



A Quality Framework for UVAR Data

Report by Task 3.2: European NAPs data quality

Status: Final draft
Version: 1.0
Date: January 29, 2025

Legal disclaimer

The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein. All images are provided by the respective partners (unless otherwise noted) and are approved for reproduction in this publication.

Document information

Project acronym	NAPCORE
Full project title	National Access Point Coordination Organisation for Europe
Grant Agreement No.	MOVE/B4/SUB/2020-123/SI2.852232
Activity no. and title	WG 3 NAP content and accessibility
Author(s)	João Montenegro, ARMIS (PT); Sónia Soares, ARMIS (PT); Lígia Conceição, ARMIS (PT)
Co-author(s)	Jasper Beernaerts, NGI (BE); Marlous Hovestad, NDW (NL); Peter Lubrich, Bast (DE)
Related to Milestone no.	M3.8
External Milestone	Yes

Document history

Version	Date	created/ modified by	Comments
0.1	07/02/2024	João Montenegro, Sónia Soares	First draft
0.2	17/04/2024	Jasper Beernaerts	Review
0.3	05/06/2024	João Montenegro, Sónia Soares	Draft
0.4	14/06/2024	João Montenegro	Draft
0.5	20/06/2024	Peter Lubrich	Review
0.6	12/07/2024	João Montenegro, Sónia Soares	Draft
1.0	25/10/2024	João Montenegro	Final draft

Action requested

- ☐ To be revised by partners involved in the preparation of the document
- ☐ For review/ approval by the Core Alignment Team
- ☐ For approval by the NAPCORE Steering Committee

Abstract

This deliverable is part of a set of quality frameworks being developed under the scope of NAPCORE WG 3. It comprises a Quality Framework dedicated to Urban Vehicle Access Restrictions (UVAR) data. The document contains a detailed description of quality criteria and metrics aligned with the characteristics and elements of UVARs. Additionally, it provides a practical example – that is – an evaluation of a real-world UVAR dataset according to the criteria proposed.

Abbreviations

Abbreviation	Meaning
CS	Congestion Charging Schemes
DG-MOVE	Directorate-General for Mobility and Transport
EC	European Commission
GIS	Geographic Information System
QGIS	Quantum Geographic Information System
ISO	International Organization for Standardization
LEZ	Low Emission Zone
LTZ	Limited Traffic Zone
MMTIS	Multimodal Travel Information Services
NAP	National Access Point



A Quality Framework for UVAR data

NAPCORE	National Access Point Coordination Organisation for Europe
PARK	Parking Regulations
PED	Pedestrian Zone
QF	Quality Framework
RTTI	Real-Time Traffic Information Services
TRO	Traffic Regulation Order
URL	Uniform Resource Locator
WG	Working Group
XML	eXtensible Markup Language
ZEZ	Zero Emission Zones



Table of contents

1. Introduction	6
1.1. Scoping.....	6
Most common UVAR types and their attributes	7
1.2. Methodology and structure	11
2. Background	13
2.1. Literature review on data quality dimensions	13
2.2. UVAR-related projects	14
3. Proposal for Quality Criteria	17
4. Proposal for Quality Assessment Methods.....	20
4.1. Generic approach	20
4.2. Aggregated view on quality criteria and metrics	20
5. Case study	22
5.1. Application of quality criteria and metrics.....	23
Accuracy-related criteria	23
Completeness-related criteria	24
Accessibility-related criteria	25
Currency-related criteria	25
Consistency-related criteria.....	25
Believability-related criteria	26
Relevance-related criteria	26
5.2. Aggregated view on quality criteria and metrics	26
6. Conclusions	28
References	29
List of Annexes	30

List of tables

Table 1 - Description of resulting quality dimensions.....	13
Table 2 - Description of quality criteria and metrics.....	18
Table 3 - Comparison between TRO and dataset regarding exemptions	24
Table 4 - Analysis of "required data coverage" - general information	24
Table 5 - Analysis of "required data coverage" – traffic regulation	24
Table 6 – Case study: Input values and respective weights for all quality metrics	27
Table 7 - Case study: resulting quality indicators	27

List of figures

Figure 1 – UVAR data value chain model (from UVAR Box [3])	7
Figure 2 – Entrance to a Low Emission Zone in Lisbon, Portugal (source: Sónia Soares/ARMIS).....	8
Figure 3 - Lisbon's Low Emission Zones scheme (source: Lisbon Municipality)	8



Figure 4 - Pedestrian Zone in Amsterdam, the Netherlands (source: Sónia Soares/ARMIS)	8
Figure 5 - Limited Traffic Zone in Rome, Italy (source: Sónia Soares/ARMIS).....	9
Figure 6 - Parking Regulation in Rome, Italy (source: Sónia Soares/ARMIS).....	10
Figure 7 - Geometry/location of UVAR dataset from the city of Lisbon	22
Figure 8 – Areas of the Polygons formed by the ground truth’s and dataset’s coordinates (in QGIS) ..	23



1. Introduction

This report summarises the elaboration by the NAPCORE Quality Activity for the field of Urban Vehicle Access Regulation (UVAR).

To facilitate such quality documentation, a dedicated Quality Activity in the NAPCORE project (WG 3/Task 3.2) is elaborating so-called Quality Frameworks (QFs) for various ITS data domains. These QFs provide, among others, commonly agreed definitions for quality criteria and guidelines on how to assess these criteria.

The recently updated Delegated Regulation on RTTI prescribes that any digitised Urban Vehicle Access Regulation (UVAR) data should be accessible via the NAP. Currently, there are different implementations of NAPs in different countries, with distinct distribution of roles and responsibilities in the provision and accessibility of RTTI and UVAR data. In the UVAR data value chain, UVAR authorities are responsible for collecting and generating UVAR DATEX II data and NAPs for its publication as a national single point of contact for service providers. The Delegated Regulation on RTTI encourages Member States to facilitate the registration of data, which can then be published on the NAPs.

For several reasons, establishing a robust data quality framework for UVAR data is paramount. UVARs play a critical role in shaping urban mobility and environmental sustainability, impacting both citizens and businesses. Ensuring the quality of UVAR data facilitates informed decision-making by policymakers, urban planners, and transportation authorities, enhancing public trust and confidence in UVAR regulations by providing transparent and credible information to stakeholders. Moreover, a comprehensive data quality framework promotes interoperability and data exchange between different systems and stakeholders, facilitating efficient implementation and enforcement of UVAR measures across jurisdictions. Ultimately, a well-established data quality framework not only enhances the effectiveness of UVAR policies but also contributes to the overall improvement of urban transportation systems and environmental quality.

Furthermore, gathering and retaining data alone doesn't generate value; the true value is realized only when collected data can contribute to tangible benefits, which rely on its quality. The goal of a UVAR value chain is to provide accurate and current UVAR information to users to increase awareness and compliance. The quality of the data and its management are of utmost importance for its credibility and acceptance by both service providers and end users. UVAR data quality management should be defined as a continuous process of monitoring and updating UVAR data characteristics based on agreed evaluation criteria for validation and classification. This process must be coordinated with local or national quality frameworks, as well as organizational, functional, and technical architectures.

This report brings together the results from the multiple stages involved in the development of a QF for UVAR data, offering a detailed illustration of its practical application. It covers each phase comprehensively, showcasing how the framework can be effectively implemented in real-world scenarios.

1.1. Scoping

More and more European cities are introducing UVARs of different types. UVARs refer to a set of measures and regulations implemented within urban areas to manage and control the access of vehicles based on various criteria, including emissions standards, vehicle types, and time-based restrictions. UVARs are designed to address environmental and traffic management objectives, primarily to reduce air pollution and congestion and improve overall urban sustainability and liveability.

A diverse set of types of access regulations is issued by cities, of which the most frequent examples are Low Emissions Zones (LEZ), Pedestrian zones (PED), Limited traffic Zones (LTZ), and Parking Regulations



(PARK). In addition, UVAR schemes are increasingly combining different types. Examples are LTZs that have emissions requirements and payment for permits, or LEZs that are implemented by a charging mechanism.

The UVAR data value chain, starting from city authorities as data sources, involves collecting, digitizing, validating, and publishing UVAR data to ensure accuracy and relevance for end users like drivers. This comprehensive model encompasses the entire process, from data collection to delivering credible and accessible information to drivers and road users. The aim is to increase awareness and compliance with UVAR regulations through continuous monitoring and updates aligned with established criteria and frameworks.

To illustrate the approach, Figure 1 exposes the ideal process, starting with the source of the information, which, in most cases, are city authorities.

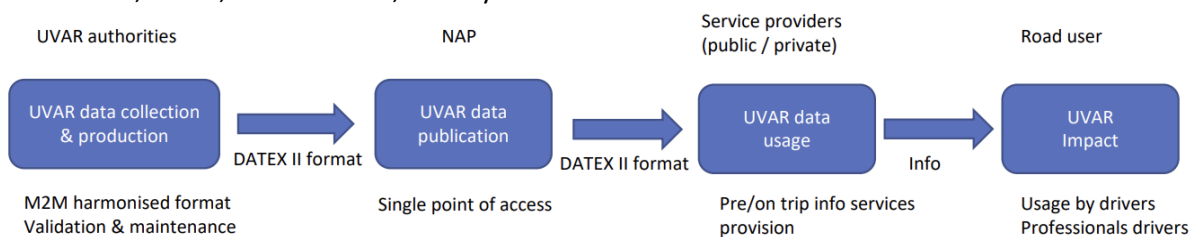


Figure 1 – UVAR data value chain model (from UVAR Box [3])

At this stage, the UVAR data is collected and must fulfil several requirements to ensure data quality. In that framework, information should be digitised and translated to a machine-to-machine (DATEX II standard) harmonised format to feed the NAP and other relevant publishing places. The responsible transport authority, operator, infrastructure manager, or transport-on-demand service provider must ensure the validation and maintenance of the information provided.

Below, a definition of the most common UVAR types is provided to enhance understanding of their main features and particularities, facilitating better assessment of data quality.

Most common UVAR types and their attributes

Low Emission Zone (LEZ): A LEZ is a designated urban area where vehicular access is restricted based on stringent air quality standards set by the UVAR scheme in place. The restrictions within LEZs typically focus on emissions levels, specifically targeting vehicles that do not meet minimum air quality standards. These standards are commonly based on European vehicle emission standards, which regulate exhaust emissions of pollutants such as NO_x, CO, VOC, HC, and PM organized into "Euro" stages or levels. LEZs may employ, in some cases, simplified national classification systems (stickers) using colours, numbers, or letters to categorize vehicles based on their compliance with these emission standards. Unlike LTZs, the primary aim of LEZs is to mitigate air pollution rather than solely addressing traffic volume or congestion. While traffic reductions can occur in the short term, the immediate objective is to accelerate the turnover of the vehicle fleet by discriminating against vehicles based on their contribution to air pollution, regardless of their mode of transport. In instances where only zero-emission vehicles are permitted, an LEZ becomes a Zero Emission Zone (ZEE). Additionally, some LEZs may have emergency measures that are activated only when air quality deteriorates to predefined emission limits, called pollution emergency schemes.



Figure 2 – Entrance to a Low Emission Zone in Lisbon, Portugal (source: Sónia Soares/ARMIS)



Figure 3 - Lisbon's Low Emission Zones scheme (source: Lisbon Municipality)

Pedestrian zones (PED): A PED is a designated urban space within a city or town, typically encompassing squares, streets, or interconnected road segments, where motorized vehicles are strictly prohibited. These areas are exclusively dedicated to pedestrian and, in some cases, bicycle use. Occasionally, permeable pedestrian zones may grant access to a limited set of vehicle categories, including emergency and police vehicles, individuals with mobility impairments holding a blue badge, residents requiring access to their garages, or delivery vehicles, often during short, off-peak time windows. Parking is prohibited within pedestrian zones, and any permitted vehicles must operate at a walking pace. The primary objectives of pedestrian zones are to enhance urban liveability, prioritize pedestrian safety, foster social interaction, and safeguard culturally and visually significant sites such as monuments and scenic landscapes.



Figure 4 - Pedestrian Zone in Amsterdam, the Netherlands (source: Sónia Soares/ARMIS)

Limited traffic Zones (LTZ): The core aim of LTZs is to restrict or regulate motorized vehicle access, aligning with objectives that prioritize urban liveability, road safety, and the preservation of cultural and natural heritage while mitigating congestion. In LTZs, the focus shifts toward reducing motorized traffic to essential and indispensable trips that serve the area's daily functioning. Typically, authorized and pre-registered user categories encompass residents, garage owners/tenants, caregivers, persons



with reduced mobility/blue badge holders, freight carriers, maintenance and servicing firms, alongside public transport, taxis, and emergency/police vehicles. These zones may also regulate parking and employ time windows to manage freight transport and loading/unloading activities. The overarching principle is to minimize motorized traffic, emphasizing walking, cycling, and public transport as preferred modes of movement. Additionally, LTZs may encompass specific vehicle categories (commonly including restrictions for lorries and coaches) or focus on vehicle characteristics such as type, weight, size, or pollution levels, addressing noise and air quality concerns.



Figure 5 - Limited Traffic Zone in Rome, Italy (source: Sónia Soares/ARMIS)

Parking Regulations (PARK): PARK aim to efficiently manage parking, control congestion, and optimize land use in cities. Parking within UVARs encompasses several facets, including permissions for parking, whether it's allowed or restricted, authorization for specific user categories such as disabled individuals or residents, designated time windows for parking, and whether it involves fees or is free. Moreover, UVAR-related permits, exemptions, and charges may encompass parking provisions and associated fees. In a broader context, UVARs oversee kerbside management, encompassing not only parking but also the regulation of loading/unloading, passengers' pick-up/drop-off, and street cleaning. Often, UVARs delineate time windows or restrictions to streamline these operations efficiently.



Figure 6 - Parking Regulation in Rome, Italy (source: Sónia Soares/ARMIS)

Understanding UVARs does not only mean understanding its types' implications but also the attributes that can characterise each regulatory system. UVARs, as diverse as they are, share core attributes that define their functionality and impact. The UVAR Box project, a pioneering initiative (explained later in section 2.2.), has contributed significantly by creating a comprehensive DATEX II model for describing UVARs, underlining the importance of a standardised framework for such regulations. Recognising these attributes is paramount as they form the basis for assessing the quality of UVAR data, an essential step in ensuring the effectiveness and reliability of UVAR systems. Consequently, a robust data quality framework requires a certain level of knowledge of these attributes, allowing for thorough evaluation and enhancement of UVAR data.

In alignment with the work carried out by the UVAR Box project in creating a comprehensive DATEX II model and a compatible structure complying with the Single Digital Gateway (SDG) Regulation, UVARs can be characterised according to the following main categories of attributes:

General information

- **City/Town Name and Country:** provides the geographical context of the UVAR, indicating the specific city or town where the regulation is enforced, along with the corresponding country. It serves as a fundamental reference point for understanding the UVAR's jurisdiction.
- **UVAR Name:** the UVAR Name is a distinctive identifier assigned to the UVAR. It helps stakeholders, including authorities and the public, easily recognise and refer to the specific UVAR scheme.
- **Broad Classification of the Type(s) of UVAR:** offers a high-level categorisation of the UVAR based on its primary objectives and restrictions. Common classifications include LEZ, PED, LTZ, Congestion Charging Schemes (CS), and PARK.
- **Status:** indicates whether the UVAR is currently active and enforceable or not. An "active" status signifies that the regulation is in effect and applicable to vehicles within the designated area, while a "not active" status suggests that the regulation is temporarily suspended or no longer in operation. This is especially important for countries, such as Italy, that contain seasonal UVARs.
- **Issuing Authority or Regulation ID:** specifies the entity responsible for implementing and overseeing the UVAR, often denoted as the "issuing authority." Additionally, it may include a unique Regulation ID, a reference code or a number associated with the UVAR, which aids in tracking and managing regulatory details.

Traffic Regulation

- **Type of Traffic Regulation:** defines the nature of the traffic regulation imposed within the UVAR. It distinguishes between different types of restrictions, such as access restrictions (limiting who can enter the zone), standing/loading restrictions (pertaining to stopping or unloading), or parking restrictions (concerning where vehicles can be parked).
- **Location:** specifies the geographical area or zone where the traffic regulation is enforced. It outlines the boundaries and spatial extent of the UVAR, often using Shape files.
- **Hours of Operation:** details the temporal aspects of the regulation, including its overall time validity. It may include start and end dates, as well as any complex structures of validity periods, or potentially triggering conditions that lead to activation/deactivation. In addition, an indication of whether the UVAR is dynamic or static is contained here.
- **Who is Affected/applicability:** who is subject to the specified traffic regulations. This typically involves a clear statement, such as "The traffic regulations apply to all motorized vehicles." If there are multiple conditions for applicability, they are presented in a structured list format.
- **Who is Exempted:** outlines all exceptions and exemptions from the traffic regulation. This section can be extensive and detailed, covering various criteria for exemptions. It may encompass exemptions based on e.g., vehicle emissions standards (e.g., Euro norms), purposes of travel, permits, delivery services, registered tradespersons, vehicle functions (e.g., postal service), or specific driver categories (e.g., residents or disabled individuals).
- **Penalty:** provides information on the penalties or fines associated with violating the traffic regulations stipulated within the UVAR. It details the consequences and repercussions for non-compliance, which can include fines, warnings, or other enforcement actions.
- **Tariffs:** details any fees or charges associated with the UVAR, such as tickets or permits that grant permission to enter the zone. It specifies when and where these permits can be purchased (e.g., before entering the zone), the methods of purchase, and any requirements for displaying physical permits, such as stickers or passes.

Additional Information

Additional information can be articulated, including alternative mobility options, URLs for further information on the UVAR, etc.

1.2. Methodology and structure

The first step was to carry out a comprehensive literature review to select the quality dimensions that would need to be considered while developing the QF. In that sense, information gathered on data quality was analysed through an extensive review of journal articles, conference papers, master's theses, book chapters, and standards. This comprehensive literature review identified 66 quality dimensions, which were organised into 9 major categories. Each dimension comprises a set of sub-dimensions that share commonalities but also exhibit distinct characteristics.

While the first phase was more theoretical and grounded in research, the next step involves defining quality criteria and use case metrics, bringing us closer to the practical formulation of a UVAR Data Quality Framework. This process began with formulating generic questions applicable across various data categories within the mobility and transport domain. These questions served as the baseline, offering insights for later producing the use case metrics specific to UVARs. Additionally, they were designed with the potential for applicability in other data quality frameworks.

To understand the particularities of UVAR data, we examined various studies and initiatives related to UVAR implementation, digitisation, maintenance, and data exchange, as detailed in the scoping and literature review sections. In this phase, the quality sub-dimensions and generic questions identified earlier are translated into a series of actionable questions, which can subsequently be processed to determine a data quality level.



Recognising that data quality is context-dependent, our approach emphasises flexibility, enabling the framework to be tailored to specific needs. Thus, each sub-dimension identified in the literature was assessed to evaluate its suitability for application to UVAR datasets. This assessment focused on tailoring the identified quality dimensions and sub-dimensions to the context of UVARs. Drawing on insights from previous UVAR initiatives like the UVAR Box and UVAR Exchange projects, this exercise involved creating UVAR-specific definitions for each dimension or sub-dimension when possible.

Next, for each dimension group, we formulated UVAR-specific criteria tailored to the characteristics of UVAR data. It is important to note that not all dimensions are included, as some were either not applicable to UVARs or considered not feasible to measure. For each criterion, one or more metrics were selected.

After consolidating a set of criteria and metrics, we developed a simple assessment method that allows us to obtain quality indicators for each quality dimension, as well as for the dataset as a whole. The method is a basic scaled weighted average that takes the evaluator's value as input and provides the quality indicators as output.

This report consolidates the outcomes of the several steps needed to achieve this QF and provides a concrete example of how to apply it. It was organised into 6 sections:

- 1. Introduction – including crucial UVAR-related definitions and explaining the process beneath the formulation of the quality assessment methods;
- 2. Background – exploring the main outcomes of the literature review on data quality dimensions, as well as gathering important information from other UVAR-related projects;
- 3. Quality criteria and metrics – demonstrating the results of the criteria and metrics developed, along with important clarifications;
- 4. Overview and explanation regarding the assessment method proposed;
- 5. Case study – a concrete example of how to apply the framework;
- 6. Discussion and outlook – including the main conclusions, lessons learnt, limitations and suggestions for future work.



2. Background

2.1. Literature review on data quality dimensions

In today's data-driven landscape, the importance of accessible, reliable, and accurate data cannot be neglected. High-quality data is fundamental to informed decision-making and efficient operations across various sectors. However, ensuring data quality is complex due to differing opinions on what constitutes data quality dimensions and the lack of systematic methodologies for developing quality frameworks. Despite extensive research over the decades, there is no consensus regarding the precise definition of data quality and its constituent dimensions.

To support a robust data quality framework, existing literature has been revisited, including scientific articles, book chapters, and standards that explore data quality dimensions. This comprehensive literature review identified 66 quality dimensions, which were then organized into a coherent structure contributing to developing effective quality frameworks. This effort resulted in a conference paper presented at the 6th International Conference on Industry 4.0 and Smart Manufacturing [4].

The outcome of this review is a consolidation of 37 quality dimensions into the following 9 major categories, which are described in Table 1.

Table 1 - Description of resulting quality dimensions

No.	Category/Group	Description/dimensions included
1	Openness	Comprises Accessibility (“how easily and quickly accessible data is”) and Availability (“extent to which information is available for download and in which formats”).
2	Exactness	Includes Accuracy (“how well the data values stored in a database align with real-world values”), Precision (“degree of detail and fitness of the measurement”), and Correctness (“the characteristic of data to be free of error or mistakes”).
3	Comprehensiveness	Deals with the wholeness and extensiveness of data. It includes Completeness (“the capacity of data to represent every meaningful state of the corresponding real-world environment”), the Amount of Data (“the quantity or volume of available data being appropriate and able to significantly represent reality”), Granularity (“degree to which the data elements can be divided”), Data Coverage (“the availability and comprehensiveness of data compared to the total universe of real-world states”), and the Variety of Data and Data Sources (“the variety of data and data sources that are available for the intended task”).
4	Coherence	Includes Consistency in data itself (“equivalence of information across data stores, applications, and systems”) and the Consistent Representation of data (“presentation of information in a uniform format that aligns with previous data, semantic rules and format”), Data Specification (“compliance with any existing legal or regulatory obligations, data standards, or business rules”), and Reliability (“the capacity of data to maintain a specified level of performance when used on specified condition”).
5	Veracity	Includes Believability, which is defined as the “extent to which information is regarded as true and credible by the user”, Traceability (“ability to track all access and changes to the data”), Objectivity (“extent to which data are unbiased unprejudiced and impartial”), and Reputation (“reputation, trustworthiness, or credibility of the source, provider or content”).
6	Promptness	Gathers dimensions focused on the frequency and speed of data exchanges to ensure it is up to date. It joins Timeliness (“time between the data becoming into existence and its availability for use”), Currency (“fit between the data's age or currency with its intended use”), and Volatility (“the period of validity in the real world, which the data aims at representing”).



7	Usability	Includes Understandability (“how clear, without ambiguity, and easily comprehended data is”), Conciseness (“how well information is compactly represented”), Interpretability (“standardizing language and representations, facilitating clear communication”), Ease of Manipulation (“assesses how data can be manipulated and applied consistently”), Ease of Operation (“evaluates the ease with which the data can be updated, maintained, and managed”), and Navigation (“assesses how easily data can be found and linked across databases”).
8	Safekeeping	Comprises Security (“extent to which access to information is restricted appropriately to maintain its security”), followed by Safety (“capability of the function to achieve acceptable levels of risk of harm”), and Integrity (“a measure of the fundamental/basic characteristics of the data”).
9	Relevance	Includes Relevance - described as “extent to which information is applicable and helpful for the desired use”-, Effectiveness (“ability to achieve specified goals with accuracy, completeness, and at a viable cost in a specified context of use”), Efficiency (“ability to quickly meet the information needs for a specific task”), Value added (“evaluates the extent which data provides advantages through its use”), Transactability (“degree to which data will produce the desired business transaction or outcome”), Usefulness (“fitness of data for a specific use”), and Uniqueness (“lack of duplicates records for the same event”).

It's crucial to recognise that not all dimensions must be applied universally when developing a Data Quality Framework. The concept of data quality as "fitness for use" reinforces the importance of context and the specific needs of data consumers. Different applications and domains may prioritise certain dimensions over others based on their specific requirements and operational environments. In this sense, the presented UVAR QF has concretised and tailored the above-mentioned categories in the context of UVAR data exchange. This means that some categories have been translated into one or several quality criteria, and a few have been neglected. However, the UVAR QF still has some variability, allowing for a flexible approach to data quality assessment. This means that this QF allows organisations to further tailor their definitions to their specific needs.

This approach provided guidelines for defining the UVAR Data Quality Framework, ensuring it accommodates the diverse requirements of UVAR data. Nevertheless, its flexibility extends beyond UVAR data, making it applicable to other quality frameworks and contexts where organizations across different sectors can leverage the same principles to develop data quality frameworks that address their specific challenges and goals.

2.2. UVAR-related projects

Coming from these general definitions of quality dimensions, in order to achieve UVAR-specific criteria, it is crucial to consult previous work regarding UVARs, i.e., previous recommendations on how to handle UVAR data. Therefore, before developing use-case-specific metrics, we delved into previous European projects that had UVARs as their core subject.

Providing relevant and effective information regarding UVAR schemes to all road users is of utmost importance to ensure rules are respected and publicly accepted. **UVAR Box** [5] is an EU-funded project (MOVE/B4/SER/2019-498/SI2.832125) providing tools to enable road users' information on access regulations to cities and regions in Europe, digitally, both ahead of and during their trip. The project started in August 2020 and ran for two years. The project began by developing the necessary structured data model, containing the necessary detailed information on UVAR schemes, the UVAR DATEX II data model. DATEX II is the standard used in Europe for the exchange of traffic information and traffic data and became the foundation for developing a software user-friendly tool, the UVAR Box Tool. Developing the UVAR structured data model, means to identify the minimum universal data



categories and information structure (including geographical scope, type of UVAR, access conditions, tariffs - prices and validity, payment options, exemptions, etc.). This model was the basis for the development of the friendly web-based tool, designed to be used by cities, or any other UVAR competent authorities, when generating digital UVAR schemes and making that information available.

Once available, the UVAR DATEX II data has the potential to become a valuable asset for policy makers, road authorities and service providers, but only if it is fully exploited. This depended on how easily accessible the data is, and how valuable the information is perceived by service providers, as well as on how this data can be incorporated into their information and navigation services. Thus, the UVAR Box project has interacted with key stakeholders to demonstrate and showcase how UVAR DATEX II data could be accessed via the network of European National Access Points (NAPs), promoting UVAR information usage on navigation services.

In parallel, the project aimed at facilitating the process for making the same information available, but in human-readable format, in accordance with the Single Digital Gateway (SDG) Regulation¹, ensuring the provision of pre-trip information.

Linked to the UVAR Box project, the **UVAR Exchange** European Parliament Preparatory action (MOVE/B3/SER/2020-716/SI2.852092) ran from August 2021 to January 2023. UVAR Exchange had two key objectives:

1. Firstly, to enhance the experience of road users by improving the communication of information to drivers in the vicinity of UVARs through improved road signs. UVAR Exchange proposed a set of harmonised physical signs to improve UVAR signage, which has been handed to the UNECE road signs working group (the international organisation responsible for road sign coordination) for their work to produce the final legally recommended UVAR road signs. Recommendations have also been made for Variable Message Signs (VMS), as well as by demonstrating the provision of information directly to a connected vehicle via C-ITS messages.
2. Secondly, UVAR Exchange addresses the technical, legal and administrative aspects of the cross-border sharing of vehicle and driver information. Enabling this would avoid the need for foreign drivers and vehicle operators to pre-register in different cities and countries. The only reason this is currently needed is when cities cannot get foreign vehicle technical information from other countries mainly due to the lack of a legal framework for UVARs. Enabling this cross-border sharing would make compliance checks and enforcement of foreign vehicles and drivers by the UVAR authorities easier, cheaper and quicker.

The work carried out under the UVAR Box and UVAR Exchange projects has been instrumental in laying the foundation for a robust and high-quality framework for UVAR data. These initiatives have played a pioneering role in digitising and harmonising UVAR data, marking a significant step forward in ensuring the effective management and dissemination of information regarding access regulations in cities and regions across Europe.

The UVAR Box Project's development of the UVAR DATEX II data model and the corresponding UVAR Box Tool has enabled Member States and cities to generate UVAR data in a standardized format. This contribution to harmonised quality requirements is pivotal in enhancing the consistency and reliability of UVAR information, making it more accessible and valuable for various stakeholders, including policymakers, road authorities, and service providers.

However, it's essential to recognize that digitizing and maintaining the quality of UVAR data also brings challenges. The process demands a considerable amount of knowledge, resources, and ongoing effort. To ensure the sustainable scaling of the UVAR Box Tool and to guarantee the long-term high quality of UVAR data, it is imperative that the tool undergoes continuous maintenance and development.

As UVAR Box and UVAR Exchange focused on harmonized machine-to-machine UVAR DATEX II data without defining a specific data quality framework, this study takes a crucial step in this direction, using



these projects as exemplars of best practices. The ongoing development of a data quality framework becomes increasingly important as UVAR schemes evolve and grow, offering road users a dependable source of precise and uniform information. This effort contributes significantly to the long-term success and widespread acceptance of UVAR schemes across Europe.



3. Proposal for Quality Criteria

The quality criteria for the UVAR QF have been defined following the process described in chapter 1.2. As highlighted during that chapter, not all quality dimensions defined by Carvalho et al. [4] were included. The authors couldn't find applicable and feasible metrics for "Usability" and "Safekeeping". Therefore, our proposed framework encompasses the remaining 7 categories, including 12 criteria and 15 metrics. Some additional explanations are given next for proper usage of the criteria and metrics.

For the dimensions included in the first group ("Openness"), we elaborated on two criteria considered adequate for the "Accessibility" dimension in the context of UVARs. One refers to a broader view of "how accessible" the dataset is, being translated into a multiple-choice metric which verifies the existence of access restrictions. The other criterion contains a single binary metric to verify if the dataset is made available through its respective NAP; since UVARs represent a datatype included in the RTTI DR, making it available at the NAP is crucial to ensure its discoverability by Service Providers.

Two quality dimensions are being assessed through a single quality metric: consistency (part of the "Coherence" group) and relevance (part of the "suitability" group). With regard to consistency, the main aspect is whether the dataset complies with the recommended data standard, which – in this case – is the DATEX II (UVAR profile). As for relevance, the only subdimension considered possible to measure (avoiding subjectivities) was data duplication, i.e., if there is significant duplication of data, then an important part of the dataset might not be relevant.

Concerning the dimensions included in the "Exactness" group, two main criteria were considered as part of the "Accuracy" dimension. One refers to location accuracy, containing two metrics encompassing the compatibility of distance and area between the coordinates of the evaluated dataset and the one established by the issuing authority (considered as ground truth). The other criterion assesses the alignment between the UVAR dataset and the ground truth (documentation) provided by the UVAR authority.

As for the "Comprehensiveness" category, each criterion assesses whether the dataset provides the required information to describe the UVAR, hence referring to the "Completeness" dimension. Three metrics are included, two of them being percentages referring to the coverage of general and traffic regulation information and another referring to the provision of relevant additional information (see section 1.1 for further information on those aspects).

When it comes to "Veracity", two criteria have been included as part of the "Believability" dimension, both only comprised of binary metrics. The first criterion is traceability, which assesses whether it is possible to track access and changes to the dataset. The second one is trustworthiness, which checks if the source of the information is reliable.

Finally, regarding the "Promptness category", since UVARs are either static or semi-dynamic data, usual criteria such as "latency" might not be of much relevance. Therefore, the authors adopted one criterion for each UVAR type, i.e., one applicable for a dynamic UVAR and another for a static one. Both refer to the frequency of updates of the dataset to adapt to real-world changes, therefore referring to the "currency" dimension.

The summary description of the criteria and respective metrics is provided in Table 2, which compiles the proposed quality dimensions, criteria and metrics, including their semantical definitions, mathematical formulations and value ranges.



Table 2 - Description of quality criteria and metrics

Quality dimension	Quality criterion	Criterion description	Metrics	Metric description	Mathematical/theoretical formulation	Possible values
Accuracy	Location accuracy	Assesses the accuracy of the dataset's coordinates in the horizontal plane (i.e., latitude and longitude) by comparing them with the ones established by the issuing authority.	Area Deviation (AD)	Defines how well the area defined in the UVAR file corresponds with the actual area defined by the issuing authority. Can be expressed through a percentage indicating the relative deviation between the area sizes defined by 1) The polygon formed by the coordinates in the UVAR dataset; 2) The polygon formed by the coordinates established by the issuing authority, considered as ground truth.	$AD = \frac{ A_{dataset} - A_{Authority} }{A_{Authority}}$	0 to 100%
			Hausdorff Distance (d_H)	The Hausdorff distance is a measure of similarity between two sets of points (in this case, coordinates). It is defined as the maximum distance of a set to the nearest point in the other set. In other words, it determines the extent to which each point of a set lies near some point of the other set, providing a metric for the greatest deviation between the sets.	$d_H := \max\{sup_{x \in X}(x, Y), sup_{y \in Y}(X, y)\}$ Where: - X: The set of coordinates from the UVAR dataset under evaluation. - Y: The set of coordinates from the map provided by the issuing authority, considered as the ground truth.	0 to positive infinity
	Validity of data	Assess if the information provided by the dataset is well aligned with the Traffic Regulation Order (TRO) or with the UVAR-related traffic signs.	Alignment with TRO (Multiclass evaluation)	Multiple choice question	Have any differences/misalignments between the information provided by the UVAR dataset (e.g., exemptions, hours of operation, UVAR type classification) and the Traffic Regulation Order (TRO) or UVAR traffic signs been detected?	0: Yes, there is significant misalignment - some elements clearly contrast with information from the TRO/traffic signs; 1: Slight differences identified between TRO/traffic signs and the dataset (not enough information to check if the dataset's information is correct) 2: No difference identified - dataset seems to match with either TRO or traffic signs
Completeness	Required data coverage	This criterion focuses on assessing the presence of essential elements, including administrative details, UVAR typology and traffic regulation details.	Percentage of general information elements covered (% general)	Measures the percentage of percentage of the data elements concerning general information that are covered by the dataset, namely: - City/Town Name and Country - UVAR name - UVAR type - Status (active or not active) - Issuing Authority or Regulation ID	$\% general = \frac{\text{count of elements covered}}{5 \text{ (total of elements assessed)}} \times 100\%$	0 to 100%
			Percentage of traffic regulation elements covered (% traffic)	Measures the percentage of the data elements concerning traffic regulation that are covered by the dataset, namely: - Type of traffic regulations (access restriction, standing/loading restrictions or parking restrictions) - Location - Hours of operation - Who is affected/applicability - Who is exempted/exceptions (if applicable) - Penalty (if applicable) - Tariffs (if applicable)	$\% traffic = \left(\frac{\sum_{i=1}^n \text{coverage score}_i}{n} \right) \times 100\%$ Where: coverage score_i is the coverage score for element i, with the following values: 0, if the element is not covered 0.5, if the element is partially covered 1, if the element is fully covered n is the total count of applicable elements.	0 to 100%
	Provision of additional information	Assesses whether additional information can be articulated (e.g., alternative	Binary classification	Yes/No question	Does the dataset include additional information (e.g., alternative mobility options, URLs for further information on the UVAR, etc.)?	0: No; 1: Yes

Quality dimension	Quality criterion	Criterion description	Metrics	Metric description	Mathematical/theoretical formulation	Possible values
		mobility options, URLs for further information on the UVAR, etc).				
Accessibility	Level of Accessibility	Assesses if the dataset is publicly accessible, and under which conditions	Multi-class evaluation	Multiple choice question	'Is the dataset publicly accessible? Are there access restrictions?	0: Dataset is not accessible 1: Dataset is accessible but content is not publicly accessible (meaning there are access restrictions) 2: Dataset is publicly accessible (no access restrictions)
	NAP accessibility	Assesses if the dataset is accessible through the National Access Point	Binary classification	Yes/No question	Is the dataset accessible through the National Access Point (NAP)?	0: No; 1: Yes
Currency	Update frequency (static UVARs)	Assess the frequency on which the (static) UVAR dataset is monitored/maintained/updated	Multi-class evaluation	Multiple choice question	How frequently is the UVAR dataset updated to reflect real-world changes?	N/A: dynamic UVAR 0: Less than once a year or no monitoring procedures established; 1: Every year 2: Every semester
	Timeliness (dynamic UVARs)	Assess the frequency on which the dynamic UVAR dataset is updated.	Multi-class evaluation	Multiple choice question	How quickly does the UVAR data update to reflect changes in access regulations, especially in dynamic scenarios like pollution emergency schemes?	N/A: Static UVAR 0: Slowly (within a day or more); 1: Moderately quickly (within hours); 2: Very quickly (within minutes);
Consistency	Compliance with data standards	Assesses whether the UVAR dataset is compliant with the recommended data standard: DATEX II	Binary classification	Yes/No question	Does the dataset comply with the DATEX II standard (UVAR profile)?	0: No; 1: Yes
Believability	Traceability	Traceability follows the lifecycle of data to track all access and changes to the data	Tracking system (Binary classification)	Yes/No question	Does the UVAR dataset have a system in place to trace all access and changes made to the information?	0: No; 1: Yes
			Retrieval of historical versions (binary classification)	Yes/No question	Is it possible to retrieve historical versions of the dataset to analyse past changes?	0: No; 1: Yes
	Trustworthiness	Assesses whether the UVAR information has been validated through a reliable source	Validation source (binary classification)	Yes/No question	Is the UVAR information sourced from the UVAR issuing authority?	0: No; 1: Yes
Relevance	Data duplication	Verifies if the dataset contains unwanted duplicate records captured for the same entity or event	Binary classification	Yes/no question	Are there duplicated information detected within the dataset	0: Yes; 1: No

4. Proposal for Quality Assessment Methods

4.1. Generic approach

In a real-life Quality Assessment for UVAR data, we assume that a dedicated person or organisation (maybe a quality manager at the data-issuing organisation) is conducting a pre-defined procedure to calculate and compile the quality criteria and metrics, as introduced in Section 3, and to document the calculation results.

Such assessment may be done by assignment (i.e., in a limited manner) or on a regular basis (e.g., every 6 months). The assessment object is supposed to be a specific UVAR dataset available on a NAP. The dataset consists of one or several data records for a specific spatial unit, maybe a city or even an entire country. The records may contain static and/or dynamic data.

It is noted that the calculation of some quality criteria requires a reference basis, also called ground truth. Such reference basis may be an official UVAR declaration (e.g. on a city's website) or Traffic Regulations Orders (TROs).

Depending on the above-mentioned prerequisites, the choice and concrete usage of the quality criteria and metrics might be different in each quality assessment. For reasons of understandability and transparency, we propose that each documentation of a Quality Assessment includes the following meta information:

- Author of Quality Assessment
- Date of Quality Assessment
- What UVAR dataset has been assessed? (Which currency? Which spatial coverage?)
- What reference basis has been used, if any?
- What quality criteria and metrics from the UVAR QF have been applied?
- Any remarks or additional notes about the calculation of the quality criteria and metrics?

The documentation of the Quality Assessment, e.g., tables with calculation results, including the above meta-information, should be available to potential data consumers. A preferred way to do so is to publish the documentation as metadata on the NAP. The NAPCORE Metadata specification¹ [6] has introduced a dedicated metadata element "Quality Information" to facilitate such documentation.

4.2. Aggregated view on quality criteria and metrics

A special task during the Quality Assessment might be an aggregation of the calculation results for individual criteria and metrics. The benefit of such aggregation is a better view on the entire quality of a UVAR dataset, and better comparability when assessing multiple UVAR datasets.

As a first step, we propose an approach that allows us to obtain aggregated quality indicators for each quality dimension. Given the input values for each metric – which are assigned by the evaluator – we use the equation (1) to calculate an aggregated quality indicator on the level of one quality dimension. It is a scaled weighted mean of the relative contributions of each individual metric to the quality indicator (these contributions are indicated by the fraction on the right side of the numerator).

$$Q_i = \frac{\sum_j \left(W_{i,j} \times A_{i,j} \times \frac{Input\ value_{i,j} - Worst_{i,j}}{Best_{i,j} - Worst_{i,j}} \right)}{\sum_j (W_{i,j} \times A_{i,j})} \quad (1)$$

Where:

¹ <https://napcore.eu/metadata/>

- Q_i – Quality indicator for the dimension i
- " i " – quality dimension identification
- " j " – the identification of the metric in each quality dimension
- $W_{i,j}$ – parameter weight for the " j "-th metric across the " i "-th dimension
- $A_{i,j}$ – applicability for the assessed dataset – binary value ... "0" for non-relevant (N/A), "1" for relevant
- $Worst_{i,j}$ – worst possible value that considered for the computation of the respective metric
- $Best_{i,j}$ – best possible value for the computation of the respective metric
- $Input\ value_{i,j}$ – the value assigned by the evaluator for the respective metric. It is worth mentioning that if the assigned value is outside of the interval $\langle Worst_{i,j}; Best_{i,j} \rangle$ (which may happen for quantitative metrics), the nearest value from that interval is used instead

It is important to note that the worst and best values considered for the computation might not be the same as the possible values outlined in Table 2 (last column). For some quantitative variables, the worst or best value would tend to infinity, which would make it impossible to perform the calculation of the quality indicator. In the UVAR QF, this only happens with the Hausdorff Distance (metric that belongs to the accuracy dimension), whose worst case is the positive infinity. For the case of that metric, we selected a reasonable value and considered it as the worst scenario, assuming that any worse values wouldn't result in significant changes in the quality of the dataset.

The assessment can be performed through the Excel sheet provided in Annex 1. The evaluator only needs to fill in the input values for each metric (column D, in yellow) and respective weights (columns D and E, in yellow), and the framework will compute the quality indicators. While working on the case study (section 5.2), we came up with values for these weights according to the expert judgment of the authors of this QF. However, although the report's case study can be used as a reference, evaluators are free to edit/adjust the weights of each metric according to their preference. Once again, we reinforce the importance of allowing for greater flexibility during the assessment process.

As a second step, after obtaining a quality indicator for each quality dimension, equation (2) details the process to obtain the general quality indicator for the dataset.

$$Q_{UVAR} = \frac{\sum_i (W_i \times Q_i)}{\sum_i (W_i)} \quad (2)$$

Where:

- Q_{UVAR} – overall quality indicator of the UVAR dataset
- " i " – quality dimension identification
- W_i – parameter weight for the " i "-th dimension
- Q_i – quality indicator of the " i "-th dimension

The weight of each dimension is calculated automatically by the Excel sheet as the average of the weights of each applicable metric belonging to that dimension. To better understand the application of the proposed criteria and assessment method, a use case example is described in the Section 5.

5. Case study

With the aim of verifying the applicability of the proposed quality framework, including its criteria and assessment methods, a UVAR dataset from the city of Lisbon (Portugal) was used as a case study. That dataset refers to a Low Emission Zone (LEZ), and it is located in Lisbon City Centre (see Figure 7).



Figure 7 - Geometry/location of UVAR dataset from the city of Lisbon

Under the UVAR Box initiative, this UVAR was digitized into a DATEX II format using the UVAR Box Tool. The dataset was created in June 2022 by ARMIS in cooperation with the Municipality of Lisbon, which provided and validated the information manually entered into the tool.

Lisbon has two LEZs. This case study focuses on Zone 1, which permits entry only to vehicles with a EURO III emission level or higher. Information about both zones and access to the relevant legislation is available on the dedicated website² [7].

Due to the familiarity with and availability of the dataset and respective geometry files (provided by the issuing authority), this use case was considered ideal for reference within this report.

² <https://informacoeseservicos.lisboa.pt/servicos/detalhe/zona-de-emissoes-reduzidas-zer>



5.1. Application of quality criteria and metrics

Accuracy-related criteria

Regarding the criterion location accuracy, two metrics were assessed: Hausdorff distance and Area Deviation. To obtain these metrics, UVAR's geographical information (coordinates) provided as a shapefile (SHP) by the municipality of Lisbon was considered the ground truth and compared to the geographical information from the evaluated DATEX II dataset (XML) using GIS tools (QGIS). The following steps were performed:

1. Extract the Coordinates (EPSG 3857) from the UVAR XML and save them as a txt file
2. Load the coordinates in QGIS as point layer "points_extracted_from_UVAR.shp"
3. Use the toolbox feature "Points to path" to generate a polyline "Path_through_UVAR_points.shp"
4. Use the toolbox feature "Lines to polygons" to generate the polygon "polygon_extracted_from_UVAR.shp"
5. Load the ground truth shapefile
6. Extract the applicable area from the shapefile and export as a new layer "ZER_ground_truth_Lisbon_municipality_applicable_zone.shp"
7. Calculate area of both "polygon_extracted_from_UVAR.shp" and "ZER_ground_truth_Lisbon_municipality_applicable_zone.shp" (see fig. below)
8. Manually compute the Hausdorff distance

This process resulted in the polygons depicted in Figure 8.

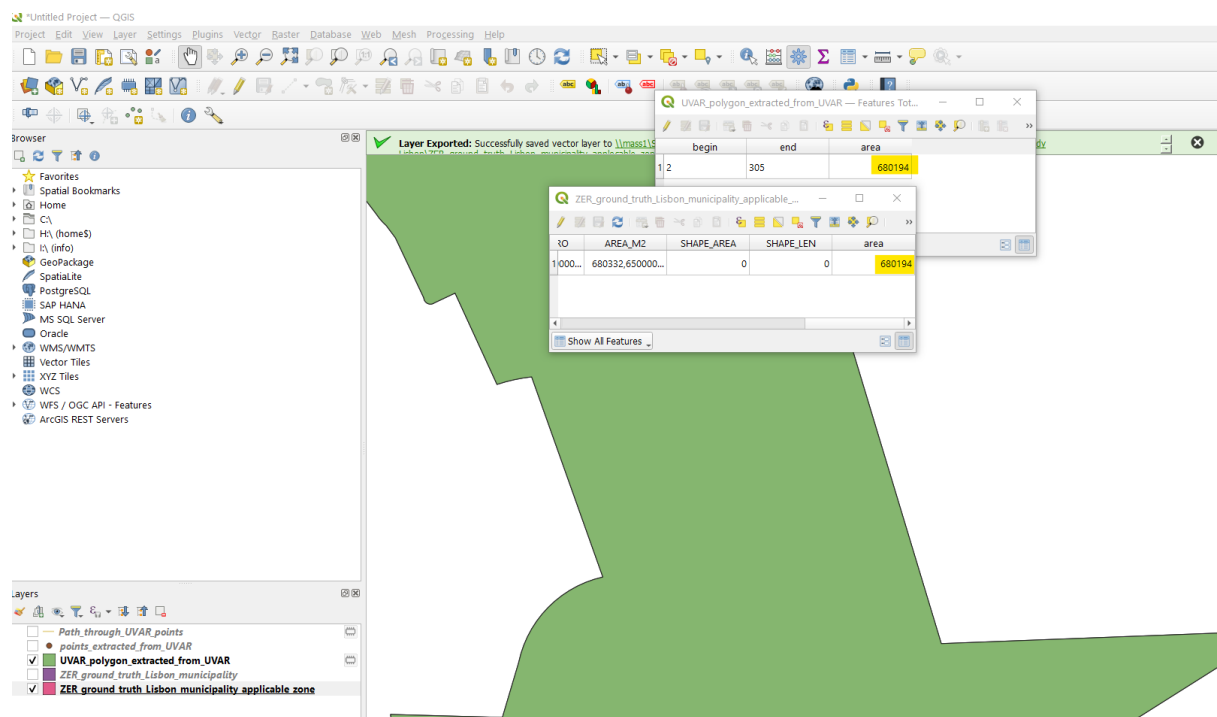


Figure 8 – Areas of the Polygons formed by the ground truth's and dataset's coordinates (in QGIS)

The coordinates of both polygons are identical, indicating a perfect match. Since their coordinates are the same, the calculation of areas resulted in the same value: 680.194 m². That said, the values assigned to the metrics are:

- **Area Deviation (AD): 0%**
- **Hausdorff distance: 0**



As for correctness, we did not identify significant misalignments between the dataset and the TRO. However, some slight differences have been observed, especially concerning the exemptions. Table 3 compares the exemptions included in the dataset and the official website.

Table 3 - Comparison between TRO and dataset regarding exemptions

TRO - exemptions	UVAR dataset - exemptions
- Emergency Vehicles - Vehicles Of Disabled - Historic Vehicles - Vehicles Of Residents - Natural Gas Vehicles - Lpg - Motorcycles - Police Vehicles - Military Vehicles - Prisoner Transport Vehicles - Armored Vehicles For Transporting Valuables	- Transport of disabled people - Road maintenance or construction - City logistics - Permanent resident - Vehicle with catalytic Converter - Historical vehicle - Emergency services - Liquid gas vehicles - Lpg vehicles - Motorcycles

Although some exemptions match, others (e.g., Prisoner Transport Vehicles) don't, which attests to the above-mentioned differences. Therefore, we assigned value **1** for the binary metric of the "validity of data" criterion ("Slight differences identified between TRO/traffic signs and dataset").

Completeness-related criteria

The completeness dimension comprises two criteria: "required data coverage" and "provision of additional information". The "required data coverage" criterion is assessed via two metrics: "percentage of general information elements covered" (analysis is presented in Table 4); and the percentage of traffic regulation elements covered (analysis presented in Table 5).

Table 4 - Analysis of "required data coverage" - general information

Element	Coverage	Information
City name/country	Yes	Lisbon/Portugal
UVAR name	Yes	UVAR_20220607_LEZ_PT_LISBON1
UVAR type	Yes	Low emission zone
UVAR Status	Yes	Active
Issuing Authority	Yes	Lisbon Municipality

Table 5 - Analysis of "required data coverage" – traffic regulation

Element	Applicability	Coverage score	Information/Observation
Type of regulation	Applicable	1	Access restriction
Location	Applicable	1	Defined as a polygon with assigned coordinates
Hours of operation	Applicable	1	UVAR is active during weekdays (Monday-Friday) from 7:00 to 21:00
Who is affected	Applicable	1	The dataset informs which vehicles are allowed to access the area: euro III emission level classification. All vehicles below that classification level are affected.
Exemptions	Applicable	0.5	The dataset contains some information regarding exemptions. However, it is incomplete: the TRO establishes a set of



Element	Applicability	Coverage score	Information/Observation
			exemptions, some of which are not covered by the dataset (partial coverage).
Tariffs	Not applicable	N/A	There are no tariffs associated with this UVAR
Penalty	Applicable	0	Although the UVAR carries a penalty as mentioned in [8], the dataset does not include information about these penalties. Therefore, the applicable penalty information is not covered.

Considering the outcome of the analyses presented in Tables Table 4 and Table 5, the application of the mathematical formulation of both metrics (see Table 2) is presented below, along with the resulting values.

$$\% general = \frac{\text{count of elements covered}}{5 (\text{total of elements assessed})} \times 100\% = \frac{5}{5} \times 100\% = 100\%$$

$$\% traffic = \left(\frac{\sum_{i=1}^n \text{coverage score}_i}{n} \right) \times 100\% = \frac{1 + 1 + 1 + 0.5 + 0}{6} \times 100\% = 75\%$$

As for the criterion “Provision of additional information”, the value of the binary metric is **1** (“Yes”), since a URL with additional information is provided through the following line of code: “<cz:urlForFurtherInformation> <https://informacoeseservicos.lisboa.pt/servicos/detalhe/zona-de-emissoes-reduzidas-zer></cz:urlForFurtherInformation>”.

Accessibility-related criteria

The assessed UVAR dataset is only accessible via the UVAR Box Tool. Therefore, it can only be downloaded by someone with an active account in the UVAR Box tool and necessary download permissions. For that reason, value assigned to the binary metric belonging to the “Level of Accessibility” criterion was **0** (“No”). As for the criterion NAP accessibility, the value **0** (“No”) was assigned to the binary metric since the dataset is not published on the Portuguese NAP.

Currency-related criteria

Since this is considered a static UVAR, there is only one applicable criterion within the timeliness dimension: “Update Frequency”. Given that the dataset hasn’t been updated since its creation, which happened in 2022 (as abovementioned), the value assigned to this metric was **0** (“Less than once a year or no monitoring procedures established”), referring to the worst-case scenario.

Consistency-related criteria

There is a single-metric criterion within the consistency dimension: compliance with recommended data standards. The recommended data standard, as detailed in Table 2, is DATEX II (UVAR profile). To carry out the assessment, first, the UVAR profile for a Low Emission Zone was generated using the DATEX II webtool³ [9]. The following settings were edited to generate the XSD files for validation:

- Source: V3.4 DATEX II TrafficRegulation
- Profile: UVARBox profiles – Low emission zones

³ <https://webtool.datex2.eu/wizard/>



- Profile location: AreaLocation
- Selection: PayloadPublication – Required attributes and ControlledZoneTablePublication

An XML validator was then used to validate the case study's dataset (XML) with the generated XSD files. It is important to state that different XML validators are available (e.g., Altova XMLspy, Oxygen XML Editor), but most of them require licence fees. After running the validation, the following message was displayed: "File C:\Users\joao.montenegro\Armis\Armis ITS Consulting - General\NAPCORE\WG3\Task 3.2 - Quality Frameworks\QF - UVAR\UVAR_20220607_LEZ_PT_LISBON1.xml is valid.". This indicates that the dataset complies with the recommended profile. Hence, value **1** ("Yes") was assigned to that metric.

Believability-related criteria

Regarding the criterion "Trustworthiness", since the dataset was produced within the scope of the UVAR Box project in partnership with the Municipality of Lisbon (Issuing Authority), it has been cross-checked and validated along with the issuing authority. Therefore, the binary value **1** ("Yes") was assigned. Still, it is worth mentioning that there is no formal validation as the dataset has not been published in a NAP.

As for the traceability criterion, the value of both binary metrics was set as **0** ("No") since it is not possible to retrieve historical versions and there is no tracking system in place. Whenever changes are made to the dataset, the previous version is replaced by the updated one, and there is no manner of consulting the earlier versions nor verifying who made the changes or which changes were made.

Relevance-related criteria

There is a single-metric criterion within the relevance dimension: data duplication. Since no data duplication has been identified, the value assigned for that metric was **1** ("No").

5.2. Aggregated view on quality criteria and metrics

The analysis of the quality criteria and metrics demonstrated in Chapter **Fehler! Verweisquelle konnte nicht gefunden werden.** provided us with the necessary input values that can be inserted into the Computation Spreadsheet (Annex 1). Table 6, extracted from the filled spreadsheet, demonstrates the values assigned to each metric. Additionally, it shows the weights assigned to each metric.

Table 6 – Case study: Input values and respective weights for all quality metrics

Quality dimension	Quality criterion	Quality metric	Weight	Input Value
Accuracy	Location accuracy	Area deviation	1	0%
		Hausdorff Distance	1	0
	Validity of data	TRO alignment	0.8	1
Completeness	Required data coverage	% general	0.7	100%
		% traffic	1	75%
	Provision of additional information	Binary classification	0.2	1
Accessibility	Level of Accessibility	Binary classification	0.6	0
	NAP Accessibility	Binary classification	0.5	0
Currency	Update frequency	Multiclass evaluation	0.5	0
	Timeliness	Multiclass evaluation	0.8	N/A
Consistency	Compliance with data standards	Binary classification	1	1
Believability	Traceability	System tracking	0.4	0
		Retrieval of versions	0.2	0
	Trustworthiness	Binary classification	0.7	1
Relevance	Data duplication	Binary classification	0.2	1

Regarding the assigned weights, the authors considered compliance with recommended data standards and metrics related to location accuracy to be the most crucial. Other metrics, such as providing additional information, having data duplication, or granting the ability to retrieve historical versions, were deemed less relevant since they do not compromise data reuse. As highlighted throughout the Section 4.2, these values were based on the authors' expertise/judgement and can be changed according to the evaluator's preference.

After filling the computation spreadsheet with those values, the quality indicators (general and respective to each dimension) were obtained. The resulting values are exposed in Table 7.

Table 7 - Case study: resulting quality indicators

Quality dimension	Parameter weight	Quality indicator
Accuracy	0.93	0.88
Completeness	0.63	0.87
Accessibility	0.55	0.00
Currency	0.50	0.00
Consistency	1.00	1.00
Believability	0.43	0.54
Relevance	0.20	1.00
General quality indicator		
		0.66



6. Conclusions

End-users expect comprehensive and accurate UVAR information integrated into their services. However, public authorities often struggle to provide the required data quality, leading to hesitation in making such data available due to concerns about its adequacy for services like navigation. In that sense, commonly agreed-upon definitions for quality criteria and guidelines for their application served as motivation for this work.

This report detailed the development of a QF for UVAR data. Based on a comprehensive literature review and the outcomes of other European UVAR projects, we formulated applicable criteria. Specific metrics for each criterion were designed to enable evaluators to obtain measurable values. We also suggested a method to obtain these metrics for a particular dataset and a procedure to aggregate values for each quality dimension. Finally, the proposed criteria and assessment methods were applied to a real-world dataset to provide future users with a practical example.

This approach emphasized flexibility, acknowledging that overall quality dimensions are not universally applicable and must be assessed for each use case. Different applications and domains may prioritize certain dimensions over others based on specific requirements and operational environments. Consequently, some quality dimensions were excluded from the final set as they were considered out of scope or difficult to measure. Nonetheless, the QF comprises seven quality dimensions, each including one or more criteria and their respective metrics.

Additionally, guidelines for defining the UVAR data QFs are provided, ensuring they accommodate the diverse requirements of UVAR data. Its flexibility extends beyond UVAR data, making it applicable to other quality frameworks and contexts. Organizations across various sectors can leverage the same principles to develop data quality frameworks that address their specific challenges and goals.

Some limitations, however, must be outlined. While the quality criteria were defined based on other EU projects and the expertise of the T3.2 partners, relevant UVAR-related aspects might have been disregarded. Although most metrics are easily applicable, the case study demonstrated that some, such as those pertaining to the location accuracy criterion, require significant effort and GIS skills to obtain. Besides, even though we suggested an assessment method to obtain indicators for each quality dimension, this QF did not establish minimum quality levels for each criterion or dimension. Similar to challenges often encountered in data quality domains, another issue is the necessity of establishing a ground truth. In the context of UVARs, it is not always straightforward to identify a document or publication that can be considered the 'ideal' or 'official' benchmark for evaluating the dataset.

Future work could explore a broader set of relevant criteria, widen baseline studies, and extend the current list of metrics. Including multiple UVARs as case studies would also be important, with the view to evaluating metrics' ranges and the eventual necessity of re-scaling them. Dedicated research to define minimum quality levels, possibly through a maturity level approach, could enhance the QF's usefulness. Furthermore, more detailed instructions may facilitate the application of specific metrics, especially those challenging for an average evaluator.



References

- [1] F. G. Alizamini, M. M. Pedram, M. Alishahi, and K. Badie, “Data quality improvement using fuzzy association rules,” in *International Conference on Electronics and Information Engineering*, 2010, pp. 468–472.
- [2] International Organization for Standardization, “ISO/IEC 25012:2008, Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model,” 2008. Accessed: Jul. 11, 2024. [Online]. Available: <https://www.iso.org/standard/35736.html>
- [3] UVAR Box Consortium, “User-friendly Information Tool on Urban and Regional Access Regulations Schemes. Deliverable 1.1-’UVAR State of Play-Report’.,” 2021. Accessed: Jul. 11, 2024. [Online]. Available: https://uvarbox.eu/wp-content/uploads/2021/12/UVARBox_WP1_Deliverable1.1_V4_07-2021_FINAL.pdf
- [4] A. M. Carvalho, S. Soares, J. Montenegro, and L. Conceição, “Data Quality: revisiting dimensions towards new framework development,” in *6th International Conference on Industry 4.0 and Smart Manufacturing [in press]*, 2024.
- [5] UVAR Box Consortium, “UVAR Box & UVAR Exchange.” Accessed: Jul. 11, 2024. [Online]. Available: <https://uvarbox.eu/>
- [6] NAPCORE, “Metadata,” mobilityDCAT-AP. Accessed: Jul. 11, 2024. [Online]. Available: <https://napcore.eu/metadata/>
- [7] Lisbon City Council, “Zona de Emissões Reduzidas (ZER).” Accessed: Jun. 20, 2024. [Online]. Available: <https://informacoeseservicos.lisboa.pt/servicos/detalhe/zona-de-emissoes-reduzidas-zer>
- [8] Urban Access Regulations in Europe, “Lisbon Urban Access Regulations.” Accessed: Jul. 11, 2024. [Online]. Available: <https://pt.urbanaccessregulations.eu/countries-mainmenu-147/portugal/lisbon>
- [9] DATEX II, “DATEX II webtool”, Accessed: Jul. 01, 2024. [Online]. Available: <https://webtool.datex2.eu/wizard/>



List of Annexes

1. Computation of quality indicators spreadsheet

