Ref. Ares(2026)4529801 - 04/05/2026

Data Space 4 Mobility Models integration

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

Deliverable number:	**D3.3**
Deliverable name:	**Data quality monitoring and assessment plan**
Work Package number:	**WP3**
Due date:	**31/03/2026**
Delivery date:	**31/03/2026**
Lead Beneficiary:	**Ingartek**
Main author:	Saioa Etxebarria

Co-funded by the European Union

Document fiche	
Authors:	Saioa Etxebarria [Ingartek]
Internal reviewers:	Isidro Arrieta [Ingartek]
Work Package:	WP3
Task:	T3.4
Nature:	R
Dissemination:	PU

Document History			
Version	**Date**	**Contributor(s)**	**Description**
V 0.0	02/02/2026	INGARTEK	Initial draft
V 1.0	27/03/2026	INGARTEK	Final draft to be reviewed
V 1.1	29/03/2026	UPM & UPC	Revision
V 2.0	31/03/2026	INGARTEK	Final version

Keywords:	Data quality, FAIR principles, data validation, data cleaning

Abstract (few lines):	This deliverable presents the data quality monitoring and assessment plan for the datasets integrated into the Mobility Data Space. It defines the methodologies, processes, and pipelines required to ensure that data complies with FAIR principles (Findable, Accessible, Interoperable, and Reusable), as well as additional quality dimensions such as validity, integrity, and spatial-temporal resolution.

Table of Contents

DEFINITIONS, ACRONYMS AND ABBREVIATIONS **5**

EXECUTIVE SUMMARY **6**

1 INTRODUCTION **7**

1.1 PURPOSE AND OBJECTIVES 7
1.2 INTENDED AUDIENCE 7
1.3 STRUCTURE OF THE DELIVERABLE 7

2 DATA QUALITY APPROACH **9**

2.1 TYPES OF DATA WITHIN THE PROJECT 9
2.2 ROLE OF METADATA IN DATA QUALITY 10

3 DATA QUALITY CONSIDERATIONS **12**

3.1 FAIR PRINCIPLES 12
3.1.1 Findability *13*
3.1.2 Accessibility *13*
3.1.3 Interoperability *14*
3.1.4 Reusability *14*
3.2 ADDITIONAL DATA QUALITY DIMENSIONS 14
3.2.1 Accuracy *15*
3.2.2 Validity *15*
3.2.3 Integrity *15*
3.2.4 Timeliness *16*
3.2.5 Completeness *16*
3.2.6 Consistency *16*
3.3 SPATIAL AND TEMPORAL ASPECTS 16

4 DATA LIFECYCLE **18**

4.1 DATA LIFECYCLE AND RESPONSIBILITIES 18

5 DATA QUALITY ASSESSMENT **20**

5.1 COMMON MITIGATION STRATEGIES 21
5.2 VERIFICATION OF DATA AVAILABILITY AND COMPLETENESS 22
5.3 VERIFICATION OF DATA USABILITY FOR MODELLING 23
5.4 VERIFICATION OF DATA CONSISTENCY AND COHERENCE 23
5.5 KPI-BASED DATA QUALITY VERIFICATION 24

6 CONCLUSIONS AND NEXT STEPS **25**

7 REFERENCES **26**

Table of figures

Figure 1. FAIR principles in Data Space context. 13
Figure 2. General Data Quality Rules. 15
Figure 3. General data lifecycle. 19
Figure 4. Data quality pipeline. 20

Definitions, Acronyms and Abbreviations

Acronym	Title
DS4MOBMODELS	Data Space for Mobility Models Integration
DPO	Data Product Offering
FAIR	Findable, Accessible, Interoperable, Reusable
KPI	Key Performance Indicator
TEN-T	Trans-European Transport Network
WP	Work Package
GA	Grant Agreement
D3.1	Deliverable 3.1 – Specifications and collection of enriched datasets
D3.2	Deliverable 3.2 – Interoperability and harmonisation framework
D3.3	Deliverable 3.3 – Data quality monitoring and assessment plan
D5.1	Deliverable 5.1 – Data discovery and governance framework
mobilityDCAT-AP	Mobility extension of the DCAT Application Profile for data catalogues

Executive Summary

The development of Deliverable D3.3 builds upon the outcomes of previous work carried out in WP3, in particular Deliverables D3.1 "Specifications and collection of enriched datasets" and D3.2 "Interoperability and harmonisation framework". While D3.1 focused on identifying and analysing available datasets and D3.2 established the standards and mechanisms for their interoperability and harmonisation, D3.3 advances this work by defining a practical framework for monitoring and assessing data quality within the DS4MOBMODELS Mobility Data Space.

This deliverable establishes a common approach to data quality that is adapted to the distributed and heterogeneous nature of the project. Rather than enforcing centralised control over data, the proposed framework relies on the coordinated contribution of data providers and project participants, supported by the use of metadata and catalogues. In this context, data quality is addressed as a context-dependent concept, closely linked to the intended use of data within specific use cases, particularly in modelling and analysis activities.

The document defines a set of verification mechanisms that can be applied throughout the data lifecycle to ensure that datasets are fit for purpose. These verifications focus on key aspects such as data availability and completeness, usability for modelling, consistency and coherence across datasets, and the ability to support the calculation of Key Performance Indicators (KPIs). In this way, data quality assessment is embedded within the operational workflows of the project, rather than treated as an isolated or purely technical process.

In addition, D3.3 builds on the FAIR principles introduced in previous deliverables, reinforcing their role as guiding elements for data management within the Mobility Data Space. Complementary data quality dimensions—such as accuracy, validity, integrity, timeliness, completeness, and consistency—are also considered to support reliable and meaningful use of data across heterogeneous sources. The role of metadata is highlighted as a key enabler for data discoverability, understanding, and appropriate use, particularly in a context where data remains under the control of its original providers.

1 Introduction

1.1 Purpose and objectives

The purpose of this deliverable is to describe the approach followed in the project to monitor and assess the quality of the datasets used within the Mobility Data Space. The project relies on datasets provided by different stakeholders and complemented, when necessary, with external data sources.

In this context, this deliverable defines a practical approach to support the understanding and appropriate use of data, with a particular focus on the role of metadata and on the treatment of datasets that are transformed or generated within the project to address specific use cases.

The main objectives of this deliverable are:

- **To describe how data quality is considered within the project**, taking into account the different data sources involved.
- **To outline the basic verifications that can be performed on the datasets**, in order to support their use within the project.
- **To ensure that datasets are described in line with FAIR principles**, in particular regarding findability, accessibility, interoperability and reusability.
- **To consider key data quality aspects**, such as validity, integrity, overall quality, and spatial and temporal resolution.
- **To describe how data transformations carried out in the use cases contribute to improving the usability of the data.**

1.2 Intended audience

This deliverable will be disseminated publicly, addressing all interested stakeholders who require an understanding of the data quality approach adopted within the project. In particular, it provides an overview of the considerations and verifications applied to assess data quality, as well as the role of metadata and data transformations in supporting the appropriate use 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 data quality approach adopted in the project**, including the different types of datasets involved and the role of metadata in supporting their discovery and use within the Mobility Data Space.

- **Section 3 introduces the main data quality aspects considered in the project**, including FAIR principles and other relevant dimensions such as validity, integrity, and spatial and temporal resolution.
- **Section 4 describes the data lifecycle,** the roles and responsibilities of the different participants, and how data is accessed and applied within the use cases.
- **Section 5 outlines the basic approaches used to assess data quality**, describing the verifications that can be performed on the datasets in the context of the project.
- **Section 6 presents a synthesis of the key conclusions of the deliverable**, highlighting the main elements of the data quality approach and their role in supporting the use of data within the project.

2 Data quality approach

The project relies on a variety of datasets provided by different stakeholders, complemented where necessary with external data sources. These datasets are used to support the development of the defined use cases, particularly in the context of modelling and analysis activities. Given this diversity of sources, data quality is approached in a practical manner, taking into account the characteristics of the available data and its intended use within the project.

In this context, data is not centrally managed or stored within the Mobility Data Space. Instead, datasets are made available through catalogue entries, supported by the publication of metadata that facilitates their discovery and understanding. Data is typically accessed from its original source, while only selected datasets—such as those resulting from transformations or processing carried out within the project—may be incorporated into the data space when they provide added value.

The data quality approach adopted in the project focuses on supporting the appropriate use of datasets, rather than enforcing strict control over their original content. This includes ensuring that datasets are adequately described through metadata, considering key quality aspects, and applying data transformations when required to meet the needs of specific use cases. The following subsections describe the main elements of this approach, including the types of data involved, the role of metadata, and the general considerations related to data quality within the project.

2.1 Types of data within the project

The project relies on a variety of heterogeneous data sources to support the implementation of the defined use cases. These data sources have been identified through the activities carried out in Deliverable D3.1, which collects information on the datasets made available by project stakeholders based on interviews and direct interactions. This initial work provides an overview of the available data landscape, including the types of data accessible, their origin, and their potential relevance for the different project activities.

Building on this analysis, the data used throughout the project can be grouped into several categories according to their origin and function:

A first category includes **datasets provided by project stakeholders**, such as public authorities, transport operators, and infrastructure managers. These datasets represent the backbone of the project and are typically registered in the catalogue with associated metadata describing their structure, update frequency, accessibility, and usage constraints. In most cases, the data remains at the source and is accessed when required, in line with data sovereignty principles promoted by the Mobility Data Space.

A second category comprises **external datasets**, which complement stakeholder data where gaps are identified. As described in D3.1, these datasets may include both open data and data obtained from third parties, including private providers, and are evaluated based on their relevance, quality, and compatibility with project requirements. While they are generally accessed directly from their original providers, they may be integrated into the project environment when necessary to support specific use cases or to ensure consistency with other datasets. In such cases, any processing or transformation applied to these datasets may lead to the generation of processed and harmonised datasets within the project context.

A third category refers to **processed and harmonised datasets**, which are generated through the integration and transformation activities carried out within the project. Given the heterogeneity identified in D3.1—particularly in terms of formats, standards, and data models—significant effort is devoted to harmonising datasets to ensure interoperability. This includes aligning data structures, standardising formats, and enriching datasets with additional attributes where required.

Finally, a fourth category includes **derived and model-generated datasets**, resulting from simulation, analytics, and modelling processes. These datasets build upon the harmonised data and are tailored to the specific requirements of the project's use cases. When relevant, they are also described and made discoverable through the catalogue, ensuring traceability and facilitating reuse.

Although different categories of data are identified, the project does not define separate data quality plans for each category. Instead, **a common data quality approach is applied across all datasets, as described in Section 5**. This approach is based on a shared set of verification strategies aligned with FAIR principles and data quality dimensions, which are applied in a flexible manner depending on the characteristics of each dataset and its role within the use cases.

2.2 Role of Metadata in Data Quality

Metadata plays a central role in supporting data quality within the project, particularly in a context where datasets originate from multiple sources and are not centrally managed. As identified in Deliverable D3.1, the available datasets are accompanied by varying levels of metadata, which provide essential information for understanding their content, origin, and potential use. Building on this, Deliverable D3.2 defines a common metadata model based on mobilityDCAT-AP, establishing a consistent way to describe datasets across the project.

From a data quality perspective, **metadata is a key enabler** for the correct interpretation and use of datasets. Rather than modifying the original data, the project relies on metadata to provide structured information that helps users

understand what the dataset contains, how it has been generated, and under which conditions it can be accessed and reused. This includes basic descriptive elements such as title and description, as well as more specific attributes such as publisher, update frequency, spatial coverage, and thematic classification, which are defined as mandatory elements within the adopted metadata model.

Metadata also plays a fundamental role in ensuring the **findability and accessibility** of datasets. By describing datasets in a structured and standardised way, metadata enables their inclusion in catalogues and supports their discovery by different users and systems. The use of a common metadata model based on mobilityDCAT-AP ensures that datasets can be consistently identified and accessed, independently of their original source. In particular, metadata elements such as identifiers, access URLs, and distribution formats provide the necessary information to locate and retrieve the data.

In addition, metadata contributes to **interoperability and reuse**, which are key aspects of data quality within the project. As defined in Deliverable D3.2, the adoption of standard vocabularies and ontology-based metadata models allows datasets from heterogeneous sources to be described in a harmonised and machine-readable way. This ensures that key concepts—such as transport modes, spatial coverage, or data standards—are consistently interpreted across different datasets and use cases, supporting their integration and combined use within modelling and analysis activities.

Finally, metadata provides essential context for assessing data quality. Elements such as update frequency, temporal and spatial coverage, data standards, and usage conditions help users evaluate whether a dataset is suitable for a specific purpose. In this sense, metadata supports a practical approach to data quality assessment, enabling informed decisions about data usage without requiring full control over the original datasets. This is particularly relevant for datasets accessed from external sources, where metadata becomes the main mechanism to understand their characteristics and limitations.

3 Data quality considerations

Data quality within the project is considered as a key element to support the effective use of datasets in the different use cases. Given the diversity of data sources involved, data quality is not approached through a single, uniform definition, but rather through a set of common considerations that help to understand and assess the suitability of datasets for their intended use.

In this context, the project adopts a pragmatic approach to data quality, focusing on aspects that are relevant for the discovery, access, and use of data. This includes the consideration of the FAIR principles—Findable, Accessible, Interoperable and Reusable—which provide a widely accepted framework to support the efficient management and reuse of data. These principles are particularly relevant in the project, where datasets are described through metadata and accessed from different sources.

In addition to FAIR principles, the project also considers other data quality dimensions that are especially relevant for mobility and transport data. These include aspects such as validity, integrity, and overall consistency of the data, as well as its spatial and temporal characteristics. These elements are important to ensure that datasets can be appropriately interpreted and used within modelling and analysis activities.

3.1 FAIR Principles

The **FAIR principles—Findable, Accessible, Interoperable and Reusable**—are widely recognised as a reference framework for data management and are particularly relevant in the context of data spaces. In these environments, data is not centrally stored but shared across multiple actors under conditions of trust, sovereignty, and controlled access. As such, **FAIR principles provide the foundation to enable data discovery, access, and reuse across a distributed ecosystem**.

In the project, FAIR principles are not enforced as strict requirements on all datasets, but are considered as guiding principles that support the effective use of data within the Mobility Data Space. Their implementation is mainly supported through the use of metadata, the data catalogue, and the transformations carried out within the use cases. In this sense, FAIR is closely linked to how data is described, accessed, and combined, rather than to the intrinsic quality of the raw datasets.

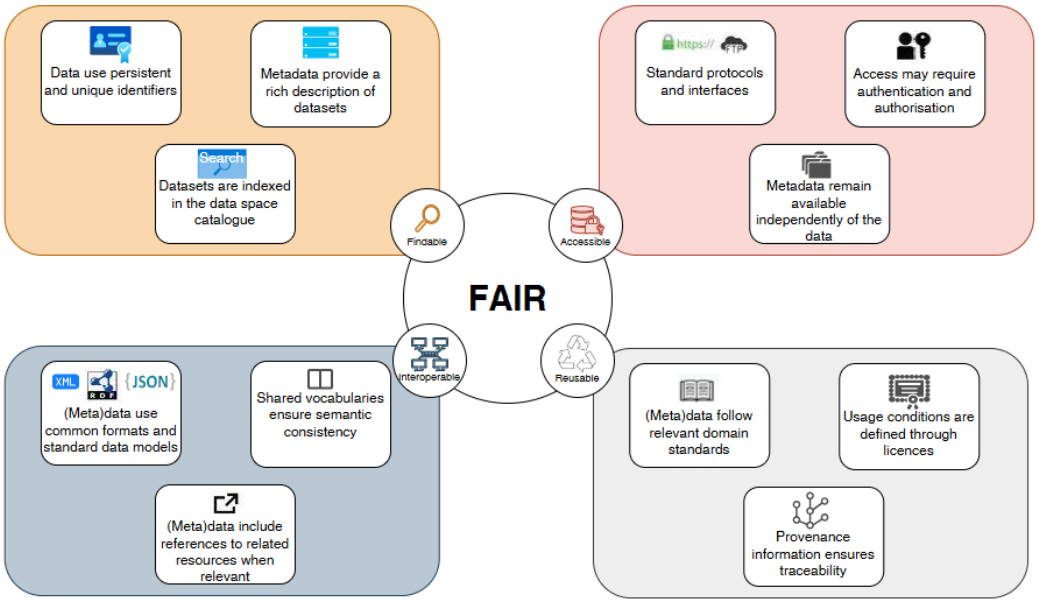


Figure 1. FAIR principles in Data Space context.

3.1.1 Findability

In a data space context, **findability is primarily achieved through the use of catalogues and metadata**. Data spaces are designed to enable participants to discover available datasets without requiring direct knowledge of their location or ownership. This is typically implemented through searchable catalogues where datasets are registered and described.

Within the project, findability is supported by the publication of metadata in the catalogue, allowing datasets from different stakeholders to be identified and explored. Metadata elements such as identifiers, descriptions, keywords, and thematic classifications enable both human users and systems to locate relevant datasets. This is aligned with FAIR principles, which emphasise that both data and metadata should be indexed and discoverable through searchable resources.

3.1.2 Accessibility

Accessibility in data spaces differs from traditional data management approaches, as data is often accessed directly from its original source rather than being centrally stored. **Data spaces promote data sovereignty, meaning that data providers retain control over how their data is accessed and used**.

In this context, accessibility is enabled through metadata, which provides the necessary information to retrieve the data, such as access URLs, formats, and usage conditions. The project does not modify access conditions but ensures that they are clearly described, allowing users to understand how data can be accessed, whether it is openly available or subject to restrictions.

3.1.3 Interoperability

Interoperability is a key challenge in data spaces, where data originates from multiple heterogeneous sources. **It refers to the ability to combine and use datasets together, even when they differ in format, structure, or semantics**.

Within the project, interoperability is supported through the adoption of a common metadata model and standard vocabularies, as defined in Deliverable D3.2. In addition, interoperability is addressed through data transformation and harmonisation activities carried out within the use cases when required. These activities enable datasets to be aligned and combined, supporting cross-domain and cross-border use cases.

3.1.4 Reusability

Reusability refers to the ability to use datasets in different contexts beyond their original purpose. In data spaces, this is closely linked to the availability of sufficient contextual information and clear usage conditions, rather than to the central storage of data.

In the project, reusability is supported by the use of metadata to describe datasets, including information about their **origin, characteristics, and conditions of use**. In addition, the generation of processed and derived datasets within the use cases contributes to enhancing reusability, as these datasets are adapted to specific analytical needs and may be made available for further use when relevant.

3.2 Additional Data Quality Dimensions

In addition to the FAIR principles, a number of complementary data quality dimensions are considered within the project to support the effective use of datasets in modelling and analysis activities. These dimensions reflect common data quality practices and are aligned with the governance approach defined for the DS4MOBMODELS data space.

As defined in Deliverable D5.1, data providers are responsible for ensuring that their data meets certain quality expectations. These aspects are not enforced through a single technical framework but are operationalised through the publication of **Data Product Offerings (DPOs), the use of metadata, and the governance processes of the data space**.

The following dimensions are considered in the project:



Figure 2. General Data Quality Rules.

3.2.1 Accuracy

Accuracy refers to the degree to which data correctly represents real-world conditions.

Within the project, accuracy is considered as a key aspect for ensuring reliable modelling results. However, given the heterogeneity of data sources, accuracy is not enforced through a single methodology but depends on the characteristics of each dataset and its provider. The role of the data space is to facilitate transparency regarding data origin, context, and limitations, allowing users to assess its suitability for their specific purposes.

3.2.2 Validity

Validity refers to the extent to which data values are consistent with their intended meaning and conform to expected formats, rules, or domain constraints.

Within the project, validity is mainly linked to the correct definition and interpretation of data, which is supported through metadata descriptions and, where applicable, shared vocabularies. This ensures that datasets can be correctly understood and used in the context of mobility modelling.

3.2.3 Integrity

Integrity relates to the internal consistency of datasets, including the coherence between different data elements and the absence of contradictions or corruption.

In the context of the data space, integrity is supported through governance processes that promote traceability, transparency, and proper data management practices. Data providers are expected to maintain the consistency of their datasets and ensure that any transformations applied to the data do not compromise its structure or meaning.

3.2.4 Timeliness

Timeliness refers to the degree to which data is up to date and available within a timeframe suitable for its intended use.

This aspect is particularly relevant in mobility scenarios, where some use cases require near real-time or frequently updated data. In the data space, timeliness is primarily the responsibility of data providers, who are expected to maintain and update their data according to its nature and intended use.

3.2.5 Completeness

Completeness refers to the extent to which all required data is available and no relevant information is missing.

Given the distributed nature of the data space, completeness may vary across datasets and use cases. Within the project, this dimension is addressed by progressively incorporating datasets from different stakeholders and by identifying potential data gaps during the implementation of the use cases.

3.2.6 Consistency

Consistency refers to the degree to which data is coherent and free of contradictions, both within a single dataset and across multiple datasets. It ensures that related data elements do not conflict with each other and that the same concepts are represented in a uniform way.

In the context of the data space, consistency is particularly relevant due to the integration of heterogeneous data sources. It is supported through the use of metadata, shared vocabularies, and governance processes that promote a common understanding of data. In addition, consistency is addressed in practice during data usage, especially when combining datasets from different providers, where alignment of formats, units, and definitions may be required.

3.3 Spatial and Temporal Aspects

Spatial and temporal characteristics are key dimensions of data quality in the context of mobility and transport applications. These aspects directly influence the suitability of datasets for modelling, simulation, and analysis, particularly in use cases that involve dynamic behaviours and cross-border scenarios.

From a spatial perspective, datasets used in the project may cover **different geographical scopes, ranging from local and urban contexts to regional, national, and cross-border levels**. The level of spatial detail, together with the consistency of spatial references across datasets, plays a key role in enabling their combined use. In practice, spatial aspects are addressed through the use of metadata describing geographical coverage and, when required, through data transformations that align spatial units or coordinate systems across datasets.

These mechanisms are applied in accordance with the interoperability framework defined in Deliverable D3.2, which promotes the use of standardised geospatial formats and reference systems. This includes formats such as GeoJSON, GML, or GeoPackage for data exchange and processing, as well as the explicit definition of Coordinate Reference Systems (CRS), such as WGS 84 (EPSG:4326) for interoperability and ETRS89-based systems for national-level accuracy.

In addition, spatial heterogeneity is a relevant consideration in the project, as datasets from different stakeholders may refer to different administrative boundaries, transport networks, or levels of aggregation. This is particularly relevant in cross-border use cases, where ensuring spatial consistency between datasets from different countries is necessary to support coherent modelling and analysis.

From a temporal perspective, **datasets may differ in terms of update frequency, temporal coverage, and level of granularity**. While some datasets provide static or infrequently updated information, others include dynamic or near real-time data. These differences can significantly affect their usability within the use cases, particularly in scenarios that require up-to-date information or time-dependent analysis.

Temporal aspects are addressed **through the use of metadata and harmonised time representations**, which provide key information such as timestamps, update frequency, and data validity periods. In line with the framework defined in Deliverable D3.2, the **ISO 8601 standard is adopted** to ensure consistent and unambiguous representation of temporal information across datasets, when required by the use cases.

4 Data Lifecycle

Understanding how data is generated, shared, and used within the project is essential to support the design and implementation of the Mobility Data Space. In this context, particular attention has been given to analysing the data **lifecycle across the different use cases**, in order to provide a global view of how data flows between actors, systems, and processes.

This global perspective is particularly **relevant for WP5 activities, such as Task 5.1 or Task 5.2, where the definition of the data space architecture requires an estimation of elements such as data providers, consumers, connectors, and interaction patterns.** By identifying how datasets are made available, accessed, and transformed within the project, it becomes possible to better understand the operational requirements of the data space and the roles of the different participants. Rather than defining a fixed or centralised workflow, the project adopts a distributed and use-case-driven approach, where data remains under the control of its providers and is accessed and processed when required. The following sections describe this data lifecycle and the associated responsibilities, providing a practical overview of how data is managed and used within the project.

4.1 Data Lifecycle and responsibilities

The management and use of data within the project follows a distributed approach in which different actors are involved at different stages of the data lifecycle. Rather than establishing a centralised data management process, the project relies on the coordinated contribution of data providers, project participants, and use case developers, each with specific roles and responsibilities.

Data providers, such as public authorities and other stakeholders, **are responsible for the original datasets**. This includes their creation, maintenance, and publication through their respective platforms or systems. Within the project, these datasets are referenced through the catalogue by means of metadata, but they typically remain at their source and are accessed when required. As such, the responsibility for the content and quality of the original data remains with the data providers.

Project participants contribute to the description and use of these datasets within the project context. In particular, they are responsible for defining and maintaining the associated metadata, ensuring that datasets can be discovered, understood, and accessed by other participants. In addition, they support the identification of relevant datasets for each use case and their appropriate use within the project activities.

In this context, **participants may also take the role of data consumers** within the Mobility Data Space. As data consumers, **they access and use Data Product**

Offerings (DPOs) for analysis, application development, or decision-making purposes. This typically involves querying the catalogue, retrieving metadata, identifying relevant datasets, and establishing the necessary data access agreements according to the defined policies before accessing and using the data within the corresponding use cases.

Data transformations and processing activities are carried out within the context of the different use cases. These activities are performed when required to **adapt the available datasets to specific modelling or analysis needs.** Participants involved in the use cases are responsible for these transformations, which may include data preparation, integration of multiple sources, or the generation of derived datasets.

This leads to **a simple data lifecycle within the project**. Datasets are first identified and described through metadata in the catalogue. They are then accessed from their original sources and used within the different use cases. When necessary, data is processed or transformed to support specific objectives. As a result of these activities, new datasets may be generated, which can be described and made discoverable through the catalogue when relevant.

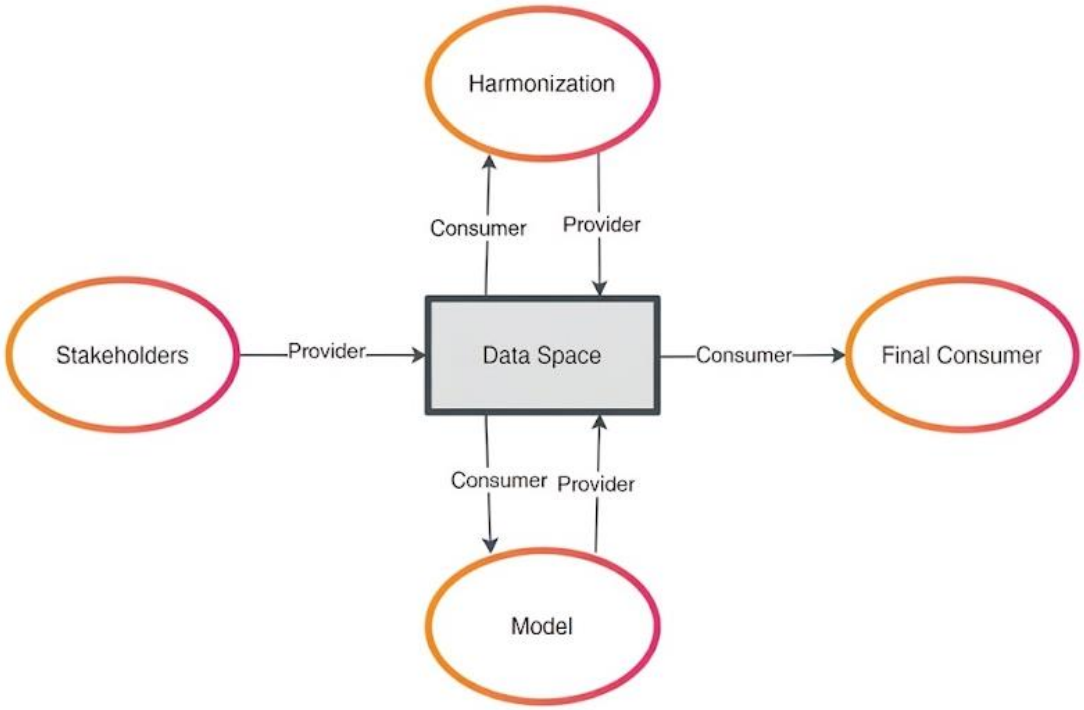


Figure 3. General data lifecycle.

5 Data quality assessment

Data quality assessment within the project is approached from a practical and use-case-driven perspective. Given the diversity of data sources involved and their different characteristics, **a common set of verification strategies is considered across the project, while allowing their application to be adapted to the specific context of each use case.**

In this context, data quality is not evaluated through isolated or purely technical checks, but through a set of **verifications that are integrated into the use of data within the different use cases.** These verifications aim to ensure that datasets are available, usable, and consistent enough to support modelling activities and generate meaningful results.

Two main perspectives guide this approach. On the one hand, **data must be sufficient and appropriate** to enable the correct functioning of the models implemented **in each use case**. On the other hand, **data must support the calculation of the Key Performance Indicators (KPIs)** defined in the project, which serve as a reference to evaluate the outcomes of the different scenarios.

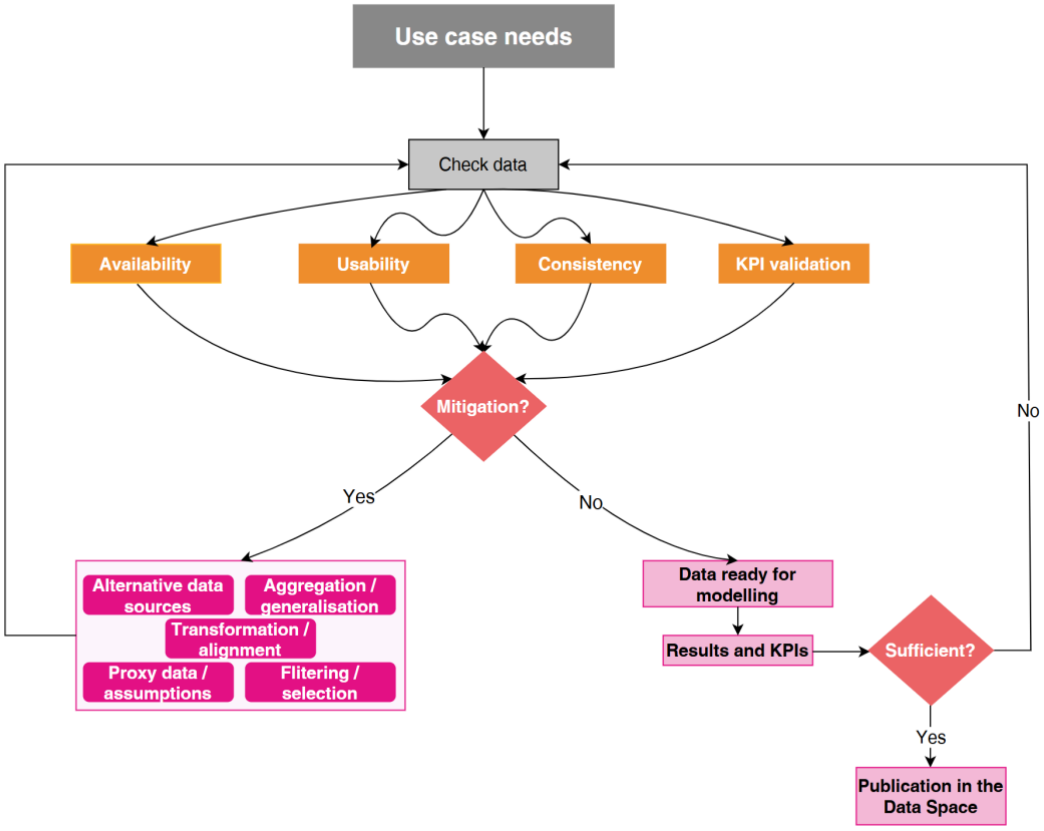


Figure 4. Data quality pipeline.

For example, in Use Case 5, which focuses on cross-border passenger transport between France and Spain, data quality assessment is directly linked to the

ability to analyse cross-border mobility and evaluate transport services and infrastructure. This use case requires the integration of multiple data sources, **including traffic counts provided by regional authorities such as the Gobierno Vasco and the Gobierno de Navarra, anonymised licence plate readings collected at border crossing points, and external datasets, including public transport timetables and infrastructure-related information**.

In this context, the proposed verification approach is applied as follows**: first, data availability and completeness are assessed by verifying that the required datasets, such as traffic counts, passenger flow information, and service frequency data, are accessible and cover the relevant spatial and temporal scope**. Then, data usability is evaluated by ensuring that **these datasets can be combined and used as inputs for the models, including the alignment of formats and aggregation levels**. Finally, **the calculation of KPIs—such as cross-border passenger volumes, service frequency, or traffic intensity—serves as a validation mechanism** to confirm that the data is suitable for the intended analysis.

This example illustrates how data quality assessment is embedded within the practical implementation of the use cases, linking data availability, modelling processes, and KPI calculation in a coherent workflow.

Based on these principles, the following sections outline a set of practical verification approaches that can be applied throughout the data lifecycle. These verifications provide a structured way to identify potential issues and ensure that data can be effectively used within the project context.

5.1 Common Mitigation Strategies

In the context of the project, data quality issues are expected to arise due to the heterogeneity of data sources, differences in data availability, and the specific requirements of each use case. Rather than preventing all possible issues in advance, the project adopts a pragmatic approach based on the identification of limitations and the application of mitigation strategies adapted to each situation.

These mitigation strategies are applied when data does not fully meet the requirements of a given use case and aim to ensure that modelling activities and KPI calculations can still be carried out. The selection of the appropriate strategy depends on the nature of the issue, the type of data involved, and the objectives of the use case.

Several common approaches can be identified across the project:

- **Use of alternative or complementary data sources**: when specific datasets are not available or are incomplete, additional sources may be incorporated to fill gaps or provide complementary information.

- **Data aggregation and generalisation**: in cases where detailed data is not available or presents inconsistencies, data may be aggregated to a higher spatial or temporal level to ensure coherence and usability.
- **Data transformation and alignment**: datasets from different sources may require adjustments in terms of format, structure, or spatial and temporal resolution to be used together within a use case.
- **Use of proxy data or assumptions**: when direct measurements are not available, proxy indicators or assumptions may be used to approximate the required variables.
- **Filtering and data selection**: in some cases, low-quality or inconsistent data may be excluded from the analysis to avoid affecting the results.

5.2 Verification of Data Availability and Completeness

The first step in assessing data quality within the project consists of **verifying that the required data is available and sufficiently complete to support the needs of each use case.** This initial verification is essential to ensure that modelling activities can be carried out and that the expected outputs, including the calculation of KPIs, can be achieved.

From a practical perspective, this involves **identifying the datasets** required for each use case and **checking whether they are accessible** from the corresponding sources. This includes verifying that the **necessary variables are available**, that the data **covers the required geographical scope**, and that **the temporal coverage is adequate** for the intended analysis. In cases where data is partially available, it is important to assess whether the available information is sufficient or whether additional data sources need to be considered.

Completeness is also assessed in terms of the presence of **missing values or gaps in the data**. Depending on the use case, certain levels of incompleteness may be acceptable, while in other cases they may prevent the correct functioning of the models or the calculation of KPIs. As such, completeness is evaluated in relation to the specific requirements of each use case.

When gaps or limitations are identified, different mitigation strategies can be applied. These may include **combining multiple data sources, using proxy data, or applying aggregation techniques** to compensate for missing information. The identification of such gaps is an important outcome of the data quality assessment process, as alows participants to take informed decisions on how to proceed with the available data.

5.3 Verification of Data Usability for Modelling

Once data availability and completeness have been assessed, the next step consists of verifying whether the available **datasets are suitable for use within the modelling processes** defined in each use case. This verification focuses on ensuring that data can be effectively used as input for the models and that it meets the minimum requirements needed for their correct execution.

This verification focuses on ensuring that the available data can be effectively used by the modelling tools and correctly interpreted within the context of each use case. This includes assessing whether the data structure is compatible with the model requirements, **whether units of measurement are consistent**, **and whether all necessary variables are present**. In addition, particular attention is given to the ability to combine datasets from different sources without introducing inconsistencies or conflicts.

Another key aspect is the alignment of data in terms of spatial and temporal resolution. Models often require data to be consistent in terms of geographical scope and time intervals. Therefore, it is necessary to verify whether the available datasets can be used together or **require adjustments, such as aggregation or resampling**, in order to be compatible.

A practical way to assess data usability is through the **execution of preliminary model runs or test scenarios**. These initial runs allow **identifying potential issues related to data input**, such as missing values, incompatible formats, or inconsistencies between datasets.

Finally, it is important to consider that data **usability is closely linked to the objectives of each use case.** Data that may not be suitable for one model could still be valid for another, depending on the level of detail and accuracy required. For this reason, this verification is carried out in a flexible manner, focusing on ensuring that the data is “fit for modelling” within each specific context.

5.4 Verification of Data Consistency and Coherence

In addition to ensuring data availability and usability for modelling, it is important to verify the internal consistency of datasets and the coherence between different data sources. These aspects are particularly relevant in use cases where multiple datasets are combined, as inconsistencies may directly affect the reliability of the results.

This verification includes checking that data values are logically consistent within each dataset. For example, it involves **identifying potential anomalies such as unrealistic values, contradictions between related variables, or duplicated**

records. These issues, if not addressed, may lead to incorrect interpretations or errors during model execution.

Coherence between datasets is also a key aspect, especially when integrating information from different sources. **This involves verifying that datasets referring to similar concepts are aligned, for instance in terms of definitions, units, or level of aggregation.** Inconsistencies between datasets—such as differences in traffic volumes, network definitions, or temporal references—may require further analysis or adjustment before being used together.

In practice, these verifications are often performed during data preparation and integration stages within the use cases. **Simple checks, comparisons between datasets, or exploratory analysis can help identify potential issues**. When inconsistencies are detected, mitigation strategies such as data filtering, aggregation, or prioritisation of certain data sources may be applied.

5.5 KPI-Based Data Quality Verification

A key approach to assessing data quality within the project is based on the ability to calculate the Key Performance Indicators (KPIs) defined in the Grant Agreement and further specified for each use case in Deliverable D2.3. **These KPIs represent the expected outputs of the use cases** and provide a practical reference to evaluate whether the available data is sufficient and appropriate for the intended analyses.

In this context, data quality is verified by checking whether all the required inputs for the calculation of the KPIs are available and usable. This includes ensuring that the necessary variables are present, that they cover the required spatial and temporal scope, and that they can be effectively integrated within the modelling processes. **If certain KPIs cannot be calculated due to missing or inadequate data, this indicates a limitation in data quality that needs to be addressed.**

In addition, the **plausibility and consistency of the calculated KPIs can also provide insights into data quality**. Unexpected or unrealistic results may reveal issues such as inconsistencies between datasets, incorrect assumptions, or limitations in the input data. In this sense, **KPI calculation acts as a validation mechanism** that reflects the combined effect of data availability, usability, and coherence.

6 Conclusions and next steps

This deliverable has presented a **practical approach to data quality** within the DS4MOBMODELS project, aligned with the characteristics of the **Mobility Data Space** and the requirements defined in the **Grant Agreement**. Rather than relying on a centralised or purely technical definition of data quality, the project adopts a **use-case-driven perspective**, where data quality is assessed in relation to its ability to **support modelling activities** and enable the **calculation of the defined KPIs**.

The analysis highlights the importance of **metadata, governance, and data lifecycle considerations** as key elements to support data quality in a distributed environment. In particular, the role of **data providers**, the use of **catalogues**, and the application of **data processing activities within the use cases** have been identified as fundamental aspects to ensure that data can be effectively **discovered, accessed, and used** across the project.

A set of **practical verification approaches** has been defined, covering aspects such as **data availability, usability for modelling, consistency, and KPI-based validation**. These approaches provide a **structured yet flexible way to assess data quality**, allowing participants to identify limitations, apply **mitigation strategies**, and ensure that datasets are **fit for their intended purpose** within each use case.

At the current stage of the project, **modelling activities are being further developed in collaboration with WP4**, with the aim of refining the requirements associated with each use case. As these models evolve, a more detailed understanding of **data needs and data flows across the different work packages** will emerge. In this context, the approach presented in this deliverable should be understood as a **general framework that will be progressively refined** based on practical implementation.

In parallel, efforts are being made to ensure that the proposed data quality approach remains aligned with the requirements of **WP5**, particularly in relation to the **design of the data space architecture and governance framework**. The definition of **data flows, roles, and interactions** described in this deliverable contributes to supporting these activities, facilitating the identification of **data providers, connectors, and integration mechanisms** required for the implementation of the Mobility Data Space.

7 References

FAIR Principles / Data Management

- European Commision
  FAIR Principles: Guiding principles to ensure data is Findable, Accessible, Interoperable, and Reusable.
  https://ec.europa.eu/isa2/actions/measuring-maturity-data_en/

- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016).
  The FAIR Guiding Principles for scientific data management and stewardship.
  Scientific Data, 3, 160018.
  https://doi.org/10.1038/sdata.2016.18
- CEUR Workshop Proceedings (2024).
  Data Quality in Data Spaces: Challenges and Approaches.
  https://ceur-ws.org/Vol-3705/paper01.pdf

Data Quality

- ISO (2019).
  ISO 8000-8: Data quality — Information and data quality: Concepts and measuring.
- ISO (2015).
  ISO 19157: Geographic information — Data quality.
- Batini, C., & Scannapieco, M. (2016).
  Data and Information Quality: Dimensions, Principles and Techniques.
  Springer.