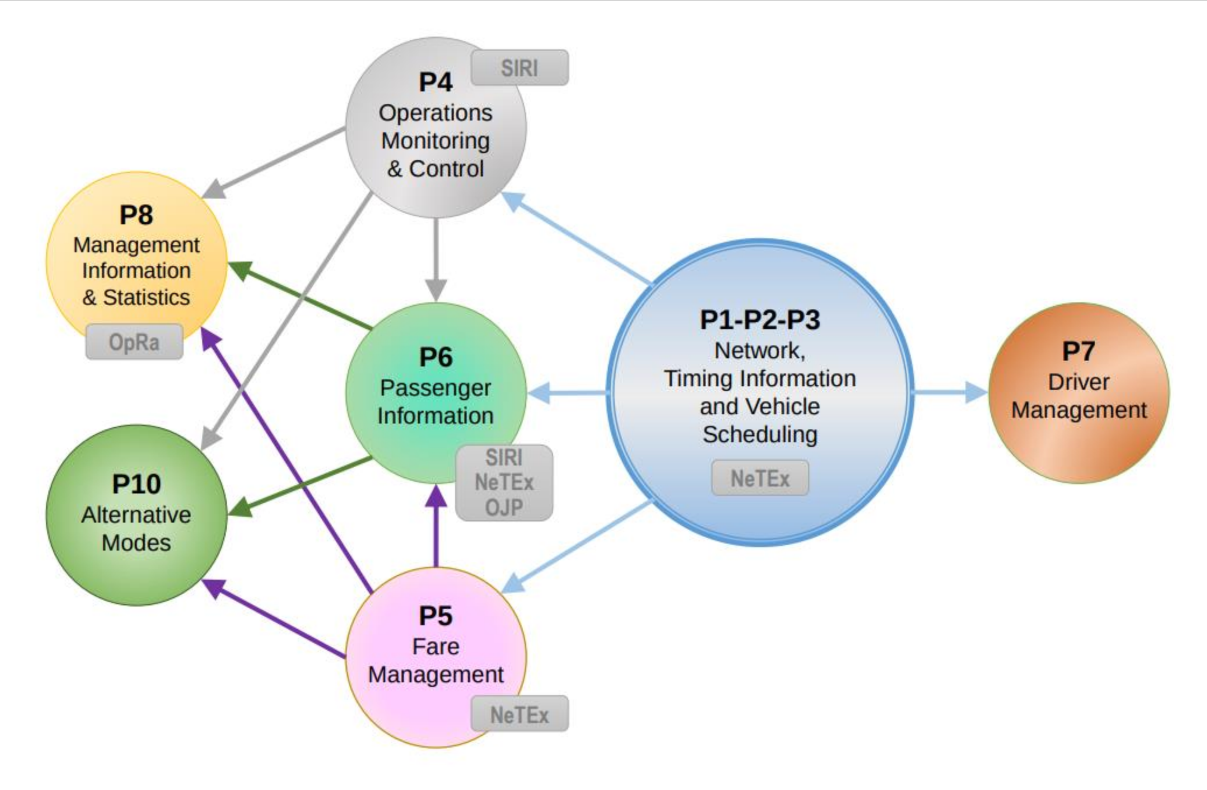

Figure 1: Transmodel's parts (source: Transmodel CEN website[8])

Another concept is currently being implemented in Transmodel: the alternatives modes concept with bike sharing, car sharing and further detailed information about e.g. cycling network (cf. P10 in the above figure).

2.1.2. Static data exchange

Static data are *"data relating to different transport modes that does not change at all or does not change often, or change on a regular basis"* (Delegated Regulation on EU-wide multimodal travel information services, Article 2). Public transport data sharing is primarily about exchanging static information (e.g. timetables, network topology and accessibility, etc.) and in this process there is a timeline requested by the European Commission with different maturity situations between Member States.

[8] http://www.transmodel-cen.eu/related-standards

2.1.2.1. NeTEx (Transmodel standard)

Network Timetable Exchange (NeTEx) is a CEN Technical Standard[9] for exchanging public transport network, schedules and related data. It provides a means to exchange data for passenger information such as stops, routes timetables and fares, among different computer systems, together with related operational data. It can be used to collect and integrate data from many different stakeholders, and to reintegrate it as it evolves through successive versions.

As detailed in the NeTEx Transmodel description page[10], the NeTEx schema can thus be used to exchange:

- Public Transport schedules including stops, routes, departures times / frequencies, operational notes, and map coordinates;
- Routes with complex topologies such as circular routes, cloverleaf and lollipops, and complex workings such as short working and express patterns. Connections with other services can also be described;
- The days on which the services run, including availability on public holidays and other exceptions;
- Composite journeys such as train journeys that merge or split trains;
- Information about the Operators providing the service;
- Additional operational information, including, positioning runs, garages, layovers, duty crews, useful for AVL and on-board ticketing systems;
- Data about the Accessibility of services to passengers with restricted mobility;
- Data is versioned with management metadata allowing updates across distributed systems;
- Fare structures (flat fares, point to point fares, zonal fares);
- Fare products (single tickets, return tickets, day, and season passes, etc.);
- Fare prices that apply at specific dates.

Who is already using NeTEx?

- In France: IDFM (Ile-de-France transport operator) plans to use NeTEx by middle of 2021. Also, most regional transport operators in France propose NeTEx exports. Some of them are actually GTFS files converted to NeTEx using GTFS to NeTEx open source tool;
- Norway has selected NeTEx at national level;
- In England: NeTEx was used for the London Olympics by Transport for London. The European Bus System of the Future 2 (EBSF11) project is also implementing NeTEx in London (by the company Hogia) (note: England also has its own system: TransXChange);
- In Italy: the 5T group in Piedmont;
- Germany and the Netherlands are finalising their profiles.

[9] http://netex-cen.eu/?page_id=11
[10] http://www.transmodel-cen.eu/standards/netex/
[11] https://cordis.europa.eu/project/id/636300/

2.1.2.2. GTFS Schedule

The General Transit Feed Specification (GTFS) data format contains two parts, the GTFS Schedule component (based on the CSV data format) and the GTFS Realtime component (based on Protocol Buffers data format).

The GTFS Schedule component, also known as GTFS "Static", defines a common format for public transportation schedules and associated geographic information. GTFS Schedule datasets let public transit agencies publish their transit data and developers write applications that consume that data in an interoperable way. A GTFS dataset is composed of a series of CSV text files collected in a ZIP file. Each file models a particular aspect of transit information: stops, routes, trips, and other schedule data. The details of each file are defined in the GTFS reference[12].

NeTEx covers a wider functional scope than GTFS: in addition to defining common rules on passenger information, NeTEx can also be used for information related to public transport operations. Nevertheless, NeTEx is used less often because of its complexity as explained in the previous section.

A converter GTFS2NeTEx in open source has been developed by HOVE on behalf of the French government. It aims to offer a tool for small and medium size stakeholders that would have difficulties to produce datasets directly in NeTEx.

2.1.3. Dynamic (real-time) data exchange

Standards for sharing real-time information (e.g. real-time passing time, incidents, occupancy, facility status, etc.) are needed to complement static exchange information. Both de jure and de facto standards are available, mainly SIRI and GTFS-RT.

2.1.3.1. SIRI (Transmodel standard)

Service Interface for Real-time Information (SIRI) is a CEN Technical Standard[13] that specifies a European interface standard for exchanging information about the planned, current or projected performance of real-time public transport operations between different computer systems. It allows pairs of server computers to exchange structured real-time information about schedules, vehicles, and connections, together with general informational messages related to the operation of the services.

---

[12] https://developers.google.com/transit/gtfs/reference
[13] https://www.siri-cen.eu/

Funded by the European Union

The information can be used for many different purposes, as detailed in the SIRI Transmodel description page[14]:

- To provide real time-departure from stop information for display on stops, internet and mobile delivery systems;
- To provide real-time progress information about individual vehicles;
- To manage the movement of buses roaming between areas covered by different servers;
- To manage the synchronisation of guaranteed connections between fetcher and feeder services;
- To exchange planned and real-time timetable updates;
- To distribute status messages about the operation of the services;
- To provide performance information to operational history and other management systems.

SIRI is vast and requires the definition of a profile, i.e. a subset that complies with the standard and meets a set of identified needs.

SIRI Profile France

In SIRI it is possible to define profiles that include only a part of the SIRI services. For example, Ile-de-France Mobilités (IDFM) has defined the SIRI Lite profile, oriented towards open data, which is based on the IDF profile but only includes certain services. A lighter French national profile (based on SIRI Lite) is being defined.

Elements, in the SIRI IDF profile, are:

- SM - Stop Monitoring: next passages;
- GM - General Message: message attached to any object (line, stop, route). Delivers information on disruptions or commercial information;
- ET - Estimated Timetable: advance/delay;
- VM - Vehicle Monitoring: vehicle position;
- PT - Production Timetable: theoretical schedules;
- CM - Connection monitoring: information on connections;
- SX - Situation Exchange: information on situations and their consequences (disturbances);
- FM - Facility Monitoring: equipment status.

2.1.3.2. GTFS Realtime

GTFS Real-time[15] is a component of GTFS (General Transit Feed Specification), which is a specification that allows public transportation agencies to provide real-time updates about their fleet to application developers. GTFS Realtime was designed around ease of implementation, good interoperability with GTFS Schedule and a focus on passenger information.

[14] http://www.transmodel-cen.eu/standards/siri/
[15] https://github.com/google/transit/tree/master/gtfs-realtime

Funded by the European Union

While SIRI is designed as an autonomous solution that does not need an external reference (NeTEx) to work, GTFS Realtime works in conjunction with GTFS Schedule: GTFS Schedule datasets define network topology, and GTFS Realtime only indicates the GTFS Schedule objects that have been updated in real-time (schedules, vehicle position, occupancy …) if they are different from their theoretical value. This approach makes GTFS Realtime lighter and therefore more efficient to process in real time than SIRI.

The GTFS Realtime data exchange format is based on Protocol Buffers[16]. Protocol Buffers are a language - and platform - neutral mechanism for serialising structured data (think XML, but smaller, faster, and simpler). The data structure is defined in a GTFS Realtime proto file, which is then used to generate source code to easily read and write your structured data from and to a variety of data streams, using a variety of languages – e.g. Java, C++ or Python.

2.1.4. Journey planning

2.1.4.1. Open API for distributed journey planning

The requirement to exchange accurate and timely information about public transport services and to implement systems able to provide multi-modal information for longer-distance journeys is crucial. Several systems have been developed to meet this requirement, based on different architectures, but the nature of the queries sent between the systems, and the content of the responses sent in return, were essentially the same. This suggests that it would be possible to define a single Open Journey Planning (OJP) API to support all distributed journey planning systems.

OJP allows a system to engineer just one interface that it can make available rather than having to engineer separate APIs for each bipartite exchange arrangement that may be required with other systems.

To summarise, the intention of the Open API is to provide an opportunity for just one universal channel to exchange information to lower-volume users.

[16] https://developers.google.com/protocol-buffers/

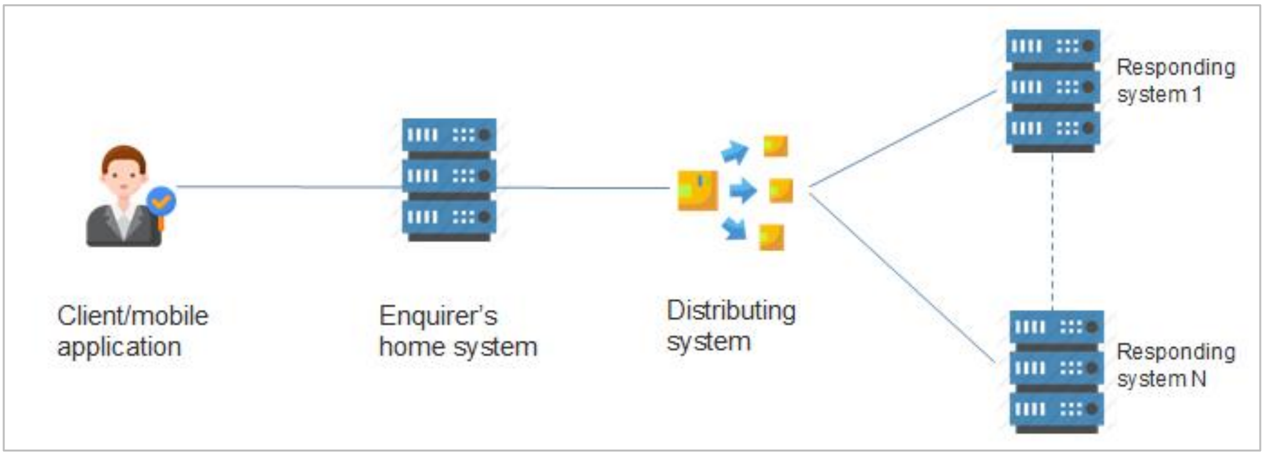

Figure 2: Principles of the Open API for distributed journey planning (HOVE)

The following descriptions assume a relatively simple and straightforward example of Distributed Journey Planning in which there are four key "actors" or "roles" in the process, as Figure 2 presents:

- Enquirer – the person asking for information;
- Enquirer's home system – the journey planning system to which the enquirer is connected;
- Distributing system – the system that distributes journey planning enquiries to other systems;
- Responding system(s) – the system(s) that respond to questions from the distributing system.

An enquirer requests a trip from a journey planning system which is referred to as the enquirer's home system – and the home system then has to recognise that the enquiry requires information from other systems. It therefore passes the enquiry to a distributing system (which may be integrated with the home system or physically separate) which makes enquiries of one or more responding system(s) from which it will glean additional information in order to put together a comprehensive trip plan to meet the enquirer's request.

Once the distributing system has the required comprehensive trip plan it returns this to the home system, which in turn presents the results to the enquirer. The home system will have a user interface through which it receives enquiries from enquirers, and through which it responds to those enquirers. It may have its own journey planning capability for a specific geographical territory (which could also cover only certain modes of transport) or it may only have sufficient intelligence to know that it needs to draw information from other journey planning systems by passing the enquiry to the distributing system.

The distributing system will need to know from which journey planning systems it needs to draw all relevant information to meet the requirements of each enquiry it receives, and it needs to be able to communicate with the relevant responding systems – which it would do using an API.

OJP has some drawbacks:

- Response times depend on the number of requests sent to journey planning system. If several journey planning systems are involved, then response time may not be satisfying enough for end users;

- If several journey planners are connected through OJP, then specific features of each journey planner cannot be used. Only a common set of features can be used, meaning that new mobility modes in particular may not be taken into account.

As transport data is not managed through a central repository, there may be issues or inconsistencies at the border between two systems, as each journey planning system will consider only local data to compute the "best" itineraries.

2.1.4.2. OpRa

Operating Raw Data and statistics exchange (OpRa) is a CEN initiative[17] with main focus on the identification of Public Transport raw data to be exchanged, gathered and stored in order to support *Study and Control* of Pubic Transport Service. This standard is currently in progress and in the long run, the work will consist in the production of a Technical Report, to document the results of the performed analysis. It will describe the recorded reality of operation, like delays and cancelled vehicle journeys, etc. either through individual measurements at a given sampling interval or in an aggregate ways (statistics).

Main areas that should be covered by OpRa are:

- Service Offer: Spatial and time coverage, offered seats, etc.;
- Service Demand: O/D matrix, load factor, etc.;
- Service Externality: pollution emissions, safety, etc.;
- Service Economy: incoming funds (sold tickets, etc.) and expenses (costs, etc.);
- Service Efficiency: lines overlaps, MTBF, delays, etc.

2.1.5. Applicability to MobiDataLab

The interest of the public transport standards for MobiDataLab are obvious as the goal of the project is to provide access to public transport data to be re-used by developers and innovators during innovation sessions, and to be able to transcode them.

Regarding how MobiDataLab will use these standards, HOVE will provide the Navitia API, the GTFS to NeTEx open source tool (developed for the French National Access Point), and other open tools aiming to analyse public transport data.

MobiDataLab will follow the evolutions of the public transport standardisation establishing connections (e.g. through the MobiDataLab Advisory Board[18]) with organisations such as the UITP

[17] Source : http://www.opra-cen.eu/faq/
[18] The MobiDataLab Advisory Board is composed of several individual experts from ongoing projects around mobility data. Its main role is to provide the MobiDataLab consortium with specific advices and strategic orientations to guide the developments of the project. More information is available on the WP1/Task 1.5

and projects such as Data4PT. HOVE is particularly involved in monitoring standardisation works for Transmodel, NeTEx and SIRI. The Figure 3 below shows a panorama of the corresponding standards.

GTFS intellectual propriety is still owned by Google and licensed under Apache 2.0. Its improvement and extension is under the management of MobilityData since 2018.

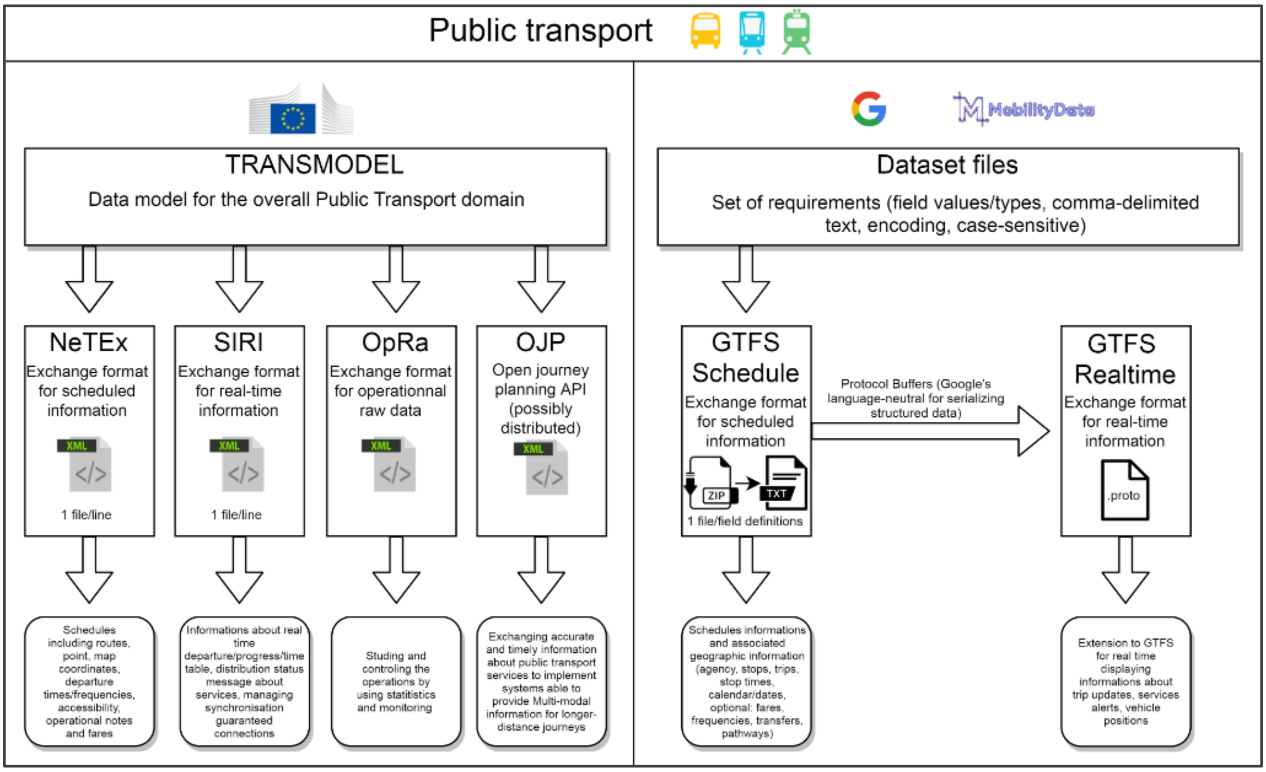

Figure 3: Panorama of standards for sharing public transport data (F. Latrubesse, AKKA)

2.2. Road traffic data

Road network operators, who are responsible for traffic management, travel safety, road conditions maintenance and user information, need good traffic knowledge, which is why quality traffic data is needed. Also, road users need to get fresh map data from the road operators to the vehicle's navigation systems. Sharing road traffic data actually depends on standards, like the DATEX II and TN-ITS described in this section.

2.2.1. DATEX II

DATEX II is the European standard for the exchange of traffic related data. It is a unified XML-based format to allow data exchange between service providers, traffic control centres, and road operators. It covers traffic and travel information such as traffic flow, traffic measures, roadworks, accidents, parking. The DATEX II specification for traffic management and traffic information has its origins

Funded by the European Union

back in the early 1990's and evolved via funding of the European Commission in 2003-2005. It is available under CEN/TS 16157. The DATEX II data model allows a number of extensions to meet specific requirements.

Who is using DATEX II?

The main use case for DATEX II is the sharing between road operators and service providers, with the focus on dynamic events. Among other examples, HERE Traffic Data Service (traffic flow and incidents) and TOMTOM Traffic Flow and Incidents are based on DATEX II.

These services are designed for:

- Server-to-server integration with traffic control centres, routing, navigation, and mapping applications;
- Real-time information about traffic incidents (accidents, road construction projects, traffic jams, travel weather warnings, road closures, etc.), their causes and impacts on travellers.

Data under the DATEX II format are being provided via National Access Points. The following map shows the distribution of DATEX II implementation across the European states:

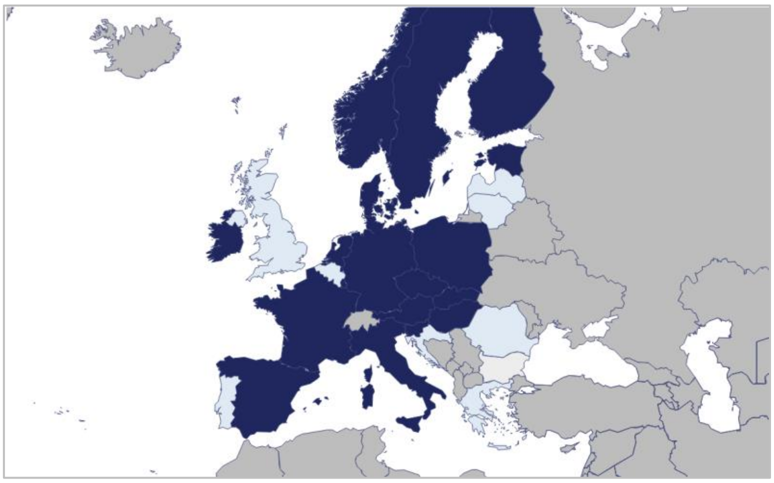

Figure 4: Implementation state for the Safety Related Traffic Information (source: DATEX II usage in NAPs[19])

Data Specification

The DATEX II data specification is available online[20] (version 3.1). The documentation consists of an object representation of the data schema, using the tool Enterprise architect.

[19] https://www.datex2.eu/naps
[20] https://docs.datex2.eu/_static/umlmodel/v3.1/index.htm

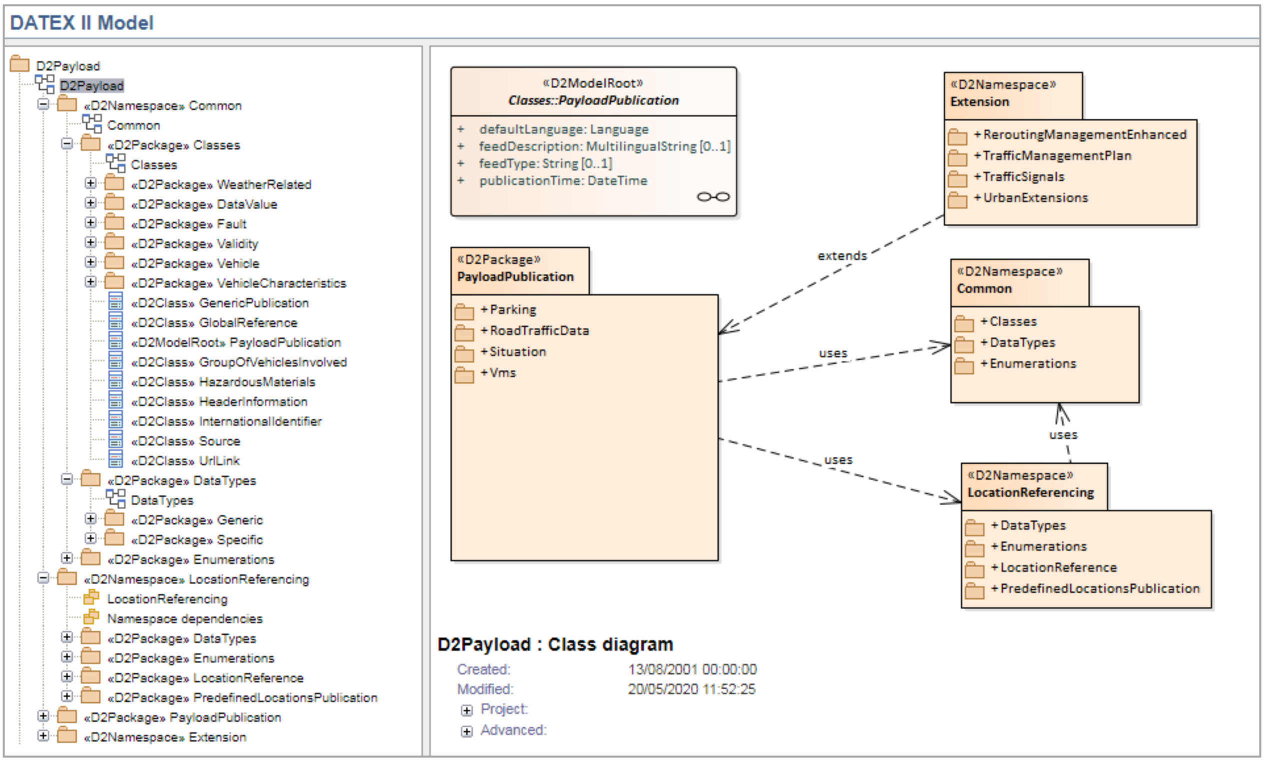

Figure 5: DATEX II Model documentation (screenshot)

This documentation shares the models using a format called XMI which allows a degree of interoperability between modelling tools and UML (as shown in Figure 5). Developers who do not need to look at the whole model can get straight to the technology specific artefacts like for example XML schemas.

Datex Organisation

The DATEX II project is organised in three layers[21]:

- Steering Group: Final responsibility, strategical and tactical decisions and representation;
- Stakeholder Advisory Board: Consists of any public or private organisation that wants to be linked to the project and advises the steering group;
- Technical Management Group: Within the technical management group, the activity leaders represent the project in technical matters and coordinate the technical activities.

2.2.2. TN-ITS

The TN-ITS CEN is a technical specification describing the exchange (of changes) of road attributes, emphasising on static road data, such as speed limits. It is maintained by the CEN, under their

21 https://www.datex2.eu/datex2/organisation

Technical Committee "Road Transport and Traffic Telematics (TC278)"[22]. The specification is built on top of the Geographic Markup Language (GML), which is an XML standard. The schema is available online[23].

2.2.2.1. Data providers

The TN-ITS Innovation Platform is coordinated by ERTICO and supported by the Joint Research Council (JRC), the CEN and the European Commission. Its current members are:

- Public Transport Authorities (Ireland, United Kingdom, Finland, Belgium, Norway, Sweden);
- Map Makers (TOMTOM, HERE Technologies, GeoJunxion).

2.2.2.2. TN-ITS GO

The EU is supporting further implementation of TN-ITS with a CEF funded project TN-ITS GO[24]. In TN-ITS GO 14 Member States will implement the TN-ITS framework on a national level. Ambition is to have data exchange for the TEN-T network[25] and bring current pilot implementation to operational processes. TN-ITS is already being used as a source for commercial digital maps. One activity in TN-ITS GO is the maintenance and development of the Technical Specification CEN TS 17268. For further development and stronger accessibility, the TN-ITS Platform is integrated together with DATEX II in the upcoming EU Project "Federation of national access points", which will start mid-2021.

2.2.3. Applicability to MobiDataLab

Road traffic standards in general and, DATEX II in particular, are relevant for MobiDataLab for the use cases about data analytics, Estimated Time of Arrival (ETA), and particularly in cases of traffic incidents. MobiDataLab will use DATEX II integrating it with the HERE platform (Real-Time Traffic API), establishing an interoperability with available streaming data.

[22] https://standards.iteh.ai/catalog/tc/cen/aa876bfe-1d4c-4f9a-8e50-2fcca8393dea/cen-tc-278
[23] http://spec.tn-its.eu/schemas/
[24] https://tn-its.eu/go-about.
[25] The Trans-European Transport Network (TEN-T) is a programme for the development of the European Union's transport infrastructure set up by the European Parliament and Council (cf. Regulation (EU) No 1315/2013)

Funded by the European Union

2.3. Micro-mobility and shared mobility

Most of the shared micro-mobility services available in many cities across Europe and beyond are using bicycles, e-scooters and other small devices. These vehicles can be found in fixed stations, or “free-floating” in public spaces (i.e., not fixed to specific locations). According to the report by MobiDataLab partner POLIS[26], "*digital technologies and geolocation play a central role in these services, and their operation generates high volumes of data, about the number, spatial distribution, and status of these vehicles, but also about who is using them, to travel where, and when.*" Two standards are mainly used to share the corresponding data: GBFS and MDS which we describe here.

2.3.1. GBFS (General Bikeshare Feed Specification)

The General Bikeshare Feed Specification[27], known as GBFS, is the open data standard for shared mobility, including shared bicycles, mopeds, scooters, and cars. GBFS is used to produce real-time data feeds in a uniform format, with an emphasis on discoverability. GBFS is intended to make shared mobility information publicly available online. Therefore, any piece of information that can result in a suer being personally identifiable is not currently and will not become part of the core specification.

GBFS was created in 2014 to describe traveller facing information of a shared mobility system. As its name suggests, GBFS is inspired by GTFS, with a similar process for governance, similar openness to the industry needs and with the similar principle that any proposed change to the specification must be produced and consumed by the industry.

The standard has evolved (v2.2 by July 2021) and can now be used to describe all the new mobility services. Ten characteristics are represented:

- Stations
- station status
- vehicles status
- deep links (to book and pay directly into mobile app of the native operator)
- contact email (of the data producer)
- versioning
- vehicle types
- geofencing (the area in which one can use the shared mobility solution)
- virtual stations and valet services (limited areas on sidewalks where one can park the shared mobility vehicle without it being an actual station)

[26] Sharing data for shared micromobility, Survey report, jan.2021, POLIS Network https://www.polisnetwork.eu/wp-content/uploads/2021/01/SHARING-DATA-FROM-SHARED-MICROMOBILITY_FINAL.pdf
[27] https://github.com/NABSA/gbfs

Funded by the European Union

- pricing

The advantages and disadvantages of GBFS are the same as those of GTFS. Advantages of GBFS are its simplicity, its openness, and the fact that it prevents vendor lock-in. Disadvantages of GBFS are the counterpart of its simplicity, i.e. it has some limitations and cannot represent everything.

2.3.2. MDS (Mobility Data Specification)

The Mobility Data Specification (MDS)[28], a project of the Open Mobility Foundation (OMF), is a set of APIs focused on dockless e-scooters, bicycles, mopeds and carshare. The goals of MDS are to provide a standardised way for municipalities or other regulatory agencies to retrieve, compare and analyse data from mobility service providers, and to give municipalities the ability to express regulation in machine-readable formats.

MDS helps cities interact with companies who operate dockless scooters, bicycles, mopeds and carshare in the public right-of-way. MDS is a key piece of digital infrastructure that supports the effective implementation of mobility policies in cities around the world.

MDS is designed to be a modular kit-of-parts. Regulatory agencies can use the components of the API that are appropriate for their needs. An agency may choose to use only agency, provider, or policy. Or they may select specific elements (endpoints) from each to help them implement their goals.

2.3.3. The main differences between MDS and GBFS

GBFS is intended for public consumption through consumer-facing applications, while MDS is intended for use only by regulators. GBFS is governed by MobilityData, whereas MDS is managed by the Open Mobility Foundation. Both are membership-based organisations which run open processes for improving and extending their respective specifications.

Table 1: Differences between MDS and GBFS

	GBFS	MDS
Primary Purpose	Enable consumer-facing applications for finding and renting shared mobility vehicles from one or more operators.	Enable regulators to monitor how mobility vehicles are deployed and used (trips taken) for the purposes of planning, program management, and operations.

[28] https://github.com/openmobilityfoundation/mobility-data-specification

Funded by the European Union

Secondary purposes	- Limited vehicle information for regulators - Research by academics and advocacy organisations.	- Feed into open data sites (w/ privacy redactions) - Assist in overall analysis and management of public right-of-way
Access	Typically available to the public	Available only to regulators
Scope	Docked, dockless, and hybrid shared mobility vehicles	Dockless mobility vehicles (support for docked and hybrid planned)
Covered vehicles	Vehicles that are currently available for rent or disabled	All vehicles deployed in the public right of way (including unavailable and on-trip)
Format	Structured CSV and JSON	Authenticated JSON APIs
Governing organisation	MobilityData	Open Mobility Foundation

2.3.4. Applicability to MobiDataLab

Micro-mobility and shared mobility standards (see Figure 6) are obviously relevant for MobiDataLab as the challenges proposed to the Living Labs participants will be provided by referring municipalities who face concrete mobility problems and data sharing issues with these new mobility modes in a urban context.

Funded by the European Union

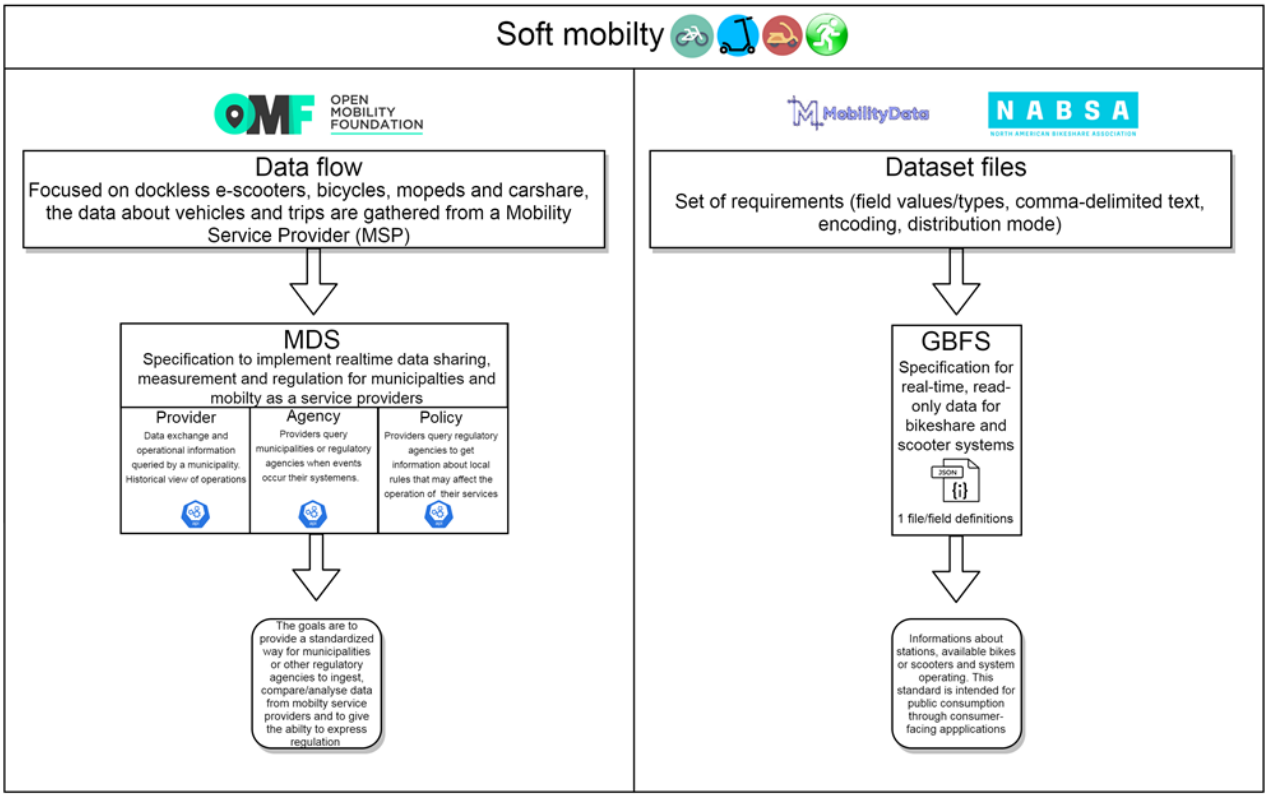

Figure 6: Panorama of standards for sharing micromobility data (F. Latrubesse, AKKA)

2.4. Ticketing data

Many public transport systems collect fares using proprietary contactless technologies. Potential disadvantages of such proprietary systems (like e.g. vendor lock-in) can be detrimental to operators, transport agencies and also to the end-user. On the contrary, solutions based on open standards can create a solid foundation for secure and efficient transport ticketing and multi-application systems. The most popular open standards are Calypso and CIPURSE™.

2.4.1. Calypso

Calypso is an international open ticketing standard, created by a group of transport operators from different countries in Europe: OTLIS-Lisbon, ACTV-Venice, STIB-Brussels, LKRKN-Constance and RATP & SNCF-Paris.[29]

[29] https://calypsonet.org/calypso-worldwide/

Funded by the European Union

Calypso is based on smartcard and contactless technologies, adapted to public transportation use. After a few years of trials, the system has been generalised in the early 2000s in major European cities. Calypso is extended now in other countries all over the world.

2.4.1.1. Use cases

A Calypso transport card typically takes the form of a smartcard, similar to the ones used for monetary transactions (as shown in Figure 7). It can also be hosted on any device with NFC (Near Field Communication) interface and computing capabilities, typically a smartphone, as long as it can store all information related to card owner and implement the Calypso authentication scheme for security.

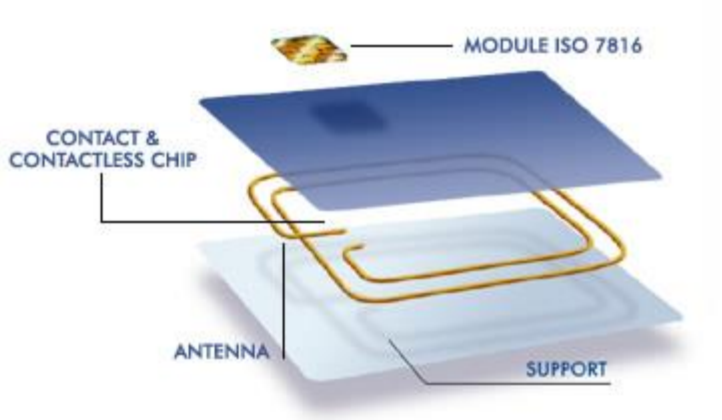

Figure 7: Calypso transport card (Calypso Handbook © 2010)

The two main uses of the Calypso transport card are:

- Reloading (one-way tickets, season tickets, etc) at vending machines, in shops, or at ticket offices in a contact and contactless way;
- Validating aboard any means of transportation – quickly and safely. The validators are used to validate entrance to (and optionally exit from) the network system.

In addition, the central system allows keeping track of the transactions, making statistics and verifying the system security and integrity.

The use of smartcards with microprocessor or other compatible devices is one of the main differences between Calypso and other e-ticketing systems where the card does not contain any information about owner and has no processing capabilities (for example London's Oyster card). This allows the Calypso standard to be more versatile and more secure.

2.4.1.2. Technology

Calypso open ticketing standard defines the secured dialogue between cards and terminal.

The technical elements constituting the Calypso specifications are:

- Calypso Secure Transaction
- Calypso generic Data Model
- Calypso generic Security Architecture
- Calypso generic API

At the base level, international standards are used to ensure transaction between card and terminal, as summarised in the following table (source: Calypso Handbook © 2010, Ticketing transaction layers):

Table 2: Summary of standards

	Layer	Standard
7	Security Management and Architecture	Calypso Security Architecture
6	Terminal Applicative Software	Calypso API
5	Data Model	Calypso Data Model
4	Card and SAM Security Mechanisms	Calypso card application
3	Card Data structure	CEN EN 1545
2	Card OS and Files structure & Commands	ISO/IEC 7816-4
1	Contactless Communication Interface and Contact Communication Interface	ISO/IEC 14443 ISO/IEC 7816-3

2.4.1.3. Calypso Networks Association (CNA)

Calypso Networks Association (CNA) is a not-for-profit organisation established in 2003 that regulates the use of Calypso technology, and delivers expert services, to promote open standards and enable their use across local, regional, national and global ticketing requirements[30].

2.4.2. CIPURSE

CIPURSE is an open security standard for transit fare collection systems. It makes use of smart card technologies and additional security measures.

[30] Sources: https://calypsonet.org , https://calypsostandard.net/ , http://cna.adcet.com/ , Calypso Handbook (available for download at http://cna.adcet.com/index.php/downloads/category/5-free-downloads?download=19:calypso-handbook)

The CIPURSE standard has been introduced in 2010 by founder members: chip supplier Infineon Technologies and smart card manufacturers Giesecke & Devrient GmbH (G&D) and Oberthur Technologies[31].

Version 2.0 of the CIPURSE Specification has been introduced in 2012. This version comprises a single, consistent set of specifications for all security, personalisation, administration and life-cycle management functions needed to create a broad range of interoperable transit applications, using various devices: not only smart cards but also NFC smartphones.

2.4.2.1. Use cases

CIPURSE standard aims at addressing

- public transport services
- collection of transport fares
- transactions related to micropayments.

Three subsets of the CIPURSE V2 standard have been defined by OSPT Alliance (see section 2.4.2.3) to address different use cases:

- CIPURSE T – Microprocessor-based transactions using smart cards, mobile phones and similar devices for more complex transit fare applications, such as monthly or annual tickets, multi-system tickets and loyalty programs.
- CIPURSE S – Tickets that can be recharged for a specific number of rides or weekly tickets
- CIPURSE L – Supports applications that use very inexpensive, disposable single-ride or daily tickets.

2.4.2.2. Technology

At the core level, CIPURSE is based on the same technical standards as Calypso:

- ISO/IEC 7816 smart card standard
- the ISO/IEC 14443 protocol layer

The CIPURSE standard also:

- Defines a secure messaging protocol;
- Identifies four minimum mandatory file types and a minimum mandatory command set to access these files;

[31] https://www.csoonline.com/article/2126207/card-makers-hope-to-shake-security-status-quo.html

Funded by the European Union

- Specifies encryption keys and access conditions;
- Is radio frequency (RF) layer agnostic;
- Includes personalisation and life cycle management, as well as system functionality to provide interoperability and fast adoption;
- Provides a security concept and guidelines.

2.4.2.3. OSPT Alliance

The OSPT Alliance is a non-profit industry organisation open to technology vendors, transit operators, government agencies, systems integrators, mobile device manufacturers, trusted service operators, consultants, industry associations and others wishing to participate in the organisation's education, marketing and technology development activities[32].

2.4.3. Collaboration between CNA and OSPT

In January 2020, CNA announced their collaboration with OSPT Alliance. Their announced end goal was to achieve convergence between their respective standards (Calypso and CIPURSE). However, as of now it seems their main output is a joint White Paper delivered in April 2021 exploring the benefits open ticketing standards can bring to account-based ticketing (ABT).

While Calypso is actively used by many operators over the world, there is only few transport operators that have adopted CIPURSE: city of Cuenca in Ecuador (transport ticketing), Korean airports (access control), Sao Paulo (highway fare collection). CIPURSE is also used for driver ID verification in Brazil, but it is no more related to ticketing[33] [34].

2.4.4. Ticketing data sharing

Both Calypso and CIPURSE define standards for electronic ticketing. However, if they define precisely the interactions between the card and the terminal they do not provide a public standard allowing data sharing. Ticketing data is available to transport authorities through API calls but it is up to the transport authorities or to integrators to define how ticketing will be structured and stored, once the raw data is retrieved from ticketing system. Such structure could be provided by using OpRa standard (see 2.1.4.2), when it will be available, since incoming funds (sold tickets) are mentioned in OpRa use cases.

---

[32] https://www.osptalliance.org/
[33] https://ticketingopenstandards.org/
[34] https://www.osptalliance.org/cipurse-in-action/

Funded by the European Union

Since Calypso and CIPURSE standards do not define a way to structure ticketing data for transport operators, they cannot be used to promote data sharing between these operators.

As of now, it is up to transport operators to make use of ticketing events to construct data sets, using proprietary formats or standard formats, for example OpRa when it will be ready. These data sets could then be used for static analysis and prediction models. For example, analysis of ticketing events can help an operator to understand and predict the flow of travellers within a particular line or day.

2.4.5. Applicability to MobiDataLab

Regarding the interest of ticketing standards for data sharing in general and for the MobiDataLab project in particular, the consortium still needs to explore how these can be used. The joint collaboration between OSPT and Calypso is promising but does not yet seem to be sufficiently advanced. However, it could converge in the next two years, before the MobiDataLab living labs take place. The MobiDataLab consortium will therefore follow this initiative closely.

Regarding how MobiDataLab would use these standards, HOVE is using the Spirtech SDK (Calypso standard) in a mobile application for the municipality of Lille in France. But even though it is possible to retrieve ticketing events, it does not seem possible to share ticketing data using this standardised SDK.

3. Data sharing standards for cross-domain interoperability

Cross-domain interoperability exists when systems from different domains interact in the exchange of information. In the present case, we consider the transport sector as a domain with its own standards (or *vertical* standards), as we have seen in the previous section. But the transport sector also follows general standards (or *horizontal* standards): for example data exchange follows web protocols (for which there are best practices), data sharing solutions aim to be deployed in the cloud – which has its own standards (for exchanging files, federating services, authenticating and controlling access) –, and mobility data are often geospatial data located in space – for the exchange of which standards are available (namely the Geographical Information Systems standards).

3.1. Best practices for sharing data on the web

3.1.1. Context

The publication of data on the web must respect best practices, which have been recommended by the World Wide Web Consortium (W3C)[35]. The data on the web best practices is an initiative from a dedicated W3C working group whose mission is:

- To develop the open data ecosystem, facilitating better communication between developers and publishers,
- To provide guidance to publishers that will improve consistency in the way data is managed, promoting the re-use of data, and
- To foster trust in the data among developers, whatever technology they choose to use, increasing the innovation potential.

Looking at these objectives, one can see that in many respects the mission of the MobiDataLab consortium is comparable to that of the Data on the web working group, except of course that it focuses on mobility data.

First of all, a distinction should be made between the two types of actors involved when publishing data on the web:

- Data publisher: publishes and shares the data
- Data consumer: reuses the data and might generate new data

A common understanding between data publishers and data consumers is fundamental. Without this agreement data publishers' efforts may be incompatible with data consumers' desires.

In particular, publishing data must follow web architecture principles, using vocabularies and standards, whether they concern the dataset itself and its metadata (as shown in Figure 1Figure 8).

[35] https://www.w3.org/TR/dwbp/

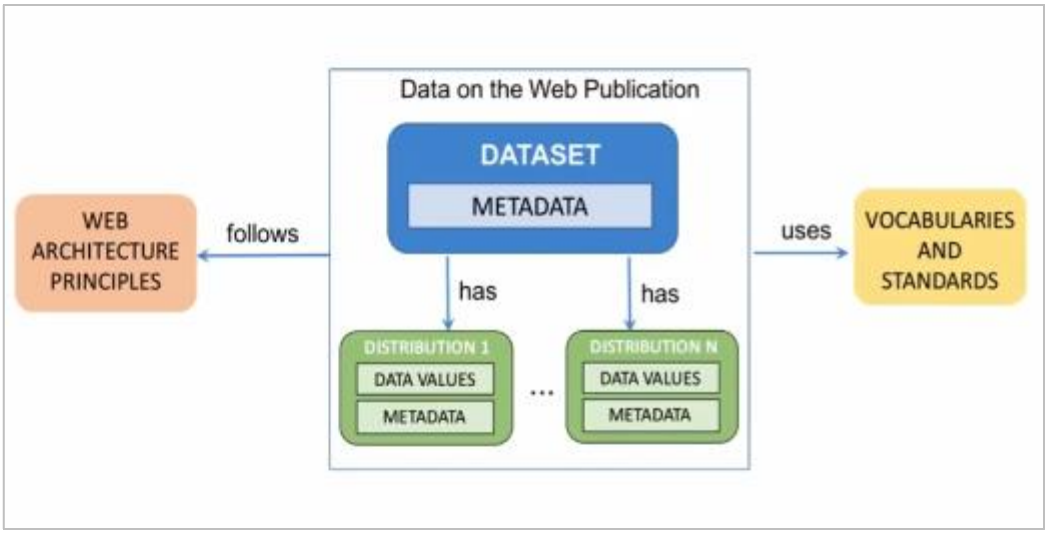

Figure 8: Principles for publishing Data on the web (C. Burle, W3C)

3.1.2. Challenges and requirements

Publishing data on the web is more than just posting data, and the W3C working group has identified 12 challenges (and 42 requirements) when it comes to publish data on the web[36]:

- Metadata (for humans and machines)
- Data licenses (how to permit and restrict access)
- Data provenance and quality (how to add trust)
- Data versioning (tracking dataset versions)
- Data identification (identifying datasets and distributions)
- Data formats (which data formats to use)
- Data vocabularies (how to promote interoperability)
- Data access (access options)
- Data preservation
- Feedback (how to engage re-users)
- Data enrichment (adding value to data)
- Data republication (reuse data responsibly)

All these requirements will be duly considered during the implementation of the MobiDataLab Transport Cloud, i.e. the prototype platform for using and fusing mobility data on the web (see WP4).

[36] https://www.w3.org/TR/dwbp/#challenges

Funded by the European Union

3.1.3. Metadata on the web

Metadata (a word composed of the Greek prefix meta, indicating self-reference; the word therefore properly means "data of/about data") is data used to define or describe another piece of data. Metadata provides additional information that helps data consumers better understand the meaning of data, its structure, the rights and license terms, the organisation that generated the data, data quality, etc. Two standards for metadata exchange are considered: DCAT and CSW.

3.1.3.1. DCAT ontology

The DCAT Application Profile for data portals in Europe (DCAT-AP) is a specification based on the Data Catalogue Vocabulary (DCAT) developed by W3C[37]. This application profile is a specification for metadata records to meet the specific application needs of data portals in Europe while providing semantic interoperability with other applications on the basis of reuse of established controlled vocabularies (e.g. EuroVoc) and mappings to existing metadata vocabularies (e.g. Dublin Core, SDMX, INSPIRE metadata, etc.). Also, DCAT-AP provides a common specification for describing public sector datasets in Europe to enable the exchange of descriptions of datasets among data portals. DCAT-AP allows:

- Data catalogues to describe their dataset collections using a standardised description, while keeping their own system for documenting and storing them;
- Content aggregators, such as the European Data Portal, to aggregate such descriptions into a single point of access;
- Data consumers to more easily find datasets through a single point of access.

DCAT-AP has an extension GeoDCAT-AP for describing geospatial datasets, dataset series and services. Another extension, StatDCAT-AP, provides specifications and tools that enhance interoperability between descriptions of statistical data sets within the statistical domain and between statistical data and open data portals.

One of the main benefits of DCAT-AP is ensuring consistency by providing a standard for the description of metadata which is published by data portals across Europe. Also, DCAT-AP assists data re-users and data providers with the following activities:

- Data re-users can get an overview of which datasets exist and which public administrations are maintaining it, in particular if the datasets are in another Member State enabling multilingual access to information. For this, data publishers and portals maintain catalogues of datasets that are made available by public administrations on their websites. The quality of the description metadata in these catalogues directly affects how easily datasets can be found.

[37] https://www.w3.org/TR/vocab-dcat-2/

Funded by the European Union

- Data providers encourage reuse of their datasets by making them searchable and accessible by listing it on one or more data portals which can significantly reduce costs.

3.1.3.2. Geospatial metadata standard (CSW)

The Catalogue Services specification is a standard maintained by the Open Geospatial Consortium (OGC). The services support the ability to publish and search collections of descriptive information for data, services, and related information objects. OGC Catalogue Services[38] support the ability to publish and search collections of descriptive information (metadata records) for geospatial data, services, and related resources. Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community.

This standard specified three classes of service operations (as stated in the Catalogue Services Specification[39]):

- OGC_Service: these class operations are the standard set of operations that every OGC service must implement in order to work with OGC Catalogue Services. The operations allow the service to provide metadata about itself via a standard service description document and also a means of retrieving catalogues records based on their unique identifier.
- Discovery: these class operations allow the service to be interrogated to retrieve run time information about the data offered by the service as well as providing means of retrieving records from the catalogue using a general predicate language to define constraints the define the subset of records to be retrieved.
- Manager: these class operations allow records to be added, modified and removed from the catalogue service.

Based on these class operations, the Catalogue Service for the Web (CSW) standard defines seven operations[40]:

- GetCapabilities (OGC_Service)
- GetRecordById (OGC_Service)
- GetDomain (Discovery)
- GetRecords (Discovery)
- Transaction (Manager)
- Harvest (Manager)
- UnHarvest (Manager)

[38] http://opengeospatial.github.io/e-learning/cat/text/index.html
[39] https://docs.opengeospatial.org/is/12-176r7/12-176r7.html
[40] https://docs.opengeospatial.org/is/12-176r7/12-176r7.html#37

Funded by the European Union

There are two equivalent encodings defined in this standard for most operations: a key-value pair (KVP) encoding suitable for use with the HTTP GET method, and an XML encoding suitable for use with the HTTP POST method. Furthermore, the specification contains discussions about how to use the XML-encoded operations with SOAP (Simple Object Access Protocol) with this standard.

The OGC CSW is available in version 3.0. The standard supports a specification of metadata for geographic data sets and enables further tooling, such as a query language to retrieve filtered information on the listed data sets. The metadata is coded following ISO 19115[41] for spatial applications and data and ISO 19119[42] for geo services[43].

Implementations:

- Pycsw 2.6.0 (reference implementation)
- CSW 2.0.2

3.1.4. Applicability to MobiDataLab

Following the best practices in data sharing on the web is paramount for the MobiDataLab project, in order to comply with the FAIR principles[44] and to make the data available on the platform more findable, accessible, interoperable and reusable.

Regarding the metadata, MobiDataLab will comply with the DCAT-AP standard and will follow the new versions of the standard with particular attention. Also MobiDataLab could include the CSW standards, based on AKKA's and HERE's previous experience with GeoNetwork[45].

3.2. Standards for data exchange in the cloud

3.2.1. Exchanging files

There are essentially two ways to exchange files between multiple cloud data centres: by using data storage services as intermediate data storages (1), or by employing direct cloud-to-cloud communication at application level (2).

41 https://www.iso.org/standard/26020.html/
42 https://www.iso.org/standard/59221.html/
43 http://docs.opengeospatial.org/is/12-168r6/12-168r6.html
44 The FAIR principles aim to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets.
45 https://geonetwork-opensource.org/

3.2.1.1. Using intermediate data storage services

In the first case, several private cloud solutions provide efficient ways to store data with 'pay-as-you-go' charging models. One of the most popular is the Amazon S3 (Simple Storage Service) service. Data stored in S3 is organised over a two-level hierarchy. At the top level are buckets, which have a unique global name and work similar to “folders” in a classical filesystem. Each bucket can store an unlimited number of data objects, and users can create, modify and read objects in buckets. Search is limited to single buckets and by using objects' names. No metadata or content-based search is provided. Therefore, for data exchange, the content to be transferred has to be known a priori by the two cloud data centres involved. S3 provides different accessing policies (with different prices) to data access. The Standard policy is for data objects that are frequently accessed and provide high availability. The Infrequent Access policy is for those data objects that are accessed rarely, but still performance is needed when they are. Finally, the Glacier policy is an extremely low-cost and durable data storage, which can require hours to access data.

Similar to Amazon’s S3, other private cloud services offer data storage. Microsoft Azure Blob can store both structured and unstructured data. Blob has two storage policies: the hot storage policy is for storing frequently accessed data, with high durability, availability and quick access time. The cool storage policy is for infrequent data access and focus on durability. Other similar storages are Google Cloud Storage and IBM Cloud Object Storage.

All the services mentioned above are private cloud services. There are alternatives that provide the same functionalities, but they require an installation to be operated. One of the most popular is Minio. Minio is an S3 open source alternative written in Go. Minio is an object storage server compatible with Amazon S3 and licensed under Apache 2.0 License. In MobiDataLab Minio can be useful if the different cloud datacentre has the possibility to install and run software.

3.2.1.2. Using direct cloud-to-cloud communication at application level

In the second case, the exchange of data between various cloud computing domains follows the same standard that governs the transferring of data over Internet entities. At the transport level, data is commonly transported by using the TCP/IP protocol. On top of this, application-specific implementation can be built. A common strategy is to define applications that talk to each other using web services standards built over the HTTP application layer.

Among these technologies, the most popular is the REpresentational State Transfer (REST). Although not strictly a standard (rather, an architectural style) is the communication technology most used when providing a general interface over the Internet. REST uses a subset of HTTP to expose web resources in a textual representation allowing them to be read and modified with a stateless protocol and a predefined set of operations. This approach allows interoperability between the computer systems on the Internet that provide these services.

Funded by the European Union

REST has the following six formal constraints[46]:

- It works in a client-server architecture (as the HTTP);
- It is stateless, as there is no concept of "session" and each REST call is independent of the other;
- Cacheability: the response to a client can be cached for performance reasons;
- It is a layered system, i.e. If a load balancer is between the client and server, there is no need to update the client or server code for REST to work;
- Servers can temporarily extend or customise the functionality of a client by transferring executable code (e.g. JavaScript);
- REST interfaces should follow certain principles of uniformity in their design.

The payload of the HTTP requests can be used to transfer data and information from one computing space to another. The payload can take different standard formats, of which the most common are: XML, YAML, CSV, JSON.

XML

The Extensible Markup Language (XML) is a markup language that defines a set of syntactic rules to define content that is human - and machine – readable. XML is widely used for the representation of arbitrary data structures in many contexts, including web services and to export data structure in machine-independent and language-independent fashion. The standard XML 1.0 was defined by the W3C and last revised on November 2008[47]. An XML schema is a description of an XML document. The schema is expressed in terms of constraints on the structure and content, but still complying with the basic syntactic constraints imposed by regular XML syntax. These constraints are generally expressed using some combination of grammatical rules governing the order of elements. Typically, each application defines its own schema.

YAML

YAML (a recursive acronym for "YAML Ain't Markup Language"[48],) is a human-readable data serialisation standard. Compared to XML, YAML does not use "tags", but it is used for similar application purposes. For example, it is commonly used for configuration files, as it allows the definition of key-value pairs of properties in a style similar to Python. Custom data types are allowed, but YAML natively encodes scalars (such as strings, integers, and floats), lists, and associative arrays (also known as maps or dictionaries). Lists and hashes can contain nested lists and hashes, forming a tree structure.

CSV

The comma-separated values (CSV) format is widely used to define structured information in a file. Each file is divided into a tabular fashion that includes rows and columns. As the name suggests,

---

[46] https://dl.acm.org/doi/10.1145/2464526.2464551
[47] https://www.w3.org/TR/xml/
[48] https://yaml.org/

Funded by the European Union

fields on each row (a record) are separated by a comma. However, the CSV file format is not fully standardised. The basic idea of separating fields with a comma clashed when the field values may also contain commas. Various CSV implementations may not handle such field data and may include escape characters or escape sequences. RFC 4180 proposes a specification for the CSV format, but the term CSV is typically used in generic delimiter-separated formats that use different field delimiters, for example, semicolons. This loose terminology can cause problems in data exchange between programs. Therefore, many applications supporting CSV files allow users to preview the first few lines of the file and then specify the delimiter character(s) or quoting rules.

JSON

JavaScript Object Notation (JSON) is a language-independent open-standard (ISO/IEC 21778:2017) format for data exchange that uses human-readable text to transmit data objects. JSON is built upon key-value pairs and arrays. It is a very common data format, with a diverse range of applications, and is more and more used as a replacement for XML.

3.2.2. TOSCA as an application standard for Cloud federation

The rise of clouds and federation of clouds in these past years has been huge. Such diffusion has been supported by several technology developments in numerous sectors. As a matter of fact, elasticity and flexibility, together with the ‘pay-as-you-grow’ principle, provided by cloud computing (due to its utility computing nature) have made it very attractive and have provided a wide range of use cases for all industries. This, in turn, has led to the development of brand-new technologies that by exploiting cloud computing, revolutionised the paradigm under which some services are provided and managed. Notable examples of such revolution are NFV (Network Function Virtualisation) and SDN (Software Defined Network), which are seen as the next step in network development and solutions for service providers to be more efficient and deliver new services faster.

Unfortunately, a lack of standardisation in the way cloud services are managed and are operating, the limited portability of cloud services and some potential issues related to cloud security, can potentially slow down further development. Moreover, such limitations can also hinder the viability, effectiveness and opportunity to define and build a federation of clouds.

The ability to migrate cloud services between different cloud platforms and even providers, a standardisation process conditioning service definitions, as well as the ability of being vendor independent (i.e., avoiding the vendor lock-in issue) can significantly help technologies based on cloud computing reach the next level.

TOSCA, an open standard by OASIS (Topology and Orchestration Specification for Cloud Applications), TOSCA helps providers do just this. It can model networks in a standardised manner, improve automation, enable portability, and more easily overcome interoperability issues. TOSCA aimed at enhancing the portability and management of cloud applications and services across their entire lifecycle.

Its actual purpose is to describe the components of an application and their interactions. In TOSCA, services and applications to be hosted on the cloud are described, including their components, relationships, dependencies, requirements, and capabilities. As such, the goals of TOSCA are:

- the automation of application deployment;
- the representation of applications in a cloud agnostic way;
- the inter-operability and re-usability of components.

Cloud federations can benefit from the usage of TOSCA as an application model which contains a generic description which then has to be converted to a deployment plan; a similar approach has been introduced before in BASMATI[49], CloudCAMP and Tosker[50].

Tosker parses an extended TOSCA YAML which describes multi-component applications for a deployment plan for Docker. TOSCA has also been used with Kubernetes to define application components along with their deployment and run-time adaptation on Kubernetes clusters across different countries, which is a case very similar with the one motivating BASMATI project.

Its adoption in the context of MobiDataLab can support the definition, translation and deployment of workflows with an extended TOSCA grammar; this idea has also been used to let Cloudify extend TOSCA grammar to describe workflows. TOSCA and its extensions have been and are currently used by a number of research projects and computing platform.

3.2.3. Identity and Access Control

Secure exchange of data requires proper management of identities, authentication and permission management. We next describe standards for the support of (federated) identity management, authentication and authorisation, role-based access control and the Amazon implementation for authorisation and access control.

3.2.3.1. SAML

The Security Access Markup Language (SAML)[51] is an open standard by OASIS that defines an XML-based framework for describing and exchanging security information between on-line business partners. The main use cases for SAML are single sign-on (SSO) and Federated Identity. SAML supports both authentication and authorisation.

In the SSO scenario, we find three entities, the user agent, the service provider (SP), which is the application to access and finally the identity provider (IdP). When configuring SAML federation, a trust relationship is established between the IdP and the SP. This trust is established via the SAML metadata. The metadata file contains accepted user identifiers and its allowable formats, along with the SAML bindings (the technical means by which the user, the IdP and the SP will exchange

[49] http://hpc.isti.cnr.it/basmati/
[50] https://github.com/di-unipi-socc/TosKer
[51] Security Assertion Markup Language (SAML) V2.0 Technical Overview – http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html

information), the acceptable authentication mechanisms, and the entities' certificates and a signature.

There are two main scenarios for authentication and authorisation: IdP-initiated and SP-initiated. In the IdP-initiated scenario, the user starts by accessing the IdP. The user is, then prompted for authentication and once done, the user can request a service. If the user is authorised, the IdP generates a SAML assertion, which contains the IdP identification, the user identification, the method of authentication (password, certificate, 2FA, MFA, etc.) and a signature by the IdP. Through the user agent, the assertion is sent to the SP using a POST message. It is the user agent who acts as the transport mechanism for the assertion. The SP verifies the assertion, maps it to a local user and then the session can start. In the SP-initiated scenario, the user starts by reaching out to the SP. Since the user is not authenticated, the SP redirects the user to the IdP using a request for authentication message. Once the user is validated, the IdP generates the SAML assertion. The assertion is sent to the SP via the user agent or an SAML assertion and the session can start.

3.2.3.2. OAuth 2.0 and OIDC

The OAuth 2.0[52] authorisation framework enables third-party applications to obtain limited access to an HTTP service, either on behalf of a resource owner by orchestrating an approval interaction between the resource owner and the HTTP service, or by allowing the third-party application to obtain access on its own behalf. In OAuth 2.0 we have four entities: a resource owner, who is the final user, a client application to which we want to grant authorisation to, an authorisation server that manages authorisation tokens and workflow, and a resource server where the resource owner's data is located.

When a native client application requires access to a resource server, it is first redirected to the authorisation server, where the resource owner is required to authenticate. The authentication mechanism is not defined by OAuth 2.0, but it can be based, for example, on SAML. Once the resource owner is authenticated, the client application receives an access token, which will be included in further requests from the client application to the resource server.

An alternative scenario is when the client application is not native but is hosted in a different application server. First, the application server must register to the authorisation server with an application name and a callback URL and receives a unique ID and a secret token. When a resource owner accesses the application (which requires data from the resource server), the application server sends a request to the authorisation server with the application ID and the resource owner ID. Then, the resource owner is required to authenticate to the authorisation server. Again, the OAuth 2.0 standard does not define any methods for authentication. Then the authorisation server returns an authorisation code to the application server (through the resource owner), which the application server uses to obtain an access token directly from the authorisation server. The access token can then be used to access the resource server on behalf of the resource owner. Note that the access

[52] OAuth 2.0, RFC6749 – https://datatracker.ietf.org/doc/html/rfc6749

tokens can be limited in the scope, that is the resources and actions that the application server is allowed to perform.

The OpenID Connect (OIDC)[53] offers an identity and authentication layer on top of OAuth 2.0. OIDC provides authentication mechanisms to OAuth 2.0, and allows third-party applications to request and receive information about the authenticated end-users.

3.2.3.3. Amazon Web Services (AWS) Identity and Access Management (IAM)

The AWS IAM[54] enables the management of access and resources for AWS services. IAM allows the creation of users, groups, roles and their permissions, using a proprietary policy language.

The access control policies in AWS IAM define *Principals*, which represent individual users, groups of users or roles, *Actions*, such as read or write, the managed *Resources*, and the access *Conditions*, which might include restrictions on users, groups, roles, kinds of resources, time, source IP, kind of authentication, etc. After a user requests an action on a resource, the AWS services populate the request with the context information, including the source IP, the groups or roles that the user belongs to, the authentication method, etc. and matches this context with the available policies. If no explicit *Allow* is found, the request is *Denied*.

AWS services provide their own authentication and authorisation mechanisms, including services for SSO and Federated Identity, and can be integrated with services compatible with SAML 2.0.

3.2.4. Applicability to MobiDataLab

The standards for data exchange in the cloud are relevant for the Architecture definition of the MobiDataLab Transport Cloud. In particular the TOSCA standard can be used to drive the definition and structuring of application that will need to process the mobility data hosted into the Transport Cloud.

As to how MobiDataLab would use these standards, not only will TOSCA be taken into account, but as it is sometimes complex to implement, Kubernetes will also be considered, at least in the API, as it is a very well designed platform that the CNR uses it in various research projects. Even though Kubernetes is a technology, it is becoming a standard.

[53] OpenID Connect – https://openid.net/connect/
[54] AWS IAM – https://aws.amazon.com/iam/

3.3. Geodata sharing standards

Mobility data are often geospatial data, where accurate location can be critical. These geospatial data can be shared according to common geographical standards. Common geographical standards relate to the data models for structuring the information (for instance raster or vector data, OpenStreetMap (OSM) data or data models defined by the INSPIRE directive), and the exchange services themselves (for exchanging raster or vector data, object location or features).

3.3.1. Geo-referenced Data Models

In order to understand the mechanisms by which location-based information can be exchanged, it is necessary to understand how the underlying information is structured. It is all the more necessary that this structuring information is itself standardised, via e.g. the raster and vector GIS data formats, the de-facto OSM standard and the INSPIRE environmental data models.

3.3.1.1. Geographical Information Systems (GIS) Data Models

There are two main data models in GIS – the Raster and the Vector data models. Each one of these data models has its own specific features and we should determine whether the Raster or Vector data model is best suited for our data and/or analytical needs.

Raster data models

Raster data are derived from a grid-based system of contiguous cells containing specific attribute information. The spatial resolution of a raster dataset represents a measure of the accuracy or detail of the displayed information. The Raster data model is widely used by non-GIS technologies such as digital cameras/pictures and LCD monitors.

Vector data models

Vector data utilises points, lines, and polygons to represent the spatial features in a map. Topology is an informative geospatial property that describes the connectivity, area definition, and contiguity of interrelated points, lines, and polygon. Vector data may or may not be topologically explicit, depending on the file's data structure.

3.3.1.2. OpenStreetMap Data Model

OpenStreetMap[55] is a crowd sourcing project for geographic map data. The project has been funded in 2004.

The data model consists of three object types (Figure 9 shows their relationships):

- nodes, which consists the geometry in spatial lat/lon coordinates in EPSG:4326;
- ways, which references to an ordered list of nodes and represents linestring geometries and polygons (linestrings where first and last point are the same reference) and
- relations, which can consist of references of each of the three objects.

Each object type can have attached tags (key/value pairs) that contains further semantic information about the object. The tagging schema underlies continuous evolution and is agreed/steered by the community. The data itself is available as XML or in a binary encoded ProtoBuf format.

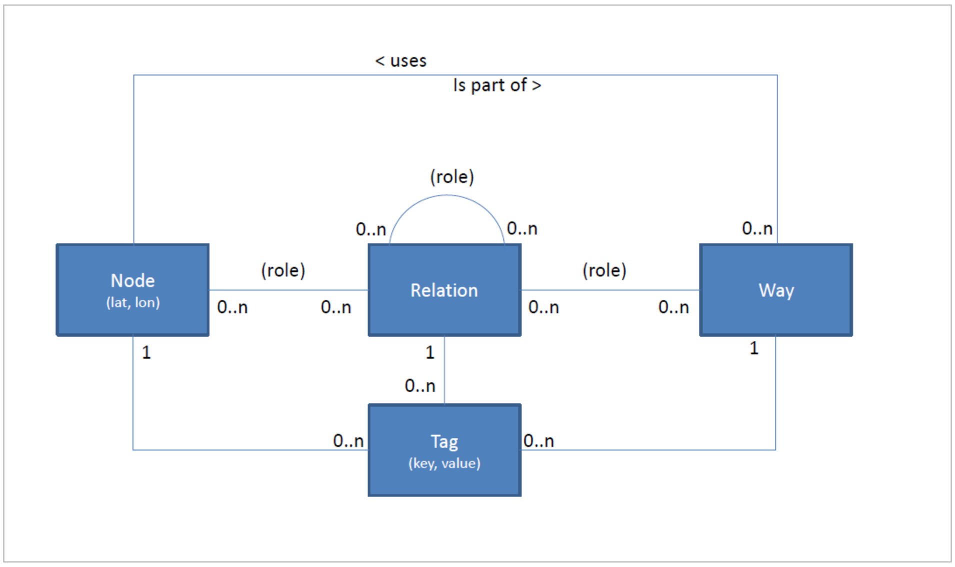

Figure 9: OSM Data Model (source: Lauer 2018)

3.3.1.3. INSPIRE Data models

The Infrastructure for Spatial Information in the European Community INSPIRE Directive (2007/2/CE[56]) laid the foundations for cross-border interoperability of environmental services and

[55] https://www.openstreetmap.org/
[56] https://inspire.ec.europa.eu/

data between EU member states. As such, the work that has been carried out in this framework is pioneering. This is why, although primarily concerned with environmental data, the INSPIRE Directive is now considered to be a model for interoperability in several sectors. This is particularly the case in transport, where a joint initiative has made it possible to coordinate the work of INSPIRE and Transmodel, as described in the following section.

The INSPIRE Directive

The INSPIRE Directive aims to create EU spatial data infrastructure for EU environmental policies and policies or activities that may have an impact on the environment. The Directive places requirements on public bodies that produce, receive, manage or update spatial datasets that cover all of the land and marine areas over which the State has to create such EU-related spatial data infrastructure. This European spatial data infrastructure will:

- enable public sector organisations to share environmental spatial information;
- facilitate public access to spatial information across Europe;
- assist in policy-making across boundaries.

INSPIRE is based on the infrastructures for spatial information established and operated by EU member states. The Directive has 34 spatial data themes[57] under which datasets are categorised.

INSPIRE support to Multi-Modal Travel Information Services

Interestingly enough, a recent joint initiative aimed to identify and analyse the overlaps and gaps existing between INSPIRE and the Transport-related domain[58].

Indeed the Delegated Regulation on EU-wide Multi-Modal Travel Information Services (MMTIS) requires that *"Transport authorities, transport operators, infrastructure managers or transport on demand service providers shall provide the static travel and traffic data and historic traffic data of the different transport modes, by using for the spatial network the requirements defined in Article 7 of Directive 2007/2/EU (INSPIRE)"*. (Article 4(1)(c))

Thanks to this work we now know how INSPIRE relates with standards used in the transport domain:

- Transmodel/NeTEx for public transport;
- TAP-TSI and RINF for rail;
- IATA for air and DATEX for road and identify the possible overlaps.

3.3.2. Standards for exchanging Geodata

[57] https://inspire-geoportal.ec.europa.eu/theme_selection.html?view=qsTheme
[58] https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/inspire-support-multi-modal-travel-information-services

Besides geo-referenced data models, the services to exchange the corresponding data are also standardised. In general, two types of services are considered: services for exchanging data in the raster format (such as the Web Map Service) and services for exchanging data in the vector format and the corresponding attributes (such as the Web Feature Service and its successor the Features API).

3.3.2.1. WMS and WMTS

WMS (Web Map Service)

The OpenGIS® Web Map Service Interface Standard (WMS)[59] provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases. A WMS request defines the geographic layer(s) and area of interest to be processed. The response to the request is one or more geo-registered map images (returned as JPEG, PNG, etc) that can be displayed in a browser application. The interface also supports the ability to specify whether the returned images should be transparent so that layers from multiple servers can be combined or not.

WMTS (Web Map Tile Service)

The Open Geospatial Consortium, Inc. (OGC), Web Map Tile Service (WMTS)[60] specification is an international specification for serving digital maps over the web using cached image tiles. WMTS services are useful when cached map or image services need to be made available in an open, recognised way across different platforms and clients.

Clients built to support the WMTS 1.0.0 specification and RESTful or KVP encoding can view and work with WMTS services. SOAP encoding is not supported. Client applications work with a WMTS service by appending parameters to the service's URL. WMTS services derived from cached map or image services support the following operations (RESTful and KVP encoding only):

- Requesting metadata about the service (GetCapabilities);
- Requesting an individual tile resource in the cache (GetTile).

3.3.2.2. WFS (Web Feature Service)

The Web Feature Service (WFS)[61] represents a change in the way geographic information is created, modified and exchanged on the Internet. Rather than sharing geographic information at the file level using File Transfer Protocol (FTP), for example, the WFS offers direct fine-grained access to geographic information at the feature and feature property level.

---

[59] http://opengeospatial.github.io/e-learning/wms/text/basic-index.html
[60] http://opengeospatial.github.io/e-learning/wmts/text/index.html
[61] http://opengeospatial.github.io/e-learning/wfs/text/basic-index.html

Funded by the European Union

This standard specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored, parameterised query expressions. Discovery operations allow the service to be interrogated to determine its capabilities and to retrieve the application schema that defines the feature types that the service offers. Query operations allow features or values of feature properties to be retrieved from the underlying data store based upon constraints, defined by the client, on feature properties. Locking operations allow exclusive access to features for the purpose of modifying or deleting features. Transaction operations allow features to be created, changed, replaced and deleted from the underlying data store. Stored query operations allow clients to create, drop, list and described parameterised query expressions that are stored by the server and can be repeatedly invoked using different parameter values.

This International Standard defines eleven operations[62]:

- GetCapabilities (discovery operation)
- DescribeFeatureType (discovery operation)
- GetPropertyValue (query operation)
- GetFeature (query operation)
- GetFeatureWithLock (query & locking operation)
- LockFeature (locking operation)
- Transaction (transaction operation)
- CreateStoredQuery (stored query operation)
- DropStoredQuery (stored query operation)
- ListStoredQueries (stored query operation)
- DescribeStoredQueries (stored query operation)

In the taxonomy of services defined in ISO 19119, the WFS is primarily a feature access service but also includes elements of a feature type service, a coordinate conversion/transformation service and geographic format conversion service.

3.3.2.3. OGC API Features

OGC API standards[63] define modular API building blocks to spatially enable Web APIs in a consistent way. OpenAPI is used to define the reusable API building blocks with responses in JSON and HTML. OGC API Features provides access to collections of geospatial data.

OGC API allows to:

---

62 https://www.ogc.org/standards/wfs
63 http://opengeospatial.github.io/e-learning/ogcapi-features/text/basic-index.html

- List the collections of data on the server that can be queried, and each describes basic information about the geospatial data collection, like its id and description, as well as the spatial and temporal extents of all the data contained. The response format (typically HTML or a GeoJSON feature collection, but GML is supported, too, and extensions can easily supply others) is determined using HTTP content negotiation. Data is returned in pageable chunks, with each response containing a next link as many collections are quite large. The core specification supports a few basic filters, in addition to bounding box filter, with extensions providing more advanced options.
- Return a single 'feature' - something in the real-world (a building, a stream, a county, etc.) that typically is described by a geometry plus other properties. This provides a stable, canonical URL to link to the 'thing'.

3.3.3. Applicability to MobiDataLab

Geodata sharing standards are of particular interest for MobiDataLab because several use cases depend on map data, needing road names, land use patterns, traffic signs, bus routes, all being available in geodata formats.

On how to use these standards, the MobiDataLab consortium will re-use extensively the OSM data and the underlying de-facto standard – which provides a different understanding of maps not following the common GIS approach. In addition, the project will include the HERE APIs providing data in several geospatial formats (ESRI shapefile, GeoJSON, etc.) e.g. for the automotive sector in particular. This is shown in Figure 10.

Regarding how MobiDataLab will follow the evolution of these standards, the consortium will draw on its expertise with these standards, mainly thanks to AKKA’s experience with the INSPIRE Directive, and HERE’s membership in the Open Geospatial Consortium (OGC). Among the developments in the OGC standards to be followed in particular there is currently a shift from the old XML to the new Features API and the MobiDataLab timeline will be in the middle of this standardisation process. Also, the INSPIRE-MMTIS joint collaboration will be carefully followed.

Funded by the European Union

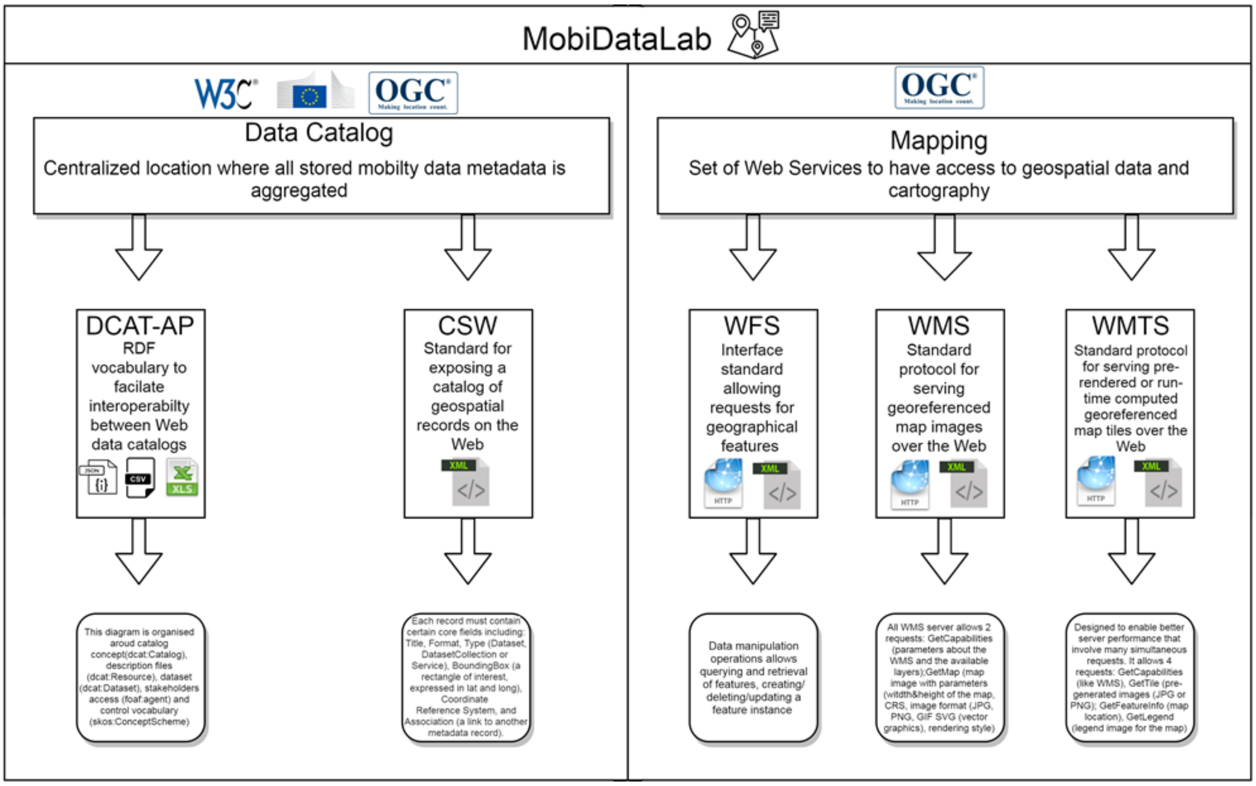

Figure 10: Panorama of standards for sharing data and geodata on the web (F. Latrubesse, AKKA)

3.4. Semantic interoperability standards

Semantic interoperability is a way of ensuring interoperability based on a common and unambiguous understanding of the information exchanged. It is achieved by means of enabling applications to interpret the meaning (semantics) of the exchanged data. The introduction of the Semantic Web concept by Tim Berners-Lee in 1999 has led the W3C (World Wide Web Consortium) to provide standards for web-scale semantic data exchange and inferencing capabilities. Such an interoperability offers very interesting perspectives for exchanging data within the transport domain (between different types of transport), but also with other domains that use similar concepts but different vocabularies (e.g. routes/itineraries are used by both tourism and transport applications).

3.4.1. Semantic web technologies

Semantic interoperability allows systems to exchange and federate data with universal meaning. Thus, interoperability specification describes how two systems or components can engage into a working interaction. Semantic interoperability focuses on describing the semantics of such interaction [Hamza et al 2019]. The most used are: semantic graphs, ontologies, Linked Open Data (which is built upon standards such as HTTP, RDF, and URIs), knowledge graphs and JSON-LD.

Funded by the European Union

Semantic Graphs are networks of semantic relationships within a database. More precisely, they define relationships and contexts between *concepts.* Relationships are stored within the graph by using suitable data representations.[64]

Ontologies. An ontology is a formalisation of a concept specifying the components such as individuals (instances of objects), classes, attributes, and relations, as well as restrictions, rules and axioms. As a result, ontologies do not only introduce a sharable and reusable knowledge representation but can also add new knowledge about the domain. An ontology can be created using computer description languages such as RDF (Resource Description Framework), RDFS (Resource Description Framework Schema) or OWL (Web Ontology Language).

Resource Description Framework (RDF) is a standard for data interchange that represents prepositional data in *triples*. Triples are *assertions* about a fact and can be used to represent highly interconnected data. Each RDF statement is a three-part structure consisting of resources where every resource is identified by a URI. Representing data in RDF allows information to be easily identified, disambiguated and interconnected[65]. SPARQL is the W3C-standardised query language for retrieving and manipulating data stored in RDF format. RDF is the format used in Semantic Web and Linked Open Data.

Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. When combined with Open Data (data that can be freely used and distributed), it is called **Linked Open Data (LOD)**. Linked Data is one of the core pillars of the Semantic Web. The Semantic Web is about making links between datasets that are understandable to both machines and humans; Linked Data provides the best practices to make these links possible. Thus, Linked Data can be seen as a set of design principles to share machine-readable interlinked data on the Web. Linked Open Data is a powerful blend of Linked Data and Open Data, in that it is both linked and uses open sources. One example of a LOD set is DBpedia.

JSON-LD stands for JavaScript Object Notation for Linked Data[66] and is a W3C Standard. It is an extension of the JSON format to embed Linked Open Data into web pages. As such, data structured according the JSON-LD format uses the JSON structure but express and convey information using LOD vocabularies (e.g. schema.org). JSON-LD makes it easy to bring legacy JSON data to the semantic web. We highlight that this is particularly useful whenever one wants to share pre-existing JSON documents according to the LOD principles (e.g. in a JSON document database, or a file).

Web Ontology Language (OWL) is a semantic web computational logic-based language designed to represent rich and complex knowledge about things and the relations between them. It also provides detailed, consistent and meaningful distinctions between classes, properties, and relationships. By specifying both object classes and relationship properties, as well as their hierarchical order, OWL enriches ontology modeling in semantic graph databases – these are also known as *RDF triplestores*. OWL, used together with a reasoner in such triplestores, enables

[64] https://flur.ee/2021/03/10/semantic-interoperability-exchanging-data-with-meaning/
[65] https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf/
[66] http://www.linkeddatatools.com/introduction-json-ld

Funded by the European Union

consistency checks (i.e., finding any logical inconsistencies) and ensures satisfiability checks (i.e., finding whether there are classes that cannot have instances).

Knowledge graphs are semantic graphs that align to an ontology. Knowledge graphs are fully integrated platforms for data unification, analytics, and sharing made possible by semantic graph formatting and ontological rules.

Semantic annotation or semantic enrichment is the process of tagging entities (e.g. documents) with relevant concepts. This makes unstructured content easier to find, interpret and reuse. Semantic annotation enriches content with machine-processable information by linking background information to extracted concepts. These concepts are unambiguously defined and related to each other within and outside the content. In the case of mobility data the semantic enrichment tags a trajectory segment to be linked to relevant contextual semantic information that can be available from heterogeneous data sources like in the MobiDataLab Transport Cloud.

Geospatial Ontologies. The role of Geospatial Ontologies has been recently discussed in [Claramunt 20]. According to this paper, to create a formal and computational data model to represent the geospatial concepts in a rich geospatial ontology - that can be generalisable across many fields and applications - is still a major challenge. A geospatial ontology for the web nowadays offers a series of functionalities towards the geospatial semantic web where a comprehensive set of geographical properties and abstractions can be both understood by different communities and, implemented.

An important recent trend is that of the development of geospatial ontologies has been closely addressed in the dual context of standard recommendations from the ISO and OGC, in order to represent geospatial concepts and properties to use on the Web.

Geospatial ontologies on the Web are largely based on several formats to implement RDF triples, such as XML, RDF, and JSON-LD. The GeoSPARQL language is a standard RDF SQL-based query that manipulates geospatial RDF data. GeoNames and LinkedGeoData are examples of datasets that cover a vast part of the world, for instance by allowing integration of large data repository such as OpenStreetMap.

3.4.2. Semantic web technologies in the transport domain

Transportation is a vast domain including physical and cyber transport network and their sub-domains. But the concept of mobility involves not only transportation, but also infrastructure, geography, payment, etc. And smart mobility multiplies even more the number of related sectors and activities linked to smart cities, such as sensors and Internet of Things. In consequence, considering smart mobility, it is easy to think of reusing existing ontology already developed across domains.

The Linked Open Vocabularies for Internet of Things (LOV4IoT) Transport Ontology Catalogue references already 73 ontology-based research projects.[67] The major difficulty remains the fact that

[67] https://lov4iot.appspot.com/?p=ontologies#transport

Funded by the European Union

some of these available ontologies are very specific and may contain certain classes and properties irrelevant to a new project one is initiating. Otherwise, other available ontologies may seem to be reusable but are not well-documented. Moreover, a recent study reveals that as the cyber transport network consists of some innovative and disruptive technologies such as sensors, connected roadways and Internet of Vehicles, autonomous robotics, MaaS, and even more to emerge in the near future, no ontology has yet been developed for these technologies.[68]

Hence, there is still space to consider developing ontologies for such new technologies and for others not yet known at present since mobility is a very fast changing field. There may also be a need to develop a general mobility ontology within the smart city context.

3.4.3. Applicability to MobiDataLab

Ontologies, Geospatial ontologies, RDF and OWL have been exploited in recent literature by CNR researchers related to semantic enrichment and mobility data analysis. One approach by Fileto et al in 2015 [Fileto15] proposed an ontology to represent both movement data and semantic dimensions that enriches trajectory data into a unique representation. This ontology enables a uniform query language exploiting SPARQL on the enriched datasets. A subsequent paper by Ruback et al. [Ruback16] proposed a conceptual framework for the semantic enrichment of movement data from the emerging Web of Data (or Linked Open Data), both as a unifying formalism and as the source of contextual data, which can be greatly useful for trajectories enrichment. Moreover, the semantic structure of such sources makes it easier to share and process enriched trajectories. The enrichment process is evaluated qualitatively by presenting a case study in the tourism domain. A more recent approach [Mello 19] proposes a conceptual model for enriched trajectories with many semantic aspects called Multiple Aspects Trajectories. This model expressed in RDF the logical data model and used NoSQL as the storage solution. These are results outcomes from the project MASTER (GA777695).

These solutions are relevant for MobiDataLab since they will enable the semantic enrichment of the Transport Cloud data, and subsequent analysis. The actual semantic enrichment and analysis will be performed by the Transport Cloud processors, and in particular the Semantic enrichment processor.

The CNR is developing research works on these standards and they will closely follow the evolutions of such standards.

---

[68] Yazdizadeh, A. & Farooq, B. Smart Mobility Ontology: Current Trends and Future Directions. Handbook of Smart Cities, Springer, 2021

4. Conclusion

This document provides a state-of-the-art study of existing data sharing standards, paving the way to the implementation of a standardised transport cloud prototype (WP4) to be used during the Living and Virtual Labs (WP5). Mobility data sharing standards have been studied, in particular for data used in the public transport, road traffic management, micro-mobility, ticketing, etc. On the other hand, standards for cross-domain interoperability (*horizontal* standards as opposed to *vertical* or sector-specific standards) have also been considered. Indeed, an important aspect of the project is the possibility to federate various cloud solutions and to build on existing European standards like e.g. DCAT-AP. For each of the standards considered, we have evaluated their applicability to MobiDataLab, taking into account their maturity, their ease or difficulty of implementation, and the proper expertise within the consortium or the advisory board.

Another aim of this standards-related task is to identify gaps to be filled. Indeed, availability of mobility data is still subject to compatibility, interoperability and standards issues. Even though public transport data are standardised, whether static or dynamic data, whether through de jure standards or de facto standards, yet some issues remain, and the standardisation work is still *in progress*. For example, the several NeTEx implementations or profiles in Europe make it difficult to harmonise in a cross-border solution. There are initiatives underway to move towards more harmonisation and interoperability in the scope of the ITS Directive 2010/40/EU and its Delegated Regulations. One example is the Programme Support Action (PSA) on the National Access Points Federation, recently awarded by DG MOVE. Beyond public transport, there are few widely adopted standards for other modes of passenger transport such as road data, ridesharing, taxis, bikes, e-scooters, electric vehicles, etc. New standards emerge, making it possible to exchange private mobility data (e.g. the MDS format), ticketing data (see the Calypso Networks Association and OSPT Alliance that are now collaborating), distributed journey planning and also Mobility as a Service (e.g. the TOMP-API). Following these standardisation efforts through a well-defined methodology will be the aim of the V2 of this document, namely the D2.5 (Report on new Mobility Data sharing standards).

Funded by the European Union

References

[Ontotext] https://www.ontotext.com/knowledgehub/fundamentals/

[Hamza 19] Hamza Baqa, Martin Bauer, Sonia Bilbao, Aitor Corchero, Laura Daniele, et al.. Towards semantic interoperability standards based on ontologies: Semantic Interoperability White Paper. 26 p., 2019, 10.13140/RG.2.2.26825.29282 . emse-02352822

[Claramunt 20] Christophe Claramunt. Ontologies for Geospatial information: progress and challenges ahead. Journal of spatial information science - JOSIS, 2020, pp.35-41. 10.5311/JOSIS.2020.20.666 . hal-03200202

[Fileto 15] Renato Fileto, Cleto May, Chiara Renso, Nikos Pelekis, Douglas Klein, Yannis Theodoridis. The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data. Data Knowl. Eng. 98: 104-122(2015)

[Ruback 16] Lívia Ruback, Marco Antonio Casanova, Alessandra Raffaetà, Chiara Renso, Vânia Maria P. Vidal. Enriching Mobility Data with Linked Open Data. IDEAS 2016: 173-182

[Mello 19] Ronaldo dos Santos Mello, Vania Bogorny, Luis Otávio Alvares, Luiz Henrique Zambom Santana, Carlos Andres Ferrero, Angelo Augusto Frozza, Geomar Andre Schreiner, Chiara Renso: MASTER: A multiple aspect view on trajectories. Trans. GIS 23(4): 805-822 (2019)

Funded by the
European Union

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](https://twitter.com/MobiDataLab)
#MobiDataLab

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

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

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