Ref. Ares(2026)1723008 - 16/02/2026

DATA SPACE 4 MOBILITY MODELS INTEGRATION

Project Acronym:	**23-ES-TG-DS4MOBMODELS**
Project ID :	**101176007**

Deliverable number:	**D3.2**
Deliverable name:	**Interoperability and harmonisation plan**
Work Package number:	**3. DATA INTEROPERABILITY AND FAIRNESS**
Due date :	**15/02/2026**
Delivery date :	**13/02/2026**

Lead Beneficiary :	**Ingartek**
Main author :	**Saioa Etxebarria**

Co-funded by the European Union

DS4MOB MODELS

Document fiche	
Authors:	Saioa Etxebarria [Ingartek]
Internal reviewers:	Isidro Arrieta [Ingartek]
Work Package:	WP3
Task:	T3.2 and T3.3
Nature:	R
Dissemination:	PU

Document History			
Version	**Date**	**Contributor(s)**	**Description**
V 0.0	27/12/2025	INGARTEK	Initial draft
V 0.1	09/02/2026	INGARTEK	Final draft to be reviewed
V 0.2	11/02/2026	UPM & UPC	Revision
V 1.1	16/02/2026	INGARTEK	Final version

Keywords:	Interoperability, Harmonization, Metadata Standards, Data model
Abstract (few lines):	This deliverable has been set up to establish and provide the guidelines and framework to ensure that data within the DS4MOBMODELS project can be effectively shared, integrated, and reused. The document defines the rules for structuring, describing, and exchanging data so that all project partners work with common standards, ensuring interoperability and harmonisation across diverse data sources.

Table of Contents

DEFINITIONS, ACRONYMS AND ABBREVIATIONS **6**
EXECUTIVE SUMMARY **7**
1 INTRODUCTION **8**
1.1 PURPOSE AND OBJECTIVES 8
1.2 INTENDED AUDIENCE 8
1.3 STRUCTURE OF THE DELIVERABLE 8
2 INTEROPERABILITY OF THE DATA CATALOGUE **10**
2.1 ONTOLOGIES AND THEIR ROLE IN METADATA 10
2.1.1 Internal structure of an ontology *11*
2.1.2 Ontologies and metadata structure *11*
2.1.3 Ontology profiles and reuse of standard vocabularies *12*
2.1.4 Adoption of mobilityDCAT-AP *12*
2.2 OVERVIEW OF MOBILITYDCAT-AP 13
2.2.1 Metadata for Catalogue *14*
2.2.2 Metadata for Catalogue Record *16*
2.2.3 Metadata for Dataset *18*
2.2.4 Metadata for Distribution *20*
2.2.5 Metadata for Agent *21*
2.2.6 Metadata for Location *21*
2.3 CURRENT INVENTORY: MANDATORY METADATA CAPTURED PER DATASET 22
3 STANDARDS AND REFERENCE FRAMEWORKS FOR DATA HARMONISATION **42**
3.1 DATA FORMATS FOR MOBILITY DATA HARMONISATION 42
3.1.1 Tabular and analytical data formats *42*
3.1.2 Structured and semi-structured data formats *43*
3.1.3 Geospatial data formats *43*
3.1.4 Domain-specific formats for mobility *44*
3.2 UNITS OF MEASURE AND QUANTITATIVE HARMONISATION 44
3.2.1 Units of measurement *45*
3.2.2 Temporal information and date-time representation *45*
3.2.3 Spatial reference systems and geographic coordinates *45*
3.3 FIELD NAMING CONVENTIONS 46
3.4 DATA TYPES AND VALUE REPRESENTATION 46
4 HARMONISATION APPROACH WITHIN THE MOBILITY DATA SPACE **47**
4.1 HARMONISATION WORKFLOW AND PROCESSING STAGES 47
4.2 CHALLENGES AND CONSTRAINTS 48
4.3 ADDED VALUE DATASETS 49
5 CONCLUSIONS AND NEXT STEPS **50**
6 REFERENCES **51**

Table of Figures

Figure 1. DCAT-AP-based dataset catalogue with controlled vocabularies for interoperable metadata. *Source: European Union.* 12
Figure 2. Minimum profile for mobilityDCAT-AP metadata. Source: Own elaboration based on mobilityDCAT-AP - Version 1.1.0 14
Figure 3. Harmonisation workflow schema 48

Table of Tables

Table 1. Mandatory properties for Catalogue 16
Table 2. Mandatory properties for Catalogue Record 18
Table 3. Mandatory properties for Dataset. 20
Table 4. Mandatory properties for Distribution. 21
Table 5. Mandatory properties for Agent. 21
Table 6. Recommended properties for Location. 22
Table 7. Selected Use Cases from previous deliverables. 23
Table 8. Collection of metadata from identified datasets. 41

Definitions, Acronyms and Abbreviations

Acronym	Title
AEMET	State Meteorological Agency
AMBici	Barcelona Metropolitan Bicycle Service
API	Application Programming Interface
ATTG	Territorial Transport Authority of Gipuzkoa
ATM T-Mobilitat	Barcelona Metropolitan Transport Authority T-Mobilitat
CRS	Coordinate Reference System
CRTM	Madrid Regional Transport Consortium
DATEX II	Data Exchange II (Traffic and Travel Data Standard)
DCAT	Data Catalog Vocabulary
DGT	General direction of traffic
EJ-GV	Basque Government
EMT Madrid	Madrid Municipal Transport Company
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
GBFS	General Bikeshare Feed Specification
GHG	GreenHouse-Gas
GML	Geography Markup Language
GTFS	General Transit Feed Specification
IGN	National Geographic Institute of Spain
IMT	Institute of Mobility and Transport
INE	National Statistics Institute
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
MITECO	Ministry for the Ecological Transition and the Demographic Challenge
MITMS	Ministry of Transport and Sustainable Mobility
NAPCORE	National Access Point Coordination Organisation for Europe
NAP	National Access Points
NeTEx	Network Timetable Exchange
OCM	Open Charge Map
ORTR	Observatoire régional des trafics routiers
OTEP	Spain – Portugal Cross-border Observatory
OWL	Web Ontology Language
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
TEN-T	Trans-European Transport Network
TMB	Barcelona Metropolitan Transport
UC	Use Case
URI	Uniform Resource Identifier
WP	Work Package
WPL	Work Package Leader
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Executive Summary

The development of D3.2 builds directly upon the results of Deliverable D3.1 “Specifications and collection of enriched datasets”, which provided an inventory of data sources, datasets, and metadata elements relevant to the project use cases. While D3.1 focused on identifying and analysing the available data assets, D3.2 advances this work by defining the interoperability and harmonisation framework required to enable their effective integration and reuse within a Mobility Data Space.

This deliverable establishes a common approach for ensuring interoperability at both metadata and data level. It describes the adoption of mobilityDCAT-AP as the reference metadata ontology, detailing its mandatory classes and properties and explaining how it is applied to consistently describe datasets identified in the project. In addition, the deliverable defines the set of standards, specifications, and reference frameworks to be applied for data harmonisation, covering aspects such as data formats, geospatial representations, units of measurement, temporal references, naming conventions, and data typing.

Beyond the definition of standards, D3.2 outlines a harmonisation approach aligned with the operational principles of a data space, where data are ingested directly from heterogeneous sources and transformed through value-adding processes. Harmonisation is presented not only as a technical requirement, but as a key enabler for value creation, supporting the development of interoperable and high-quality data products that underpin the project use cases and contribute to the sustainability and business models of the Mobility Data Space.

The scope, structure, and content of D3.2 are closely aligned with the use cases defined in Deliverable D3.1, including the incorporation of the Barcelona metropolitan use case, ensuring continuity and coherence across the project’s technical work. Furthermore, the framework defined in this deliverable provides the foundation for Task 3.4 on data quality assurance according to FAIR principles and directly informs Deliverable D3.3 “Data quality monitoring and assessment plan”, which will operationalise data quality verification, cleaning, and validation pipelines.

The main target group for this deliverable is the project consortium and technical partners involved in data integration, harmonisation, and platform development, as it provides the methodological and technical basis required to implement interoperable, harmonised, and reusable mobility data services within the DS4MOBMODELS project. Once public, the secondary audience is external stakeholders.

1 Introduction

1.1 Purpose and objectives

The purpose of this deliverable is to define the **Interoperability and Harmonisation Plan** of the project, establishing a common and coherent framework to ensure that the datasets identified and analysed in previous phases can be consistently represented, integrated and reused within the Mobility Data Space. This deliverable provides a structured overview of the **standards, data models and harmonisation approaches** applicable to the project datasets, together with the methodological principles guiding their adoption and implementation.

This deliverable builds directly upon the results of **Deliverable D3.1**, which presented an inventory of data sources and datasets and assessed their metadata, quality, accessibility and readiness for integration. In addition, it maintains continuity with the mobility and logistics use cases defined and validated in **Deliverable D2.3**, ensuring that the selected interoperability standards and harmonisation strategies are aligned with the functional requirements and information needs identified for each use case.

To achieve these objectives, the deliverable focuses on the following key aspects:

- **Define the interoperability framework of the project data catalogue**, specifying the metadata models to be applied at catalogue, dataset and distribution levels, in alignment with the Mobility Data Space.
- **Identify and describe the standards applicable** to the datasets addressed in the project, covering metadata, data representation and semantic interoperability.
- **Define the harmonisation plan**, describing the methods, processes and transformations required to achieve syntactic and semantic alignment with the selected standards and data models.
- **Address interoperability challenges and constraints** identified in previous phases, identifying issues related to data heterogeneity, incomplete standardisation and differing levels of data maturity.

1.2 Intended audience

This deliverable will be disseminated publicly, addressing all interested stakeholders who require an understanding of the interoperability framework and harmonisation strategy adopted within the project. In particular, it provides insight into the standards, data models and methodologies applied to ensure consistent representation and integration of heterogeneous mobility-related datasets within the Mobility Data Space.

1.3 Structure of the deliverable

This deliverable is structured as follows:

- **Section 2** describes the interoperability framework of the project data catalogue, including the metadata models adopted for the catalogue, datasets and data distributions.
- **Section 3** presents the main standards, specifications and reference frameworks commonly applied for data harmonisation within the project context.
- **Section 4** defines the harmonisation approach to be followed, including the methods, tools and processes applied to achieve syntactic and semantic alignment with the selected standards.

- **Section 5** presents a synthesis of the key results and conclusions of the analysis carried out in this deliverable, highlighting the main insights obtained and outlining the elements that will inform and support the subsequent phases of the project development.

2 Interoperability of the Data Catalogue

The interoperability of the DS4MOBMODELS data catalogue aims to ensure that the datasets identified and analysed in Deliverable D3.1 can be consistently described, easily discovered, and effectively reused across different organisations, pilots, and external Mobility Data Space initiatives. Interoperability at catalogue level is a prerequisite to enable cross-domain data sharing, avoid vendor- or platform-specific metadata descriptions, and support the integration of DS4MOBMODELS datasets within wider European mobility and logistics data ecosystems.

In line with the objectives of WP3, this is achieved through (i) the adoption of **standardised metadata vocabularies** and (ii) the provision of **machine-readable metadata descriptions**, allowing automated processing, validation, harvesting, and exchange of metadata between heterogeneous catalogues and platforms. This approach ensures that dataset metadata can be interpreted in a uniform way by both humans and machines, independently of the system or organisation publishing the data.

In practice, interoperability at data catalogue level is addressed through the following complementary dimensions:

- **Syntactic interoperability**, which ensures that metadata are expressed using common serialisation formats and exchange mechanisms. By relying on widely adopted Semantic Web technologies and RDF serialisations, metadata can be automatically processed, validated, and exchanged between catalogues without the need for bespoke transformations.
- **Semantic interoperability**, which focuses on guaranteeing a shared and unambiguous meaning of metadata elements. This is achieved through the use of standard metadata vocabularies, controlled vocabularies, and explicit conceptual models, enabling different stakeholders to describe datasets in a harmonised manner and to interpret key concepts (such as themes, transport modes, spatial coverage or access conditions) consistently across catalogues.
- **Organisational interoperability**, which addresses the alignment of publication and governance practices related to the data catalogue. This includes the definition of clear roles and responsibilities for metadata provision and maintenance, agreed update and validation processes, and common rules for dataset publication. These measures ensure that metadata remain accurate, reliable, and up to date over time, thereby reinforcing trust and long-term reuse of the catalogued datasets.

2.1 Ontologies and their role in metadata

At the heart of semantic interoperability are ontologies. An **ontology** can be defined as a formal and explicit representation of knowledge within a specific domain, capturing both the relevant concepts and the of information that can be reused across different systems, organisations, and use cases.

In practical terms, an **ontology establishes a common vocabulary** that data publishers and data consumers can rely on when describing and interpreting data. By formally defining concepts (or classes) such as Dataset, Catalogue, Distribution, or Agent, as well as their associated properties (for example title, publisher, format, or spatial coverage), an ontology ensures that these terms have a precise and unambiguous meaning. This is particularly important in distributed data environments, where metadata is produced by multiple actors following different internal practices.

For metadata management, this shared understanding is essential. If two or more data catalogues adopt the same ontology, a Dataset described in one catalogue is semantically equivalent to a Dataset described in another. This semantic alignment enables automated processing, comparison, and integration of metadata across platforms, which is a key requirement for large-scale data spaces such as the Mobility Data Space.

2.1.1 Internal structure of an ontology

Beyond providing a shared vocabulary, ontologies are characterised by a well-defined internal structure that determines how knowledge is formally represented. An ontology is not merely a list of terms; rather, it is composed of several interrelated elements that together enable consistent interpretation, reasoning, and reuse of information.

At the core of an ontology are **classes**. Classes represent the fundamental ideas or entities within a given domain. In the context of metadata and data catalogues, typical examples include Catalogue, Dataset, Distribution, or Agent. These classes define the types of things that can exist in the domain and provide the backbone for structuring metadata descriptions.

Closely related to concepts are **relationships**, which describe how concepts are connected to one another. Relationships can express hierarchical structures, such as subclass relations, or semantic associations, such as a dataset being published by an organisation or being distributed as a specific format. These relationships form a domain model or vocabulary and allow metadata to move beyond flat descriptions towards a graph-based representation of knowledge.

Ontologies may also include **functions**, which represent specific types of relationships where the value of one element is determined based on others. While less prominent in metadata vocabularies, functions can be used to derive or assign values, such as categorisation rules or temporal associations.

Another key component is **instances**. Instances are concrete real-world entities that belong to a given class. For example, a specific mobility dataset published by a transport authority is an instance of the class Dataset, while the authority itself is an instance of Agent. In metadata catalogues, instances are typically identified using persistent URIs, enabling them to be referenced consistently across systems.

Finally, ontologies may define **axioms** and constraints, which are statements that are always considered true within the domain. These axioms can be used to express rules, consistency conditions, or logical restrictions. In practical terms, they support validation mechanisms and help ensure that metadata descriptions conform to the intended model.

2.1.2 Ontologies and metadata structure

In the context of data spaces and cross-domain interoperability, **ontologies are typically implemented using Semantic Web standards** such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL). These standards provide a formal and machine-readable way to represent metadata as interconnected graphs rather than isolated records.

By relying on **RDF-based representations**, metadata elements such as datasets, organisations, distributions or spatial references are identified through globally unique and persistent identifiers (URIs). This approach enables metadata to be unambiguously referenced, linked and reused across different catalogues and platforms, supporting the principles of Linked Data.

For metadata management, this graph-based representation is particularly relevant. It allows metadata descriptions to be extended, combined and enriched over time without breaking existing structures. New properties or relationships can be added by reusing existing ontologies or vocabularies, ensuring both backward compatibility and semantic consistency.

As a result, ontologies used for metadata representation do not operate in isolation. Instead, they are typically composed of multiple interoperable vocabularies, each addressing a specific aspect of the description, such as cataloguing, thematic classification, temporal and spatial coverage, or organisational roles.

2.1.3 Ontology profiles and reuse of standard vocabularies

In large-scale interoperability initiatives, and particularly in European data spaces, ontologies are rarely developed from scratch. Instead, a common practice is to define ontology profiles that reuse and constrain existing standard vocabularies.

An ontology profile can be understood as a tailored specification that identifies which classes, properties and controlled vocabularies from one or more reference ontologies are applicable in a given context, and how they should be used. Profiles may define mandatory, recommended or optional elements, as well as additional constraints to ensure consistency across implementations.

This profiling approach offers several advantages. It ensures alignment with widely adopted standards, facilitates interoperability beyond the boundaries of a single project, and reduces the implementation effort for data providers and consumers. At the same time, it allows projects to address their specific requirements without compromising semantic compatibility.

In the domain of metadata for data catalogues, this approach is exemplified by DCAT-AP and its domain-specific extensions, which provide profiles of the W3C DCAT ontology adapted to the European and sectorial contexts.

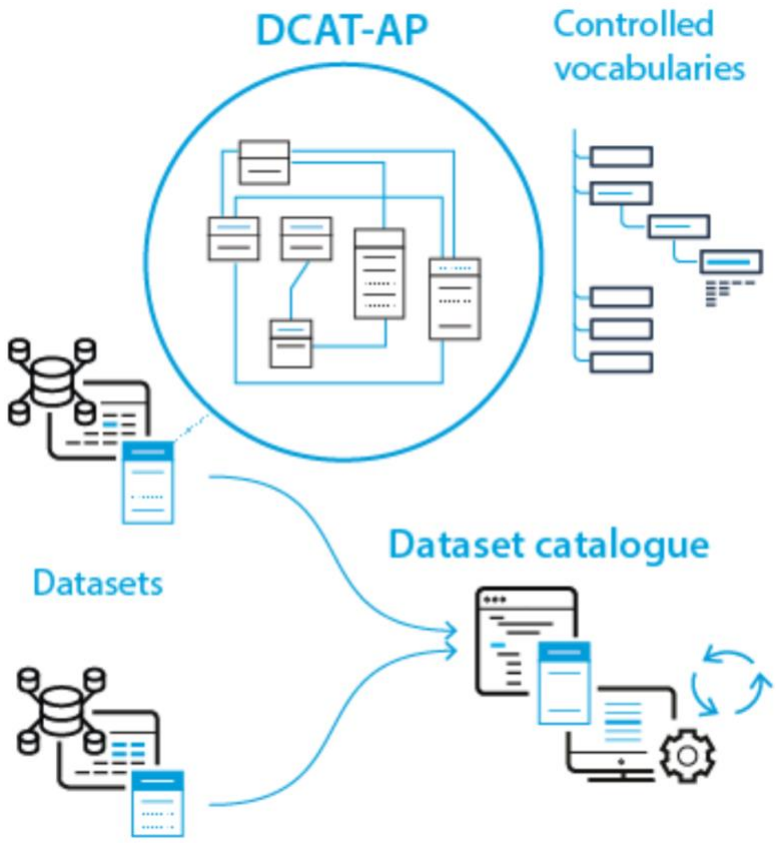


Figure 1. DCAT-AP-based dataset catalogue with controlled vocabularies for interoperable metadata. *Source: European Union.*

2.1.4 Adoption of mobilityDCAT-AP

Building on the concepts introduced in the previous sections, metadata interoperability within DS4MOBMODELS is approached through the adoption of an ontology-based data model defined as a profile of existing standard vocabularies. This approach is fully aligned with current best practices in

European data spaces, where semantic interoperability is achieved by reusing and constraining widely adopted ontologies rather than introducing project-specific models.

As established in Deliverable D3.1, **mobilityDCAT-AP** has been identified as the reference data model for describing dataset metadata within the project. In the context of the present deliverable, mobilityDCAT-AP is therefore considered not only as a metadata specification, but as a formal ontology profile that provides a semantically consistent and machine-readable representation of the information collected across the different use cases and pilots.

MobilityDCAT-AP builds upon the W3C DCAT ontology and its European application profile DCAT-AP, extending them with mobility- and transport-specific concepts and controlled vocabularies. By reusing core semantic building blocks such as DCAT, Dublin Core Terms, SKOS, and FOAF, and by integrating domain-specific extensions, mobilityDCAT-AP exemplifies the ontology profiling principles described in the previous sections. It enables the representation of datasets, distributions, organisations, spatial and temporal coverage, and thematic classifications in a harmonised and interoperable manner.

Within DS4MOBMODELS, the role of this ontology profile is to provide a common semantic framework for metadata description, ensuring that datasets originating from heterogeneous sources can be consistently described, discovered, and interpreted across catalogues and platforms. The focus of Deliverable D3.2 is therefore not on re-evaluating the suitability of mobilityDCAT-AP, but on detailing its conformation and application to the project context, including the specific classes, properties, and vocabularies that are used to represent the collected information.

2.2 Overview of mobilityDCAT-AP

In this section, the mandatory elements of the mobilityDCAT-AP metadata model are analysed, focusing on the core classes, properties, and controlled vocabularies required to achieve conformance with the specification. The analysis is based on the minimum mobilityDCAT-AP profile and serves as the foundation for the detailed descriptions of each metadata class provided in the subsequent sections.

The **minimum mobilityDCAT-AP profile considering only mandatory elements is described in the following schema**. The model is centred around three core classes inherited from DCAT-AP: `dcat:Catalog`, `dcat:Dataset`, and `dcat:Distribution`. These classes represent, respectively, the data catalogue as an entry point, the datasets described within the catalogue, and the concrete access mechanisms through which datasets are made available.

In addition to these core entities, the profile includes supporting classes required to provide essential contextual information. These include `dcat:CatalogRecord`, used to represent catalogue-level metadata records, `foaf:Agent`, used to describe publishers and responsible organisations, and `dct:Location`, used to capture the spatial coverage of datasets.

Mobility-specific extensions are introduced through additional mandatory properties that reference controlled vocabularies expressed as SKOS concept schemes. In particular, datasets are required to be classified using the `mobilitydcatap:mobilityTheme` property, which links datasets to a shared mobility-themed taxonomy, while distributions may reference the applicable mobility data standard through the `mobilitydcatap:mobilityDataStandard` property. These controlled vocabularies are essential to ensure consistent interpretation of metadata values across catalogues and platforms.

In the Figure 2, the classes and properties highlighted in blue indicate the elements that have been introduced, extended, or made mandatory by mobilityDCAT-AP with respect to the core DCAT-AP specification. These elements represent the **domain-specific adaptations required to capture mobility- and transport-related characteristics that are not explicitly covered by the generic DCAT-AP model.** Dotted boxes refer to the controlled vocabularies to be used as values for the properties referencing them.

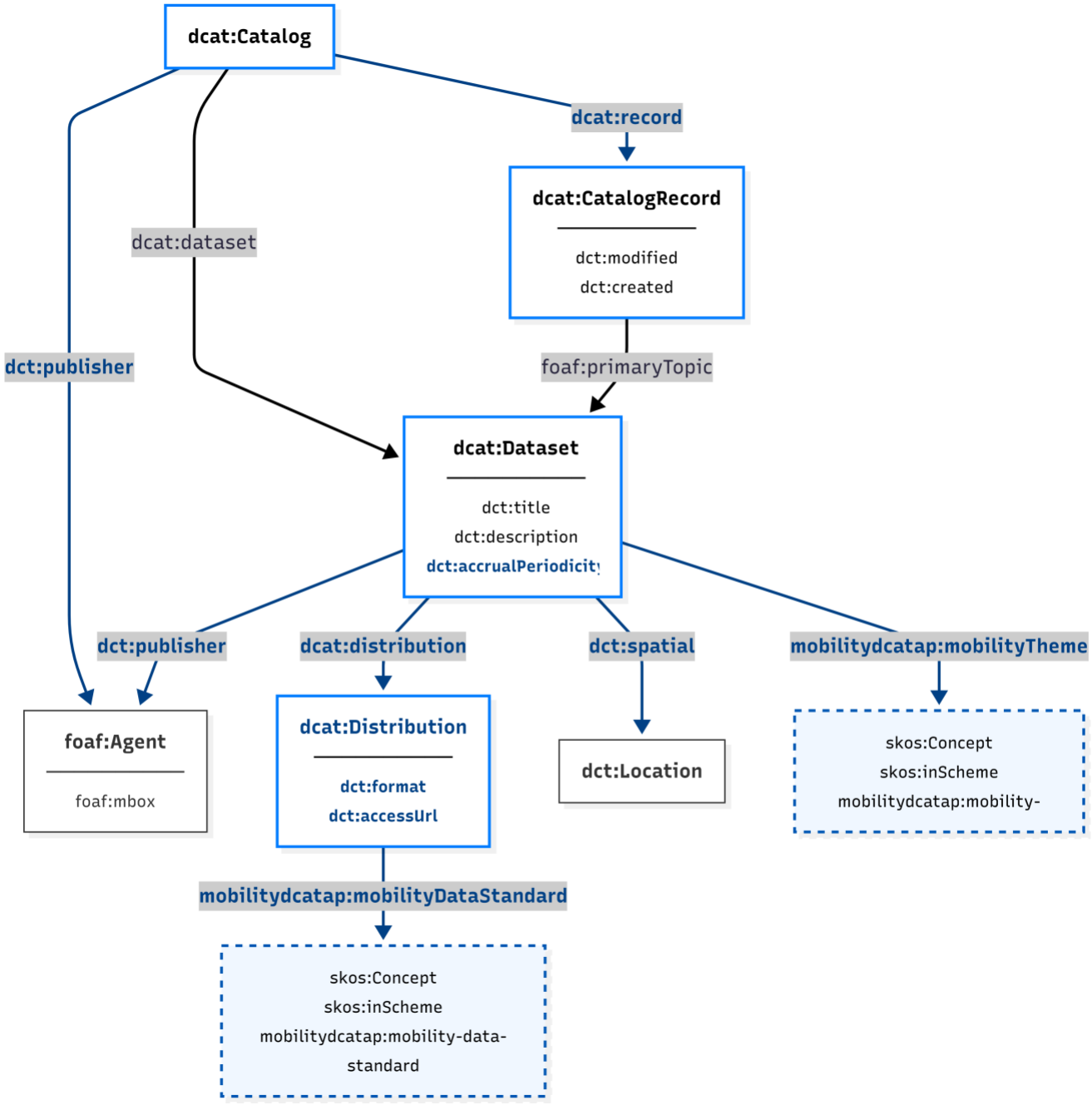


Figure 2. Minimum profile for mobilityDCAT-AP metadata. Source: Own elaboration based on mobilityDCAT-AP - Version 1.1.0

2.2.1 Metadata for Catalogue

The Catalogue is a core and mandatory component of the mobilityDCAT-AP metadata model, as it represents the entry point through which mobility datasets are made discoverable within a data portal or catalogue service. In mobilityDCAT-AP, a catalogue corresponds to a concrete mobility data portal where datasets and their associated distributions are described using structured metadata and exposed to data users for discovery and access.

As defined by the `dcat:Catalog` class, the catalogue provides high-level descriptive information about the data offering as a whole and establishes the organisational and contextual framework in which datasets are published. It groups together one or more datasets, links them to their corresponding

metadata records, and identifies the organisation responsible for the publication of the catalogue. Within the minimum mobilityDCAT-AP profile, the catalogue ensures a consistent and interoperable starting point for metadata exchange, upon which the mandatory metadata elements described in the following subsection are applied to guarantee conformance with the specification.

The following tables (Table 1) describe the mandatory metadata properties required for the description of a Catalogue in mobilityDCAT-AP.

Title	
Description	This property provides a human-readable name for the mobility data catalogue.
Type	`rdfs:Literal` – Text string (UNICODE), optionally with language tags.
Property	`dct:title`
Obligation	Mandatory
Comments	The title may be repeated to provide representations in multiple languages.
Example	`<dct:title xml:lang="en">Mobility Data Catalogue</dct:title>`

Description	
Description	This property provides a textual description of the catalogue, including its scope and the type of mobility datasets it contains.
Type	`rdfs:Literal` – Text string (UNICODE), optionally with language tags.
URI	`dct:description`
Obligation	Mandatory
Comments	The description should allow users to understand the purpose of the catalogue and its contents.
Example	`<dct:description xml:lang="en">Catalogue aggregating datasets related to mobility and transport.</dct:description>`

Publisher	
Description	This property identifies the entity responsible for making the catalogue available and maintaining its metadata.
Type	Resource – Instance of `foaf:Agent`
URI	`dct:publisher`
Obligation	Mandatory
Comments	The publisher can be an organisation or public authority. Additional information (e.g., name, email) may be provided through the referenced foaf:Agent resource.
Example	`<dct:publisher rdf:resource="http://example.org/organisation/ds4mobmodels"/>`
Homepage	
Description	This property provides the main webpage of the mobility data catalogue.
Type	Resource – Instance of `rdfs:Resource`
URI	`foaf:homepage`
Obligation	Mandatory
Comments	The homepage should point to a human-readable web page where users can browse the catalogue.

Example	`<foaf:homepage rdf:resource="https://data.example.org/mobility-catalogue"/>`

Spatial / Geographic	
Description	This property refers to a country code where the data platform is hosted.
Type	Resource – Instance of `dct:Location`
URI	`dct:spatial`
Obligation	Mandatory
Comments	The value should represent the hosting country (typically expressed via a URI identifying the country/territory).
Example	`<dct:spatial rdf:resource="https://www.geonames.org/2510769/kingdom-of-spain.html"/>`

Dataset	
Description	This property links the mobility data portal with a dataset that is described at the mobility data portal.
Type	Resource – Instance of `dcat:Dataset`
URI	`dcat:dataset`
Obligation	Mandatory
Comments	A catalogue shall reference at least one dataset. This property may be repeated to link multiple datasets.
Example	`<dcat:dataset rdf:resource="http://example.org/dataset/traffic-counts"/>`

Record	
Description	This property links the catalogue to one or more catalogue records describing individual datasets.
Type	Resource – Instance of `dcat:CatalogRecord`
URI	`dcat:record`
Obligation	Mandatory
Comments	Each catalogue record provides metadata about the inclusion of a dataset in the catalogue. This property may be repeated.
Example	`<dcat:record rdf:resource="http://example.org/catalogue-record/record-1"/>`

Table 1. Mandatory properties for Catalogue

2.2.2 Metadata for Catalogue Record

The Catalogue Record is a mandatory component of the mobilityDCAT-AP metadata model and **represents the metadata entry through which an individual dataset is registered and managed within a mobility data portal**. Each catalogue record refers to a single dataset and its associated distributions, providing information about the lifecycle and maintenance of the metadata entry itself rather than about the dataset content.

In mobilityDCAT-AP, the importance of catalogue records is explicitly reinforced compared to DCAT-AP, as metadata records are considered essential for the operation of mobility data portals and National Access Points. For this reason, the obligation of the `dcat:CatalogRecord` class is raised to mandatory,

ensuring that each dataset described in a catalogue is accompanied by a corresponding catalogue record. The following tables (Table 2) describe the mandatory metadata properties required for Catalogue Records to ensure conformance with mobilityDCAT-AP.

Primary topic	
Description	This property links the catalogue record to the dataset that is described by this metadata entry.
Type	Resource – Instance of `dcat:Dataset`
URI	`foaf:primaryTopic`
Obligation	Mandatory
Comments	Each Catalogue Record shall refer to exactly one dataset as its primary topic.
Example	`<foaf:primaryTopic rdf:resource="http://example.org/dataset/traffic-counts"/>`

Language	
Description	This property specifies the language of the metadata in the catalogue record.
Type	Resource – URI identifying a language `dct:LinguisticSystem`
URI	`dct:language`
Obligation	Mandatory
Comments	The value should be a language identifier from a controlled vocabulary (e.g., ISO 639-1 / ISO 639-2 as published in a recognised authority list). This property refers to the language used in the metadata record, not the language of the dataset content.
Example	`<dct:language rdf:resource="http://publications.europa.eu/resource/authority/language/ENG"/>`

Creation date	
Description	This property indicates the date and time when the catalogue record was created in the catalogue.
Type	Literal – `xsd:date` or `xsd:dateTime`
URI	`dct:created`
Obligation	Mandatory
Comments	It refers to the creation of the metadata record (not necessarily the creation of the dataset itself).
Example	`<dct:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2026-02-04T10:15:00Z</dct:created>`
Update / Modification date	
Description	This property indicates the most recent date and time on which the catalogue record was modified.
Type	Literal – `xsd:date` or `xsd:dateTime`
URI	`dct:modified`
Obligation	Mandatory
Comments	It SHOULD be generated by the system, whenever a platform user updates the metadata entry. If there has been no update yet, the value from the property "creation date" SHOULD be taken.

Example	`<dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2026-02-04T12:30:00Z</dct:modified>`

Table 2. Mandatory properties for Catalogue Record

2.2.3 Metadata for Dataset

The Dataset is a mandatory core class of the mobilityDCAT-AP metadata model and represents a coherent collection of mobility-related data that is published or curated by a single source. In the context of mobility data portals, **a dataset corresponds to the conceptual data offering**, independently of the specific technical means through which the data is accessed or distributed.

Within mobilityDCAT-AP, each dataset is described using structured metadata that enables its discovery, interpretation, and reuse across platforms and organisations. A dataset may be associated with one or more distributions, which provide the concrete access mechanisms to the data, but the dataset itself captures the semantic and descriptive aspects of the data collection, such as its title, description, thematic classification, spatial and temporal characteristics, and responsible parties. The following tables (Table 3) specify the mandatory metadata properties for describing a Dataset in accordance with mobilityDCAT-AP.

Title	
Description	This property provides a human-readable name for the dataset.
Type	Literal – Text string (UNICODE), `rdfs:Literal`.
URI	`dct:title`
Obligation	Mandatory
Comments	The title may be repeated to provide representations in multiple languages.
Example	`<dct:title xml:lang="en">Traffic Counts on National Roads</dct:title>`

Publisher	
Description	This property identifies the entity responsible for publishing the dataset.
Type	Resource – Instance of foaf:Agent
URI	`dct:publisher`
Obligation	Mandatory
Comments	This information can be natively created by a data platform, then corresponding to the entity that is registered to the data platform and has the role of a data publisher.
Example	`<dct:publisher rdf:resource="http://example.org/organisation/road-authority"/>`

Description	
Description	This property provides a textual description of the dataset, explaining its content and purpose.
Type	Literal – Text string (UNICODE), `rdfs:Literal`.
URI	`dct:description`
Obligation	Mandatory
Comments	The description should enable users to understand what the dataset contains and how it can be used. It may be repeated to provide representations in multiple languages.

Example	`<dct:description xml:lang="en">Dataset containing aggregated traffic counts collected on national roads.</dct:description>`

Frequency	
Description	This property indicates how frequently the dataset is updated.
Type	Resource – URI from a controlled vocabulary `dct:Frequency`
URI	`dct:accrualPeriodicity`
Obligation	Mandatory
Comments	The value should be taken from an authority list such as the EU frequency vocabulary.
Example	`<dct:accrualPeriodicity rdf:resource="http://publications.europa.eu/resource/authority/frequency/DAILY"/`

Spatial Coverage	
Description	This property describes the geographic area covered by the dataset content.
Type	Resource – Instance of `dct:Location`
URI	`dct:spatial`
Obligation	Mandatory
Comments	This property refers to the spatial coverage of the dataset itself and must not be confused with dct:spatial at Catalogue level. Several options of controlled vocabularies are provided.
Example	`<dct:spatial rdf:resource="https://www.geonames.org/2510769/kingdom-of-spain.html"/>`

Mobility Theme	
Description	This property classifies the dataset according to a mobility-specific thematic category.
Type	Resource – `skos:Concept`
URI	`mobilitydcatap:mobilityTheme`
Obligation	Mandatory
Comments	The value shall be taken from the controlled vocabulary of mobility themes defined by mobilityDCAT-AP.
Example	`<mobilitydcatap:mobilityTheme rdf:resource="https://w3id.org/mobilitydcat-ap/mobility-theme/traffic-volume"/>`
Dataset Distribution	
Description	This property links the dataset to one or more distributions through which the data is made available.
Type	Resource – Instance of `dcat:Distribution`
URI	`dcat:distribution`
Obligation	Mandatory
Comments	Each dataset shall be associated with at least one distribution. This property may be repeated.

Example	`<dcat:distribution rdf:resource="http://example.org/distribution/traffic-counts-api"/>`

Table 3. Mandatory properties for Dataset.

2.2.4 Metadata for Distribution

The Distribution is a mandatory component of the mobilityDCAT-AP metadata model and represents a concrete way in which a dataset is made available to data consumers. **A distribution describes the technical characteristics required to access the data**, such as the access URL, the format, or the communication method used to retrieve the data. A single dataset may be associated with one or multiple distributions, each corresponding to a different access mechanism or data representation.

While in DCAT-AP the Distribution class is classified as recommended to allow for cases where a dataset may not be directly accessible, mobilityDCAT-AP explicitly raises the obligation of this class to mandatory. This reflects the expectations of mobility data portals and National Access Points, where datasets are intended to be directly accessible through one or more concrete distributions. The tables below (Table 4) specify the mandatory metadata fields for Distributions, as required by the mobilityDCAT-AP standard for interoperability.

Access URL	
Description	This property provides the URL through which the distribution can be accessed. It may point to a landing page, an API endpoint, or a service interface used to retrieve the data.
Type	Resource – URI `rdfs:Resource`
Property	`dcat:accessURL`
Obligation	Mandatory
Comments	The access URL does not necessarily have to be a direct download link; it may refer to a webpage or service describing how to access the data.
Example	`<dcat:accessURL rdf:resource="https://api.example.org/traffic-counts"/>`

Format	
Description	This property indicates the file format or technical representation of the distribution.
Type	Resource – URI (typically from a controlled vocabulary) `dct:MediaTypeOrExtent`
URI	`dct:format`
Obligation	Mandatory
Comments	The value should refer to a standard media type or file format identifier.
Example	`<dct:format rdf:resource="http://publications.europa.eu/resource/authority/file-type/CSV"/>`

Mobility Data Standard	
Description	This property specifies the mobility data standard or specification that the distribution conforms to.
Type	Resource – SKOS Concept (from the mobility data standard controlled vocabulary)
URI	`mobilitydcatap:mobilityDataStandard`
Obligation	Mandatory

Comments	The value shall be taken from the controlled vocabulary for mobility data standards to support consistent interpretation and discovery. If no established mobility data standard is applied, an information "proprietary standard" is requested.
Example	`<mobilitydcatap:mobilityDataStandard rdf:resource="https://w3id.org/mobilitydcat-ap/mobility-data-standard/gtfs"/>`

Rights	
Description	This property specifies the rights or licence under which the distribution is made available, defining the conditions for access and reuse of the data.
Type	Resource – URI identifying a licence or rights statement `dct:RightsStatement`
URI	`dct:rights`
Obligation	Mandatory
Comments	As a minimum, it should include an information on the type of conditions for access and usage.
Example	`<dct:rights rdf:resource="http://creativecommons.org/licenses/by/4.0/"/>`

Table 4. Mandatory properties for Distribution.

2.2.5 Metadata for Agent

The Agent is a mandatory supporting class in the mobilityDCAT-AP metadata model and is used to represent entities that play a role in the publication, management, or description of metadata. **An agent may correspond to an individual, an organisation, or another legal entity**, and can be associated with catalogues, catalogue records, datasets, or distributions in different roles, such as publisher or contact point.

MobilityDCAT-AP adopts the `foaf:Agent` class to ensure compatibility with existing metadata standards and to support consistent identification of responsible entities across catalogues and platforms. When the agent represents an organisation, the use of the W3C Organization Ontology is recommended to further describe organisational structures and attributes.

Name	
Description	This property provides the name of the agent, identifying the individual or organisation responsible for a catalogue, dataset, or distribution.
Type	Literal – Text string (UNICODE), `rdfs:Literal`.
URI	`foaf:name`
Obligation	Mandatory
Comments	The name should be a human-readable label identifying the agent. It may be repeated to provide representations in multiple languages.
Example	`<foaf:name xml:lang="en">Example Transport Authority</foaf:name>`

Table 5. Mandatory properties for Agent.

2.2.6 Metadata for Location

The Location class is a mandatory supporting component in the mobilityDCAT-AP metadata model and is used to represent spatial regions or named places associated with catalogues or datasets. It provides a structured way to describe geographic references in metadata, enabling consistent identification and interpretation of spatial information across mobility data portals and federated catalogues.

In mobilityDCAT-AP, locations may be represented either by referring to named places through controlled vocabularies or, less preferably, by using geographic coordinates, in line with DCAT-AP

spatial guidelines. For mobility data portals, the use of named places is explicitly encouraged, as mobility datasets are often provided by public authorities operating at well-defined administrative levels such as countries, regions, or municipalities. To ensure semantic consistency and interoperability, identifiers for locations should preferably be expressed using recognised controlled vocabularies, such as EU authority tables, NUTS codes, or GeoNames URIs. Where no suitable identifier is available, clear geographic names may alternatively be provided using human-readable labels. The following tables (Table 6) describe the recommended metadata properties for Location instances in mobilityDCAT-AP, as no mandatory properties are defined for this class.

Gazetteer	
Description	This property MAY be used to specify the gazetteer to which the Location belongs.
Type	Resource – Instance of `skos:ConceptScheme`
URI	`skos:inScheme`
Obligation	Recommended
Comments	This property is analogue to an addition by GEODCAT-AP. It can be used to indicate which reference gazetteer/vocabulary is used to identify the place (e.g., GeoNames, NUTS, EU authority tables).
Example	`<skos:inScheme rdf:resource="http://sws.geonames.org/"/>`

Geographic Identifier	
Description	This property is used to provide a geographic identifier for the Location (i.e., an identifier from a controlled vocabulary).
Type	Literal – `rdfs:Literal`
URI	`dct:identifier`
Obligation	Recommended
Comments	Use identifiers from controlled vocabularies such as EU authority tables, NUTS, or GeoNames. (Even though it's a literal, it is commonly expressed as a URI string.)
Example	`<dct:identifier>http://sws.geonames.org/2510769/</dct:identifier>`

Table 6. Recommended properties for Location.

2.3 Current inventory: mandatory metadata captured per dataset

At the time of preparation of this deliverable, a consolidated Mobility Data Space or centralised metadata catalogue has not yet been deployed. Therefore, the metadata collected so far have been gathered directly from the datasets identified in the project inventory, rather than through a common catalogue infrastructure. For this reason, the current inventory focuses on the mandatory metadata elements defined for the `dcat:Dataset` class in mobilityDCAT-AP, which represent the minimum information required to describe datasets in a consistent and interoperable manner.

It is important to note that the identified datasets are not isolated assets but are directly linked to one or more project use cases. **The use cases defined and analysed in Deliverable D2.3 are recalled below (Table 7), complemented by the inclusion of the Barcelona metropolitan use case**, which had not yet been fully defined at the time of preparation of that deliverable.

Use Case Id	Use Case Title
National Pilot	
UC Winter	Winter service management: Emergency parking for heavy vehicles
UC13 Nac	Highway and main road congestion monitoring
Regional Pilot: *Andalusia, Valencia and Catalonia Region*	
UC16 And	Transport emissions evaluation [Andalusia Region]
UC21 CV	Low Emission Zone monitoring [Valencia Region]
UC22 Cat	Active mobility monitoring [Catalonia Region]
Local / Metropolitan Pilot: *Madrid City*	
UC3 Mad	Intermodal node demand analysis
Cross-border Pilot - *Portugal*	
UC5 Port	Cross-border passenger transport
UC10 Port	Cross-border freight transit analysis
UC18 Port	Freight emissions
UC19 Port	Charging infrastructure
Cross-border Pilot - *France*	
UC5 Fra	Cross-border passenger transport
UC10 Fra	Cross-border freight transit analysis
UC18 Fra	Freight emissions
UC19 Fra	Charging infrastructure
Local / Metropolitan Pilot: *Barcelona City*	
UC3 Bcn	Intermodal node demand analysis

Table 7. Selected Use Cases from previous deliverables.

The Table 8 presents the identified datasets together with the mandatory dataset-level metadata elements captured at this stage. For clarity and readability, some metadata elements are shown in simplified form, using conceptual labels instead of their full URI references, while remaining fully aligned with the mobilityDCAT-AP specification. However, for specific elements—namely update frequency, mobility theme, and spatial coverage—values have been assigned using the corresponding controlled vocabularies defined or referenced by mobilityDCAT-AP, ensuring semantic consistency and alignment with the interoperability framework.

At this stage, the table does not include information related to dataset distributions. This is due to the fact that a common Mobility Data Space or catalogue infrastructure has not yet been deployed within the project, and concrete access mechanisms for the datasets will be defined and described at a later phase once the data space becomes operational.

UC	Title	Description	Publisher	Frequency	+mobility theme	Spatial / geographic coverage
UC Winter	Winter road safety parking areas	Winter road maintenance parking areas dataset describing designated facilities on the national road network that provide safe parking for heavy vehicles during adverse weather conditions	MITMS	ANNUAL	parking-service-and-rest-area-information; truck-parking-locations-and-conditions	kingdom-of-Spain
UC Winter	European Safe and Secure Truck Parking Areas	This dataset provides information related to road freight transport and truck stopping locations in Europe, focusing on safety, security, working conditions for drivers, and risks such as cargo theft associated with parked heavy goods vehicles.	ESPORG	OTHER	parking-service-and-rest-area-information; truck-parking-locations-and-conditions	Europe
UC Winter|UC13 Nac| UC16 And|UC21 CV| UC22 Cat|UC3 Mad| UC3 Bcn|UC18 Port| UC18 Fra	Mobility study using Big Data	This dataset is based on anonymized mobile phone positioning data, collected in full compliance with Spanish data protection regulations (Organic Law 3/2018 on Personal Data Protection and the guarantee of digital rights). The content is organized by study type and further structured into subfolders according to different spatial zoning levels (municipal, district, and GAUS). For each zoning level, the dataset includes folders by product type, quality metadata, and the spatial zoning used, along with its associated information.	MITMS	MONTHLY	other	kingdom-of-Spain

UC Winter|UC13 Nac| UC16 And|UC22 Cat| UC21 CV|UC3 Mad| UC3 Bcn|UC5 Port| UC10 Port|UC18 Port| UC19 Port|UC5 Fra| UC10 Fra|UC18 Fra| UC19 Fra	Spanish Transport Network	Road transport networks (urban and interurban roads), railways, inland waterways, air and cable transport, and their intermodal connections, with national coverage.	IGN	UNKNOWN	network-geometry-and-lane-character	kingdom-of-Spain
UC Winter|UC13 Nac| UC18 Port|UC18 Fra	Meteorological data	Meteorological and climatological information from AEMET.	AEMET	10MIN	road-weather-conditions	kingdom-of-Spain
UC Winter|UC13 Nac| UC22 Cat|UC3 Mad| UC3 Bcn|UC5 Port| UC5 Fra	GHG Emissions Inventory	This dataset contains the annual Greenhouse Gas (GHG) Emissions Inventory prepared by the Spanish Emissions Inventory and Projections System. It reports anthropogenic GHG emissions by source and removals by sinks, in accordance with the UN Framework Convention on Climate Change (UNFCCC), the Paris Agreement, and EU Regulation (EU) 2018/1999 on the Governance of the Energy Union and Climate Action.	MITECO	ANNUAL	environmental-standards-for-vehicles	kingdom-of-Spain

UC Winter|UC13 Nac| UC16 And|UC21 CV| UC22 Cat|UC3 Mad| UC3 Bcn|UC5 Port| UC5 Fra|UC10 Fra| UC10 Port|UC18 Fra| UC18 Port	Vehicle fleet by municipality	Information on the composition and age of the vehicle fleet, as well as on the driver population registry.	DGT	ANNUAL	vehicle-details	kingdom-of-Spain
UC Winter|UC13 Nac| UC19 Port|UC19 Fra	Electric vehicle charging infrastructure	The Open Charge Map dataset provides open and crowd-sourced information on electric vehicle charging infrastructure worldwide. It includes the location and characteristics of charging stations, such as connector types, power capacity, access conditions, and operational status, supporting the analysis and planning of electric mobility infrastructure.	Open Charge Map	CONT	availability-of-charging-points-for-electric-vehicles	Europe
UC Winter|UC13 Nac| UC5 Fra	GTFS for Interurban Public Transport Navarra	This dataset contains authorized interurban public passenger transport services managed by the Government of Navarre, provided in GTFS format.	Gobierno de Navarra	ANNUAL	public-transport-scheduled-transport	Navarra
UC Winter|UC13 Nac| UC5 Fra	Static GTFS for Interurban Public Transport Gipuzkoa	Scheduled timetables in GTFS format for the ATTG network	GeoEuskadi	WEEKLY	public-transport-scheduled-transport	Gipuzkoa
UC Winter|UC13 Nac| UC5 Fra	Real-Time GTFS for Interurban Public	Real-time timetables in GTFS-RT format for the ATTG network	GeoEuskadi	UPDATE_CONT	public-transport-scheduled-transport	Gipuzkoa

	Transport Gipuzkoa					
UC Winter|UC13 Nac| UC5 Fra	Static GTFS for Interurban Public Transport Aquitania	Scheduled timetables in GTFS/NeTEx format for the Txik Txak network	Syndicat des Mobilités Pays Basque-Adour	DAILY	public-transport-scheduled-transport	pays-Basque
UC Winter|UC13 Nac| UC5 Fra	Real-Time GTFS for Interurban Public Transport Pays Basque	Real-time timetables in GTFS-RT format for the Txik Txak network	Syndicat des Mobilités Pays Basque-Adour	UPDATE_CONT	public-transport-scheduled-transport	pays-Basque
UC Winter|UC13 Nac| UC5 Port	Static GTFS for Interurban Public Transport Extremadura	Scheduled timetables in GTFS format for the Extremadura network	NAP transporte multimodal	UNKNOWN	public-transport-scheduled-transport	Extremadura
UC Winter|UC13 Nac| UC5 Fra|UC5 Port	Static GTFS long distance FlixBus	File containing the timetables for the entire European FlixBus transport network	Flixbus France	DAILY	public-transport-scheduled-transport	Europe
UC Winter|UC13 Nac| UC5 Fra|UC5 Port	Static GTFS long distance BlaBlaCar Bus	List of stops, timetables, and routes for the entire European BlaBlaCar Bus network in GTFS format, updated daily.	BlaBlaCar Bus	DAILY	public-transport-scheduled-transport	Europe
UC Winter|UC13 Nac| UC5 Fra|UC10 Fra|	Traffic counts at detectors in Euskadi	Counts of light and heavy vehicles recorded by detectors operated by the Basque Government	Gobierno Vasco	5MIN	traffic-volume	euskal-autonomia-erkidegoa

UC18 Fra|UC19 Fra						
UC Winter|UC13 Nac	Traffic volumes on Spanish State Road Network	The Traffic Map aims to provide key information on traffic volumes and recorded accidents across the Spanish State Road Network.	MITMS	ANNUAL	traffic-volume	kingdom-of-Spain
UC Winter|UC13 Nac| UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Traffic volumes on Euskadi	Traffic density in the Basque Country during 2024 (traffic counts)	Open Data Euskadi	ANNUAL	traffic-volume	euskal-autonomia-erkidegoa
UC Winter|UC13 Nac	Traffic volumes on Navarra Network	Road segments in Navarra with data on Average Daily Traffic (ADT), Average Daily Heavy Traffic (ADHT), and the corresponding road section to which they are assigned.	IDENA	ANNUAL	traffic-volume	Navarra
UC Winter|UC13 Nac	Location of the detectors on the Navarra Network	Location of traffic counting stations used to record vehicle flows, including data on Average Daily Traffic (ADT), Average Daily Heavy Traffic (ADHT), and the road segment to which each station is assigned in Navarra.	IDENA	UNKNOWN	locations-and-stations	Navarra
UC Winter|UC13 Nac| UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Location of the detectors in Euskadi	Location of traffic counting stations used to record vehicle flows and the road segment to which each station is assigned in Euskadi.	Gobierno Vasco	UNKNOWN	locations-and-stations	euskal-autonomia-erkidegoa
UC13 Nac	Toll sections in	Location of toll sections of interest in Spanish State Road Network	MITMS	UNKNOWN	identification-of-tolled-roads	kingdom-of-Spain

	Spanish State Road Network					
UC16 And	Traffic volumes on Andalucia Network	Data from the 2024 General Traffic Count Plan in Andalucia in SHAPE and spreadsheet formats.	Junta de Andalucia	ANNUAL	traffic-volume	Andalusia
UC16 And	Historical traffic volumes on Andalucia Network	The traffic maps present, for each province, the traffic volumes on the different road segments, as well as the Average Daily Traffic (ADT) and the percentage of heavy vehicles recorded at each counting station.	Junta de Andalucia	ANNUAL	traffic-volume	Andalusia
UC16 And	Household Mobility Survey 2011 Andalucia	2011 Social Survey: Mobility in the Urban Regions of Andalusia	Junta de Andalucia	AS_NEEDED	other	Andalusia
UC16 And	Static GTFS for Interurban Public Transport Andalucia	This dataset contains authorized interurban public passenger transport services managed by the Government of Andalucia, provided in GTFS format.	Red de Consorcios de Transporte de Andalucía	DAILY	public-transport-scheduled-transport	Andalusia
UC16 And	Logistics areas in Andalucia	Geographic information related to specific facilities and infrastructures aimed at strengthening the economic and productive fabric of the Autonomous Community of Andalucia (productive spaces, CADE centers, technological spaces, logistics areas, extractive activities, etc.).	Instituto de Estadística y Cartografía de Andalucia	UNKNOWN	freight-and-logistics	Andalusia
UC16 And|UC21 CV| UC22 Cat|UC18	Survey on Road Freight	This dataset provides statistical tables from the 2024 Permanent Road Freight Transport Survey. The tables are available in PDF and Excel formats.	MITMS	ANNUAL	freight-and-logistics	kingdom-of-Spain

Fra| UC18 Port	Transport 2024					
UC16 And|UC21 CV| UC22 Cat|UC18 Fra| UC18 Port	Road freight transport flows in Spain	Domestic road freight transport by Spanish-registered vehicles (tonnes and tonne-kilometres). Origin–destination relationships by autonomous community.	MITMS	ANNUAL	freight-and-logistics	kingdom-of-Spain
UC16 And	Information on Spanish State Ports	Annual port traffic data for the main Spanish ports	MITMS	ANNUAL	freight-and-logistics	kingdom-of-Spain
UC21 CV	Delimitation of the LEZ in Valencia	Geospatial layer for delineating Low Emission Zones in the city of Valencia	Ajuntament de València	AS_NEEDED	other-access-restrictions-and-traffic-regulations	Valencia
UC21 CV	Surface area of the municipalities of the Comunitat Valenciana	This dataset defines the municipalities of the Valencian Community, based on municipal boundary lines registered in the Central Cartographic Registry (RCC) from the National Geographic Institute (IGN). It incorporates geometric improvements made by the Valencian Cartographic Institute (ICV) and the coastline from the ICV's 1:5,000 cartography.	Generalitat Valenciana	AS_NEEDED	other	comunitat-valenciana
UC21 CV|UC3 Mad| UC3 Bcn	Population in Spanish municipalities	Population by sex and municipality	INE	ANNUAL	other	kingdom-of-Spain

UC21 CV	Standardized Digital Street Map of the Valencian Community	The Valencian Standardized Digital Street Map is a comprehensive alphanumeric and geographic dataset of streets and building addresses for all municipalities in the Valencian Community. It provides continuous territorial coverage and integrates information from major public address sources (Cadastre, INE, Cartociudad, and CV05). The dataset includes streets, street segments, and building numbers, covering both urban and rural areas, and is continuously maintained and updated by the Valencian Cartographic Institute.	IDEV	ANNUAL	road-classification	comunitat-valenciana
UC21 CV	Bicycle routes in the city of Valencia	Geographical information of CarrilBici and Ciclocalle in the city of Valencia	Ajuntament de València	AS_NEEDED	cycle-network-data	Valencia
UC21 CV	Bicycle routes in the Comunitat Valenciana	Includes all non-motorized infrastructures currently in service across the Valencian territory.	Generalitat Valenciana	AS_NEEDED	cycle-network-data	comunitat-valenciana
UC21 CV	Traffic volumes on Comunitat Valenciana	2024 traffic map of the roads in the Comunitat Valenciana	Generalitat Valenciana	ANNUAL	traffic-volume	comunitat-valenciana
UC21 CV	Bycicle traffic volumes on Comunitat Valenciana	2024 bycicle traffic map of the roads in the Comunitat Valenciana	Generalitat Valenciana	ANNUAL	cycle-network-data	comunitat-valenciana

UC21 CV	Air quality monitoring stations data in the Comunitat Valencian a	Air pollutant measurements from air quality monitoring stations in the Comunitat Valenciana	Generalit at Valencian a	DAILY	parameters-needed-to-calculate-environmental-factors	comunitat-valenciana
UC21 CV	Traffic intensity by road segments in the city of Valencia	This layer describes traffic intensity across the different areas of the city of Valencia.	Ajuntame nt de València	UNKNOWN	traffic-volume	Valencia
UC21 CV	Static GTFS for interurban bus services in the Comunitat Valencian a	These CSV and GTFS files contain all routes and itineraries operated by the different interurban public bus services concessioned by the Generalitat Valenciana.	Generalit at Valencian a	DAILY	public-transport-scheduled-transport	comunitat-valenciana
UC21 CV	Static GTFS for urban bus services in city of Valencia	Regular urban bus services in the city of Valencia	Ajuntame nt de València	DAILY	public-transport-scheduled-transport	Valencia
UC21 CV	Static GTFS for urban bus services in Castelló de la Plana	Regular urban bus services in the city of Castelló de la Plana	Ajuntame nt de Castelló de la Plana	UNKNOWN	public-transport-scheduled-transport	Castello-de-la-plana

UC21 CV	Static GTFS for urban bus services in the city of Alicante	Regular urban bus services in the city of Alicante	Vectalia Movilidad	DAILY	public-transport-scheduled-transport	Alicante
UC21 CV	Spanish national logistic nodes	The Reference Geographic Information Transport Network is a three-dimensional, intermodal network with national coverage, defined and published in accordance with the INSPIRE Directive, and updated within the framework of cooperation of SCNE.	SCNE	AS_NEEDED	network-detailed-attributes	kingdom-of-Spain
UC21 CV	Parking facilities in the city of Valencia	This dataset provides information on parking facilities in the city of Valencia, including their locations and key characteristics.	Ajuntament de València	AS_NEEDED	car-parking-locations-and-conditions	Valencia
UC21 CV	Bicycle parking facilities in the city of Valencia	This dataset provides information on bike parking facilities in the city of Valencia, including their locations and key characteristics.	Ajuntament de València	AS_NEEDED	bike-parking-locations	Valencia
UC21 CV	Location of loading and unloading zones in the city of Valencia	This dataset provides the geographic location of designated loading and unloading zones in the city of Valencia.	Ajuntament de València	AS_NEEDED	availability-of-delivery-areas	Valencia
UC22 Cat|UC3 Bcn	Daily Mobility Survey in the province of Barcelona	The EMEF analyses weekday mobility patterns of residents aged 16 and over in the province of Barcelona. It also maintains the historical data series for the Barcelona Metropolitan Region (RMB) and the city of Barcelona.	OMC	ANNUAL	other	provincia-de-Barcelona

UC22 Cat	Traffic volumes on Catalonia Road Network	Traffic counting systems are used to measure traffic intensity and record the volume of vehicles passing along the roads where they are installed. These systems make it possible to calculate the Average Daily Traffic (ADT), which represents the average number of vehicles passing a given point on a road each day.	GENCAT	ANNUAL	traffic-volume	Catalunya
UC22 Cat	GHG emissions in Catalonia	This dataset provides annual data on greenhouse gas (GHG) emissions by gas and by sector, division, class, and subclass, following the IPCC emissions inventory methodology.	GENCAT	UNKNOWN	environmental-standards-for-vehicles	Catalunya
UC22 Cat	Main atmospheric pollutants in Catalonia	This dataset provides information on the main atmospheric pollutants in Catalonia, based on measurements collected across the region.	IDESCAT	ANNUAL	environmental-standards-for-vehicles	Catalunya
UC22 Cat	Emergency parking location and capacity in Catalonia	Truck parking areas concessioned by the Department of Territory, Housing and Ecological Transition of Catalonia	GENCAT	UNKNOWN	parking-service-and-rest-area-information; truck-parking-locations-and-conditions	Catalunya
UC22 Cat	Road freight transport with origin or destination in Catalonia	This dataset provides information on road freight transport flows with origin or destination in Catalonia.	IDESCAT	ANNUAL	freight-and-logistics	Catalunya
UC22 Cat|UC3 Bcn	GTFS for public transport	Aggregates all public transport supply information in Catalonia	T-Mobilitat	DAILY	public-transport-scheduled-transport	Catalunya

	in Catalonia					
UC3 Bcn	Traffic volumes by road in the city of Barcelona	Average Daily Traffic (ADT) on weekdays at different traffic counting points across the road network of the city of Barcelona.	ATM	ANNUAL	traffic-volume	Barcelona
UC3 Bcn	Basic demand data in the city of Barcelona	This dataset provides basic demand data for the Barcelona metropolitan area	TMB	ANNUAL	other	Barcelona
UC3 Bcn|UC3 Mad	Employment contracts registered by Spanish municipalities	Statistical data from SEPE on registered employment contracts, disaggregated by municipality, sex, type of contract, and economic activity sector.	SEPE	ANNUAL_2	other	kingdom-of-Spain
UC3 Bcn|UC3 Mad	Registered unemployment by Spanish municipalities	Statistical data from SEPE on registered unemployment, disaggregated by municipality, sex, age, and economic activity sector.	SEPE	ANNUAL_2	other	kingdom-of-Spain
UC3 Bcn	Urban planning map of Catalonia	This dataset provides the urban planning map of Catalonia, showing land-use classifications and planning designations across the territory.	GENCAT	ANNUAL	other	Catalunya
UC3 Bcn	FGC Park-and-ride facilities	This dataset identifies the FGC stations where Park-and-Ride facilities are available.	FGC	CONT	park-and-ride-stops	provincia-de-Barcelona
UC3 Bcn	AMB Park-and-ride facilities	Park-and-ride facilities located near public transport hubs (e.g. metro, tram, bus) supported by AMB.	AMB	AS_NEEDED	park-and-ride-stops	provincia-de-Barcelona

UC3 Bcn	AMBici GBFS	Standardized GBFS dataset providing static and real-time information on the AMBici bike-sharing system, including station locations, vehicle availability, pricing plans, and operational restrictions.	AMB	CONT	bike-sharing-locations-and-stations	provincia-de-Barcelona
UC3 Bcn	Bicing Barcelona GBFS	Standardized GBFS dataset providing static and real-time information on the Bicing bike-sharing system, including station locations, vehicle availability, pricing plans, and operational restrictions.	Ajuntament de Barcelona	CONT	bike-sharing-locations-and-stations	provincia-de-Barcelona
UC3 Mad	Static GTFS for the Madrid Interurban Bus Network	Dataset containing public transport schedules and the associated geographic information for the Interurban Bus Network of the Community of Madrid.	CRTM	AS_NEEDED	public-transport-scheduled-transport	comunidad-de-Madrid
UC3 Mad	Static GTFS for the Madrid Metro Network	Dataset containing public transport schedules and the associated geographic information for the Metro network.	CRTM	AS_NEEDED	public-transport-scheduled-transport	comunidad-de-Madrid
UC3 Mad	Static GTFS for the Light Rail / Tram Network	Dataset containing public transport schedules and the associated geographic information for the Light Rail / Tram network.	CRTM	AS_NEEDED	public-transport-scheduled-transport	comunidad-de-Madrid
UC3 Mad	Static GTFS for the Commuter Rail Network	Dataset containing public transport schedules and the associated geographic information for the Commuter Rail network.	CRTM	AS_NEEDED	public-transport-scheduled-transport	comunidad-de-Madrid
UC3 Mad	Static GTFS for the Urban Bus Network	Dataset containing public transport schedules and the associated geographic information for the Urban Bus Network of the Community of Madrid.	CRTM	AS_NEEDED	public-transport-scheduled-transport	comunidad-de-Madrid

UC3 Mad	Static GTFS for the EMT Network	Dataset containing public transport schedules and the associated geographic information for Madrid's Urban Bus Network (EMT).	CRTM	AS_NEEDED	public-transport-scheduled-transport	Madrid
UC3 Mad	CRTM Daily Demand Evolution by mode	Dataset on the evolution of demand by transport mode for the CRTM	CRTM	DAILY	other	comunidad-de-Madrid
UC3 Mad	Land use in the Community of Madrid	Information on land classification from the original, currently valid general urban planning of each municipality in the Community of Madrid, including incorporated planning.	IDEM	AS_NEEDED	other	comunidad-de-Madrid
UC3 Mad	Traffic volume in the Community of Madrid	The study focuses on obtaining the Average Daily Traffic (ADT), expressing the number of vehicles that travel on average per day along different road sections.	Comunidad de Madrid	ANNUAL	traffic-volume	comunidad-de-Madrid
UC3 Mad	Park-and-Ride facilities in the city of Madrid	Dataset containing information on municipal public Park-and-Ride parking facilities in the city of Madrid	EMT	AS_NEEDED	park-and-ride-stops	Madrid
UC3 Mad	Park-and-Ride facilities in the city in the Community of Madrid	Parking facilities in the Community of Madrid linked to public transport	CRTM	AS_NEEDED	park-and-ride-stops	comunidad-de-Madrid
UC3 Mad	Bicimad GBFS	File compiling the General Bikeshare Feed Specification (GBFS) of the municipal Bicimad service.	EMT	AS_NEEDED	bike-sharing-locations-and-stations	Madrid

UC3 Mad	Bicycle Sharing Systems in the Community of Madrid	Bicycle Sharing Systems and Public Transport in the Community of Madrid	CRTM	AS_NEEDED	bike-sharing-locations-and-stations	comunidad-de-Madrid
UC3 Mad	2018 Household Mobility Survey of the Community of Madrid	Result tables from the Household Mobility Survey of the Community of Madrid (EDM2018) by various variables: households, individuals, trips, and stages.	CRTM	UNKNOWN	other	comunidad-de-Madrid
UC3 Mad	Air quality real-time data in the city of Madrid	The Madrid City Council's Integrated Air Quality System allows real-time monitoring of atmospheric pollution levels across the municipality.	CRTM	CONT	environmental-standards-for-vehicles	Madrid
UC5 Port|UC10 Port| UC18 Port|UC19 Port	Transport Networks of Portugal	Transport networks at a 1:200,000 scale for Portugal. The main features included in this dataset are the railway network, the road network, airports, and maritime transport.	Direção-Geral do Território	UNKNOWN	network-geometry-and-lane-character	Portuguese-republic
UC5 Port|UC10 Port| UC18 Port|UC19 Port	Traffic intensity at Extremadura network 2023	This dataset provides traffic intensity data for the road network in Extremadura for the year 2023.	Junta de Extremadura	ANNUAL	traffic-volume	Extremadura
UC5 Port|UC10 Port| UC18 Port|UC19 Port	Vehicle fleet by municipality in Portugal	This dataset provides the distribution of the insured vehicle fleet by vehicle circulation district, vehicle category, year of manufacture and circulation district, year of manufacture and vehicle category, and municipality of circulation and vehicle category.	ASF	ANNUAL	vehicle-details	Portuguese-republic

UC5 Port|UC10 Port| UC18 Port|UC19 Port	Secure parking facilities inventory Spain	This dataset includes facilities that have reported the initiation of the certification procedure for secure parking areas.	MITMS	AS_NEEDED	parking-service-and-rest-area-information; truck-parking-locations-and-conditions	kingdom-of-Spain
UC18 Port	Meteorological data Portugal	This dataset provides access to different sources of meteorological information in Portugal.	IPMA	CONT	road-weather-conditions	Portuguese-republic
UC10 Port|UC18 Port	Freight transported by foreign vehicles in Portugal	Tonne-kilometres (tkm) of foreign heavy goods vehicles in Portugal	INE Portugal	AS_NEEDED	freight-and-logistics	Portuguese-republic
UC10 Port|UC18 Port| UC19 Port	Spatial Distribution of National Emissions	This dataset provides a spatially disaggregated distribution of Portugal's national emissions inventory	APA	UNKNOWN	environmental-standards-for-vehicles	Portuguese-republic
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Transport Networks of France	A homogeneous vector description of the different landscape elements in France with decametric accuracy.	IGN France	UNKNOWN	network-geometry-and-lane-character	republic-of-France
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Road traffic on the main corridors in Nouvelle-Aquitaine	This dataset provides road traffic data for the main transport corridors in Nouvelle-Aquitaine	DREAL Nouvelle-Aquitaine	UNKNOWN	traffic-volume	nouvelle-Aquitaine
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Annual Average Daily Traffic on the French	The Annual Average Daily Traffic (AADT) for a road section is calculated as the yearly average number of vehicles traveling on that section per day, in both directions. These traffic data are typically	Ministère de la Transition écologique	ANNUAL	traffic-volume	republic-of-France

	national road network	accompanied by an estimated percentage of heavy goods vehicles.				
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Static and real-time GTFS feeds for railways in France	For regional public rail services (TER), national public services (Intercités), and commercial services (TGV).	SNCF	CONT	public-transport-scheduled-transport	republic-of-France
UC5 Fra|UC10 Fra| UC18 Fra	Vehicle fleet by municipality in France	Detailed data on the French vehicle fleet as of January 1 for the years 2011–2025, broken down by vehicle type, fuel/energy, Crit'Air sticker, region/department/municipality, vehicle age, gross vehicle weight (MMA), user's professional activity, and annual distance travelled.	SDES	ANNUAL	vehicle-details	republic-of-France
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Heavy goods vehicle parking areas in France	This dataset provides static data (ID, location, total parking spaces, price) for truck parking areas located on France's concession motorways.	Bison Futé	AS_NEEDED	parking-service-and-rest-area-information; truck-parking-locations-and-conditions	republic-of-France
UC5 Fra|UC10 Fra| UC18 Fra|UC19 Fra	Emission factors for different road transport modes in France	This dataset provides pollutant emission factors in g/km for all road vehicle categories (passenger cars, light commercial vehicles, heavy-duty vehicles, and motorized two-wheelers), with multiple subcategories and a wide range of traffic conditions (road type, speed limit, and traffic congestion level).	ADEME	UNKNOWN	environmental-standards-for-vehicles	republic-of-France
UC10 Fra|UC18 Fra	Freight by mode and origin	Tonne-kilometres of freight transported by mode (rail, inland waterways, road), by region of origin in France	Ecolab	UNKNOWN	freight-and-logistics	republic-of-France

	region in France					
UC10 Fra|UC18 Fra	Data on road freight transport (TRM) in France and Europe.	Road freight transport (TRM) data comes from surveys conducted in EU countries.	SDES	UNKNOWN	freight-and-logistics	republic-of-France
UC18 Fra	Meteorological data France	This dataset provides access to different sources of meteorological information in France.	Meteo France	CONT	road-weather-conditions	republic-of-France

Table 8. Collection of metadata from identified datasets.

3 Standards and reference frameworks for data harmonisation

While the previous section focused on the semantic representation of metadata through the mobilityDCAT-AP ontology, data interoperability within DS4MOBMODELS also requires the harmonisation of the data itself. In the context of a data space, data harmonisation plays a key role in generating added value from data originating from heterogeneous sources, as it enables datasets to be combined, enriched, and reused in a coherent manner. This added value is a fundamental enabler of advanced services and analytics and directly contributes to the sustainability and business models of the data space.

Data harmonisation underpins the creation of interoperable and high-value data products by addressing syntactic, structural, and quantitative aspects of the data, such as formats, spatial representations, units of measurement, temporal references, and naming and typing conventions. By applying common standards at these levels, the project reduces integration effort, improves data quality and usability, and enables the development of advanced analytics and services that support the business models of the data space.

3.1 Data formats for mobility data harmonisation

The harmonisation of data formats is a fundamental prerequisite for interoperability within a data space. In the mobility domain, datasets are produced by heterogeneous actors using different technologies and internal data representations. Without a minimum level of alignment at the format level, the integration and reuse of data across systems and use cases becomes costly and error prone.

In the context of a Mobility Data Space, the adoption of common and well-established data formats facilitates data ingestion, validation, transformation, and combination, thereby increasing the overall value of the data. This section describes the main data formats commonly used for mobility datasets, focusing on open, standardised, and widely supported formats that enable syntactic interoperability and scalable data processing.

3.1.1 Tabular and analytical data formats

Many mobility datasets are inherently tabular in nature, such as traffic counts, parking occupancy statistics, vehicle flows, or aggregated mobility indicators. For these datasets, tabular formats provide a simple and widely supported means of data exchange and processing.

Comma-Separated Values (CSV) is commonly used for publishing tabular mobility data due to its simplicity and broad tool support. Its structure is defined by RFC 4180, and it is particularly suitable for static or periodically updated datasets intended for wide dissemination and human inspection. To support interoperability, CSV files should follow consistent conventions regarding delimiters, character encoding (UTF-8), and column naming.

For large-scale datasets or analytical use cases, columnar formats offer significant advantages. **Apache Parquet** is an open columnar storage format designed to optimise storage efficiency and query performance. While CSV is well suited for data exchange and transparency, Parquet is more appropriate for high-volume mobility datasets that are consumed by data processing pipelines, simulation models, or machine learning applications.

Within a Mobility Data Space, these formats are complementary rather than alternative. The same dataset may be published in CSV for exchange and inspection purposes, and in Parquet for analytical exploitation, while maintaining semantic consistency through aligned schemas and shared metadata definitions.

3.1.2 Structured and semi-structured data formats

Mobility datasets are increasingly exchanged through APIs and event-based systems, where structured or semi-structured formats are required. These formats are particularly relevant for real-time services, regulatory reporting, and interoperable data exchange between transport authorities and mobility platforms.

JavaScript Object Notation (JSON) is widely used for representing structured mobility data, particularly in web services and real-time data feeds. JSON provides a lightweight, hierarchical structure based on key–value pairs and nested objects, making it well suited for RESTful APIs and streaming interfaces. However, as JSON itself does not enforce a schema by default, its effective use in a data space context requires the definition of clear schemas (e.g. JSON Schema) and consistent naming and typing conventions to ensure interoperability across providers.

For XML-based data exchange, several mobility-specific standards rely on formally defined XML schemas (XSD), which precisely define the structure, data types, and relationships between elements. **DATEX II**, for example, is structured around a well-defined information model for road traffic and travel data, organised into packages covering traffic conditions, incidents, infrastructure status, and measurement data. Its XML schema enforces strict validation rules and supports complex hierarchical relationships between traffic events, locations, and temporal information.

Similarly, **NeTEx** defines a comprehensive and modular XML-based model for public transport networks, timetables, stops, routes, and fares. Its structure is based on reusable components and supports complex multimodal transport systems, enabling consistent representation of static public transport information across operators and regions. **SIRI**, closely aligned with NeTEx, focuses on real-time public transport information and defines structured XML messages for vehicle monitoring, service updates, and estimated timetables.

The choice between JSON and XML often depends on the use case: JSON is typically preferred for lightweight APIs and real-time services, while XML remains relevant for formal data exchange standards and regulatory contexts.

3.1.3 Geospatial data formats

Spatial information is a core component of mobility data, as datasets often refer to transport networks, infrastructure elements, locations, routes, or geographic areas. To ensure spatial interoperability across systems and platforms, geospatial data formats should be based on **recognised international standards** and support consistent handling of geometries, coordinate reference systems, and spatial attributes.

GeoJSON, standardised as **RFC 7946**, is a widely adopted format for representing geographic features and their attributes using JSON. It is particularly suitable for lightweight web-based applications, APIs, and data portals, and is commonly used to publish mobility-related spatial data such as parking locations, charging points, traffic sensors, or zones. GeoJSON benefits from broad tool support and ease of integration in web environments, but it is primarily designed for relatively simple geometries and datasets of moderate size.

For more complex or large-scale geospatial datasets, standards developed by the **Open Geospatial Consortium (OGC)** are often preferred. **Geography Markup Language (GML)**, formalised as **ISO 19136**, provides a rich and expressive XML-based encoding for geographic information and is widely used in infrastructure-related datasets and INSPIRE-compliant services. GML supports complex geometries, coordinate reference systems, and feature relationships, making it suitable for detailed representations of transport networks and infrastructure, albeit at the cost of higher verbosity.

In addition to exchange-oriented formats, **file-based geospatial formats** remain highly relevant in mobility workflows, particularly for data preparation, analysis, and integration in Geographic Information Systems (GIS). The **ESRI Shapefile** format, while not an ISO standard, is a long-established de facto

standard supported by virtually all GIS tools and widely used to represent vector-based transport networks, road segments, and spatial assets. However, Shapefile has known technical limitations, such as file fragmentation and attribute constraints.

To address these limitations, **GeoPackage**, standardised as **OGC GeoPackage (OGC 12-128r18)**, has emerged as a modern and robust alternative. GeoPackage is an open, standards-based format built on SQLite that supports vector and raster data, multiple layers, spatial indexing, and metadata within a single file. GeoPackage is particularly well suited for large mobility datasets, offline use cases, and scenarios where data need to be exchanged or processed efficiently across different tools and platforms.

Within a Mobility Data Space, these geospatial formats are complementary rather than mutually exclusive. **GeoJSON and GML** are typically used for interoperable data exchange and API-based access, while **Shapefile and GeoPackage** support operational GIS workflows and large-scale spatial analysis. By supporting multiple geospatial representations aligned with common semantic models and metadata descriptions, the data space can accommodate diverse usage scenarios while maintaining spatial interoperability.

3.1.4 Domain-specific formats for mobility

In addition to generic data formats, the mobility sector relies on a number of **domain-specific formats** that define not only how data are encoded, but also how mobility concepts are structured and interpreted. These formats typically encapsulate both syntactic and semantic harmonisation, making them particularly valuable for interoperability within specific transport subdomains.

In the public transport domain, the **General Transit Feed Specification (GTFS)** has become a widely adopted de facto standard for publishing static timetable data, including routes, stops, calendars, and fare information. GTFS is based on a bundle of CSV files with a predefined structure and shared semantics, enabling straightforward processing while ensuring consistent interpretation across systems. Together with **GTFS-Realtime**, which defines a protocol for exchanging real-time service updates, GTFS supports a broad range of journey planning, passenger information, and operational use cases.

Complementing GTFS at a more formal and comprehensive level, **NeTEx (CEN/TS 16614)** is a European standard that defines an XML-based data model and exchange format for public transport networks, timetables, and fares. NeTEx supports complex public transport structures, multimodal networks, and detailed fare models, and is closely aligned with other CEN standards such as SIRI.

In the road transport and traffic management domain, **DATEX II (CEN/TS 16157)** provides a standardised framework for the exchange of traffic and travel information. DATEX II defines both a conceptual data model and an XML-based encoding, covering aspects such as traffic conditions, incidents, roadworks, and infrastructure status. Its widespread adoption by European road authorities makes it a key standard for harmonising traffic-related datasets and enabling interoperable traffic information services.

For shared and active mobility services, **GBFS (General Bikeshare Feed Specification)** has emerged as the de facto standard for publishing information about bike-sharing systems, and increasingly other shared mobility services such as e-scooters. GBFS is based on JSON and RESTful APIs and provides standardised feeds for station information, vehicle availability, and system status. Its lightweight structure and real-time orientation make it particularly suitable for integration into mobility platforms and data spaces supporting multimodal and shared mobility use cases.

3.2 Units of measure and quantitative harmonisation

Mobility datasets frequently include quantitative values such as distances, speeds, travel times, traffic volumes, timestamps, or spatial references. To enable meaningful comparison, aggregation, and integration of such data across sources and use cases, these values should be expressed using

harmonised units, formats, and reference systems. Quantitative harmonisation is therefore a key enabler of interoperability and reliable data analysis within a Mobility Data Space.

This section presents the main standards applied to ensure consistency in the representation of numerical values, temporal information, and spatial references in mobility datasets.

3.2.1 Units of measurement

The representation of physical quantities in mobility data—such as length, speed, time, energy, or emissions—requires the use of consistent and internationally recognised units of measure.

The **International System of Units (SI)**, standardised under ISO 80000, provides the reference framework for expressing physical quantities in a harmonised manner. The adoption of SI units ensures that numerical values can be correctly interpreted and compared across datasets, independently of their origin or the systems producing them. Typical examples in the mobility domain include metres or kilometres for distance, seconds or hours for time, and metres per second or kilometres per hour for speed.

For machine-readable representations and automated processing, the **Unified Code for Units of Measure (UCUM)** is often used in conjunction with SI units. UCUM defines unambiguous, standardised codes for units of measure, enabling systems to automatically interpret, validate, and convert quantitative values when required. This is particularly relevant in data spaces, where datasets may be combined and processed by heterogeneous platforms and analytical tools.

3.2.2 Temporal information and date-time representation

Time is a critical dimension of mobility data, as datasets frequently include timestamps, temporal coverage, update frequencies, and validity periods. To avoid ambiguities in the interpretation of temporal information, harmonised date and time representations are essential.

ISO 8601 defines an internationally recognised standard for representing dates, times, and durations in a machine-readable format. Its use ensures consistent interpretation of temporal values across datasets and systems, including the handling of time zones and daylight saving time. ISO 8601 is widely adopted in mobility datasets for timestamps, observation times, and temporal coverage indicators.

For example, a timestamp indicating the moment when a traffic measurement was recorded can be expressed as 2026-02-05T08:30:00Z, where the date, time, and time zone are unambiguously defined.

3.2.3 Spatial reference systems and geographic coordinates

Quantitative harmonisation also extends to the representation of geographic coordinates and spatial reference systems. Mobility datasets often include point locations, linear geometries, or spatial extents that must be interpreted consistently across platforms and across national boundaries.

The use of standardised **Coordinate Reference Systems (CRS)** ensures that geographic coordinates are expressed within a well-defined spatial framework. At international level, **WGS 84 (EPSG:4326)** is the most commonly used CRS for mobility data, as it is the reference system employed by GPS devices, web-based mapping services, and many geospatial APIs. Its widespread adoption makes it the de facto standard for exchanging mobility-related geographic information, particularly in cross-border and web-based contexts.

In national and regional contexts, more specific projected coordinate reference systems are often used to support higher spatial accuracy and local data processing needs. In Spain, the official geodetic reference system is **ETRS89**, and mobility datasets are commonly provided using projected CRS such as **ETRS89 / UTM zone 30N (EPSG:25830)** or, depending on the geographic area, **ETRS89 / UTM zone 29N or 31N**. These CRS are defined under European and national geodetic standards and are widely used by public authorities for transport and infrastructure data.

OGC and ISO standards within the **ISO 19100 series** define common principles and mechanisms for the representation, transformation, and documentation of spatial reference systems. By clearly specifying the CRS used and, where necessary, providing transformations between reference systems, mobility datasets can be integrated and analysed consistently, ensuring spatial interoperability within the Mobility Data Space.

3.3 Field naming conventions

In addition to harmonising data formats, units of measurement, and reference systems, data interoperability also depends on the consistent naming of data fields within datasets. Field naming conventions play an important role in ensuring that datasets from different sources can be easily understood, combined, and processed without requiring extensive data preparation or manual interpretation.

Within the context of a Mobility Data Space, it is recommended to adopt **simple, consistent, and machine-friendly naming conventions** for dataset fields. Field names should be expressed using lowercase characters, avoid spaces and special characters, and follow a consistent word-separation strategy such as **snake_case** (e.g. `traffic_volume`, `average_speed`, `vehicle_type`). This convention is widely supported across data processing tools and programming languages and facilitates automated data handling.

Field names should be **self-descriptive** and aligned, where possible, with the semantic definitions used in metadata models and domain standards. Abbreviations should be avoided unless they are widely understood within the mobility domain, and pluralisation should be applied consistently for fields representing collections or counts.

3.4 Data types and value representation

In addition to harmonising data formats, units of measurement, and field names, interoperability also depends on the **consistent representation of data types and values** within datasets. Differences in how numeric, textual, temporal, or boolean values are encoded can significantly hinder data integration and automated processing, even when datasets share the same structure and semantics.

Within the context of a Mobility Data Space, it is recommended to adopt clear and consistent conventions for data types. Numeric values representing measurements, counts, or quantities should be encoded as numbers rather than strings, using a dot (`.`) as decimal separator in accordance with international conventions. Textual values should be encoded as strings using UTF-8 character encoding, while boolean values should be represented using explicit and consistent values (e.g. `true` / `false`).

Special attention should be given to the representation of **missing or unknown values**. The use of explicit null values or empty fields is preferred over placeholder values such as `-1`, `9999`, or textual markers (e.g. `"N/A"`), which may introduce ambiguity. Temporal values should follow ISO 8601 formats, and numeric values should be expressed using harmonised units as described in Section 3.2. By applying consistent data type and value representation rules, datasets from different providers can be more easily validated, combined, and reused within the data space.

4 Harmonisation approach within the Mobility Data Space

The harmonisation approach adopted in DS4MOBMODELS is designed in line with the operational principles of a Mobility Data Space, where data are ingested directly from heterogeneous source systems and transformed to generate **added value**. Rather than enforcing a single data model or format at source level, the project follows a principle of **non-intrusive ingestion**, preserving the original datasets while applying harmonisation processes as a distinct and value-generating layer.

The primary objective of harmonisation is to enable datasets from different providers to be **combined, compared, and reused** in a consistent manner. By aligning data formats, field names, units of measurement, reference systems, and semantic classifications, harmonisation reduces integration effort and increases the usability of data across organisational and technical boundaries. In this context, harmonised datasets constitute enhanced data products that go beyond the original raw data and form a core asset of the data space.

A key principle of the approach is that harmonisation is **use-case driven**. The use cases defined within the project represent concrete scenarios where value can be extracted from data once interoperability barriers are removed. As such, harmonisation activities are prioritised and scoped based on the requirements of these use cases, ensuring that the applied transformations directly contribute to the creation of value and to the sustainability of the data space.

4.1 Harmonisation workflow and processing stages

The harmonisation process within DS4MOBMODELS follows a structured workflow that reflects the lifecycle of data within a data space, from ingestion to reuse. This workflow is designed to support both syntactic and semantic alignment while maintaining traceability between original and harmonised data.

The first stage consists of **data ingestion**, where datasets are collected directly from source systems in their native formats. At this stage, no transformations are applied, ensuring that original data remain unchanged and can be referenced when needed. This is followed by an **initial validation and profiling phase**, during which datasets are analysed to identify formats, structures, value domains, and potential inconsistencies.

Subsequently, **syntactic harmonisation** is applied, addressing aspects such as format normalisation, field naming conventions, data type consistency, unit conversion, and the use of common spatial reference systems. These activities operationalise the harmonisation principles and standards described in **Section 3**, ensuring that datasets originating from heterogeneous sources can be processed uniformly by downstream systems and tools within the data space.

Building on this foundation, **semantic harmonisation** focuses on aligning data values and structures with shared vocabularies, classifications, and reference frameworks, such as controlled code lists, geographic gazetteers, and domain-specific taxonomies. This step directly builds upon the interoperability framework and ontology-based metadata model described in **Section 2**, enabling consistent interpretation of data across sources and supporting cross-dataset analysis, integration, and reuse within the Mobility Data Space.

Finally, harmonised datasets are made available within the data space as **enhanced data products**, ready to be consumed by applications, analytics, and services. Throughout the workflow, traceability is maintained between harmonised outputs and original source data, supporting transparency, governance, and future evolution of the harmonisation processes.

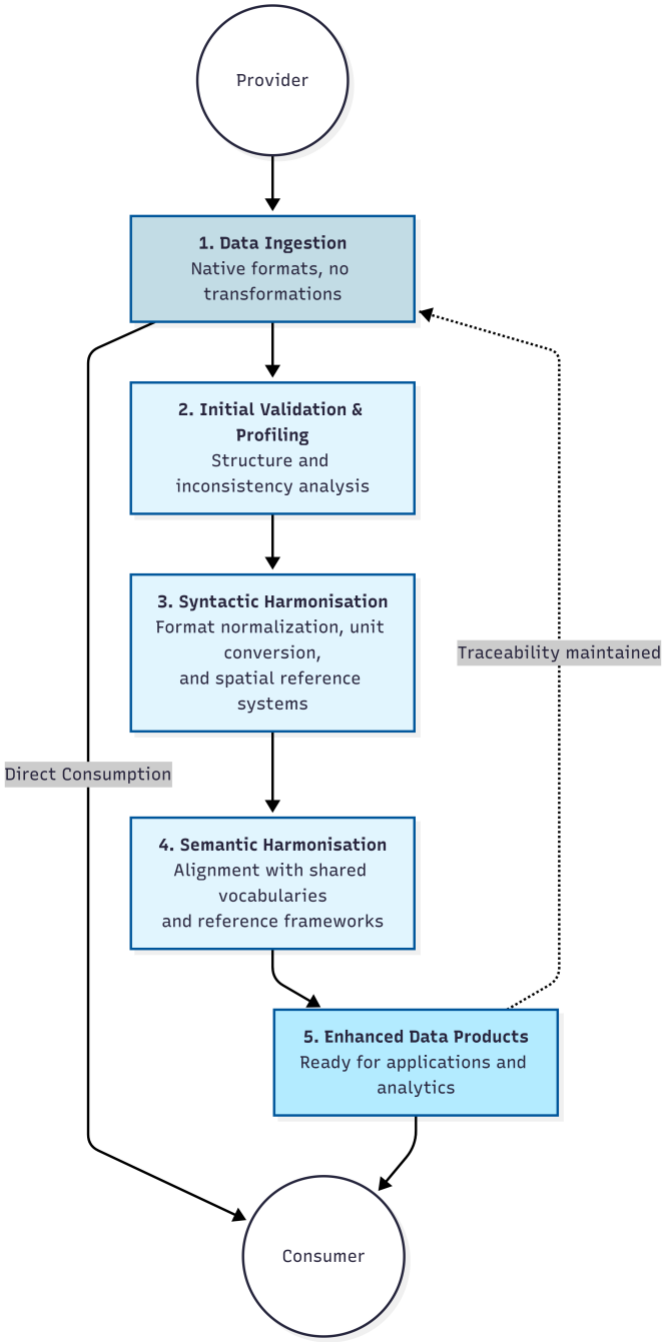


Figure 3. Harmonisation workflow schema

4.2 Challenges and constraints

The interoperability framework defined in this deliverable is designed to operate in a context characterised by heterogeneous data sources, varying levels of standardisation, and differing degrees of technical and organisational maturity among data providers. In a Mobility Data Space where datasets are ingested directly from source systems, interoperability cannot be assumed as a given condition; rather, it must be understood as a progressive objective shaped by existing constraints and structural differences across the mobility ecosystem.

One of the primary challenges identified in previous phases of the project is **data heterogeneity**. The datasets considered within DS4MOBMODELS originate from different organisations, domains, and technical environments. They are produced using diverse formats (e.g. CSV, JSON, XML-based standards), distinct structural models, different coordinate reference systems, heterogeneous units of measurement, and inconsistent naming conventions. In addition, datasets may vary in granularity, temporal resolution, and update frequency. This structural diversity increases integration complexity and may limit the extent to which datasets can be directly combined or analysed without prior alignment.

A second challenge relates to **incomplete or uneven standardisation** across data sources. While some datasets adhere to established sectoral standards such as DATEX II, NeTEx, or GBFS, others rely on locally defined schemas, partial implementations of standards, or legacy data models developed for internal operational purposes. Even when standards are referenced, their implementation may differ in scope or depth, leading to inconsistencies in optional elements, code lists, or structural components. This uneven adoption of standards creates interoperability gaps that cannot be resolved solely through formal specification alignment.

A further constraint concerns the **differing levels of data maturity and quality** among providers. Some organisations maintain well-documented datasets with comprehensive metadata, stable update cycles, and clearly defined governance procedures. Others may provide datasets with limited documentation, incomplete attributes, ambiguous definitions, or irregular update mechanisms. Variations in data maturity affect not only technical integration but also the reliability and interpretability of the information made available within the data space.

Additional constraints may arise from organisational or operational factors, such as limited technical capacity, legacy systems, or evolving data governance practices. Recognising these challenges is essential to situate the harmonisation approach within realistic operational conditions.

4.3 Added value datasets

Within a Mobility Data Space, harmonisation is not merely a technical necessity but a **key mechanism for value creation**. By transforming heterogeneous datasets into interoperable and semantically aligned data products, harmonisation enables new forms of data reuse that would otherwise be costly or infeasible. This added value is central to the economic and operational viability of the data space.

The use cases defined in the project represent concrete expressions of this value creation. Each use case demonstrates how harmonised data can support advanced analytics, decision-making processes, or innovative mobility services that rely on the integration of data from multiple sources. In this sense, use cases are not independent of the harmonisation approach but are the **direct outcome of the value unlocked through harmonised data**.

By reducing data preparation effort, enabling cross-domain analysis, and supporting scalable reuse, harmonisation contributes to the development of sustainable data-driven services. These services, in turn, underpin the business models of the data space, whether through improved operational efficiency, enhanced public services, or the creation of new commercial opportunities. As the data space evolves, the harmonisation framework established in DS4MOBMODELS provides a foundation for continuously generating value through the onboarding of new datasets and the expansion of use cases.

5 Conclusions and next steps

This deliverable has defined the interoperability and harmonisation framework to be applied within DS4MOBMODELS, establishing a coherent approach to metadata modelling, data harmonisation standards, and harmonisation processes in the context of a Mobility Data Space. The analysis has shown that semantic interoperability, supported by mobilityDCAT-AP, and data harmonisation, based on internationally recognised standards, are key enablers for integrating heterogeneous mobility datasets and transforming them into interoperable and reusable data products.

A central insight of this work is that harmonisation activities constitute a value-generating layer within the data space. By ingesting data directly from source systems and applying syntactic and semantic transformations, the project creates enhanced datasets that underpin the defined use cases and contribute to the sustainability and business model of the Mobility Data Space. The use cases themselves represent concrete expressions of the value unlocked through interoperable and harmonised data.

The framework and processes defined in this deliverable provide the foundation for the subsequent activities under **Task 3.4**, which focuses on data quality assurance in accordance with the **FAIR principles**. Building on the harmonised data structures and reference standards established here, Task 3.4 will define and implement data quality pipelines addressing validation, cleaning, consistency checks, and bias assessment, ensuring that datasets published through the Mobility Data Space are findable, accessible, interoperable, and reusable.

These activities will be consolidated in **Deliverable D3.3 – Data quality monitoring and assessment plan**, which will specify the quality checks, metrics, and pipelines to be applied to each data source. Together, Deliverables D3.2 and D3.3 form a coherent progression from interoperability and harmonisation design to operational data quality assurance, supporting the publication of high-value, trustworthy mobility datasets and enabling their effective reuse by actors across the mobility ecosystem.

6 References

Technical Standards and Ontologies:

- mobilityDCAT-AP: The reference metadata model and ontology profile for the project. https://w3id.org/mobilitydcat-ap
- ISO 8601: International standard for date and time representation. https://www.iso.org/iso-8601-date-and-time-format.html
- DATEX II (CEN/TS 16157): Standard for the exchange of traffic and travel information. https://datex2.eu/
- NeTEx (CEN/TS 16614): European standard for public transport network and timetable exchange. https://netex-cen.eu/
- GTFS (General Transit Feed Specification): De facto standard for static timetable data. https://gtfs.org/
- GBFS (General Bikeshare Feed Specification): Standard for shared mobility service information. https://gbfs.org/
- WGS 84 (EPSG:4326): Global Coordinate Reference System for mobility data. https://epsg.io/4326
- ETRS89 / UTM zone 30N (EPSG:25830): Official geodetic reference system for Spanish datasets. https://epsg.io/25830
- OGC GeoPackage (OGC 12-128r14): Standards-based format for exchanging vector and raster spatial data. https://www.geopackage.org/
- Apache Parquet: Columnar storage format used for high-volume analytical datasets. https://parquet.apache.org/
- UCUM (Unified Code for Units of Measure): Standard for unambiguous codes for units of measure. https://unitsofmeasure.org/

Legal and Policy Frameworks:

- INSPIRE Directive (2007/2/EC): European directive establishing an infrastructure for spatial information. https://knowledge-base.inspire.ec.europa.eu/legislation/inspire-directive_en
- FAIR Principles: Guiding principles to ensure data is Findable, Accessible, Interoperable, and Reusable. https://ec.europa.eu/isa2/actions/measuring-maturity-data_en/