A methodological framework for managing and updating controlled vocabularies in mobilityDCAT-AP*

Julien Martineau[1,*], Mario Scrocca[2], Peter Lubrich[3], Petr Bureš[4] and Jasper Beernaerts[1]

[1]National Geographic Institute, Avenue de Cortenbergh 115, Brussels, 1000, Belgium
[2]Cefriel, Viale Sarca 226, 20126 Milano, Italy
[3]Bundesanstalt für Straßen- und Verkehrswesen (BASt), Brüderstraße 53, 51427 Bergisch Gladbach, Germany
[4]TamTam Research s.r.o., Bergerova 1117/4, Praha 8, 182 00, Czech Republic

Abstract

Controlled vocabularies are essential to ensure semantic consistency and interoperability in heterogeneous metadata ecosystems such as European mobility data platforms. Within mobilityDCAT-AP, a reference metadata specification for mobility data portals, they are treated as core components enabling harmonised metadata and cross-border data exchange across National Access Points. This paper outlines a collaborative framework for their governance and maintenance within the NAPCORE initiative, coordinated by a dedicated team representing multiple member states. Based on SKOS representations and a lightweight Excel-driven workflow, producing both RDF serialisations and human-readable documentation, the approach supports both domain expert contributions and automated publication. This framework illustrates a scalable and reusable approach to managing controlled vocabularies as key semantic assets in domain-specific metadata standards, for resolving persistent issues relating to metadata quality, including inconsistent implementation and limited semantic interoperability.

Keywords

mobilityDCAT-AP, DCAT-AP, National Access Points, NAPCORE, ITS Data Exchange, semantic web standards, controlled vocabularies, metadata interoperability

1. Introduction

Controlled vocabularies play a foundational role in metadata ecosystems, as they support semantic consistency, interoperability, and discoverability by constraining the set of admissible values for selected metadata properties [1]. Their role is particularly relevant in heterogeneous and federated settings, where metadata are produced and exposed by multiple organisations and need to remain understandable and reusable across platforms. In the mobility domain, this need is especially evident due to the fragmentation of data portals and National Access Points (NAPs), which publish metadata about datasets and services with different scopes, interfaces, and modelling practices [2]. In such a context, controlled vocabularies are essential not only to reduce ambiguity in metadata values, but also to enable more effective cross-portal discovery, harmonised indexing, and machine-actionable metadata processing.

Within this landscape, mobilityDCAT-AP[1] has been defined as a reference metadata specification for mobility data portals, building on DCAT-AP and GeoDCAT-AP to foster interoperable metadata publication and exchange in Europe [3, 4, 5]. The development of mobilityDCAT-AP showed that controlled vocabularies cannot be considered a secondary implementation detail. Rather, they constitute a core component of the specification, as they operationalise many of the domain agreements needed to describe mobility datasets consistently and reusable. In particular, in the context of NAPs they are also

SEMANTICS 2026 : 22nd International Conference on Semantic Systems, September 15–17, 2026, Ghent, BE
*Corresponding author.
julien.martineau@ngi.be (J. Martineau); mario.scrocca@cefriel.com (M. Scrocca); lubrich@bast.de (P. Lubrich); petr.bures@tamtamresearch.com (P. Bureš); jasper.beernaerts@ngi.be (J. Beernaerts)
https://ngi.be (J. Martineau)
0000-0002-8235-7331 (M. Scrocca); 0000-0002-0023-1234 (P. Lubrich); 0000-0002-3797-0970 (P. Bureš)

[1]mobilityDCAT-AP. Available at: https://w3id.org/mobilityDCAT-AP

relevant to perform an assessment of compliance with respect to the Delegated Regulations defining which type of data must be made available by EU Member States.

The identification of the relevant value spaces required a combination of requirement elicitation, analysis of existing standards and regulations, reuse of available vocabularies whenever possible, and the introduction of new vocabularies only where mobility-specific needs were not already covered. This process led to the definition and maintenance of several controlled vocabularies[2] covering key metadata aspects such as transport mode, mobility theme, update frequency, intended information service, and access conditions [5].

This paper focuses on the methodological framework adopted for the creation, governance, evolution, and publication of controlled vocabularies in the mobilityDCAT-AP context. Rather than discussing controlled vocabularies only as supporting artefacts, we treat them as first-class semantic resources whose lifecycle must be explicitly managed through collaborative processes, transparent governance, and appropriate technical publication solutions. In particular, we describe how stakeholder contributions were collected and consolidated, how a reuse-first strategy was applied in the definition of the vocabularies, and how the resulting artefacts were implemented and maintained through SKOS-based representations and spreadsheet-supported editorial workflows [6, 5]. By reflecting on this experience, the paper aims to provide not only an account of the work carried out for mobilityDCAT-AP, but also a reusable methodological reference for other domain-specific metadata specifications that need to manage controlled vocabularies in a sustainable and interoperable way.

2. Context and motivation

NAPCORE[3] is a European initiative that coordinates and harmonizes more than 30 National Access Points (NAPs) to improve mobility data exchange across Europe. It works to standardize data formats, align technical interfaces, and strengthen interoperability. The organization also supports EU Member States in implementing common procedures and strategies, helping build a long-term, unified platform for intelligent mobility.

The NAPs were launched with the European ITS directive 2010/40/EU of 7 July 2010 which requires Member States to ensure the availability, collection and regular updating of road, traffic and travel data. It also requests the member states to coordinate to develop harmonised specifications (including standardised terminology) ”to ensure the compatibility, interoperability and continuity for the deployment and operational use of ITS” (art. 6 in [7]).

Concerning metadata, an initial NAP-focused initiative for harmonisation brought to the definition of the Coordinated Metadata Catalogue (CMC) [8]. However, such a specification was not aligned with best practices for machine-readable, interoperable data catalogues in Europe, as defined by the DCAT-AP [9]. Within NAPCORE, a dedicated working group defined the mobilityDCAT-AP specification, a mobility-related extension of DCAT-AP that takes domain-specific requirements into account. The full methodology followed for the definition of the mobilityDCAT-AP and a description of the resources made available can be found in [5].

The NAPCORE initiative still has a dedicated working group, which consists of 13 member states, responsible for maintaining and evolving mobilityDCAT-AP and related resources. Within this working group, the *Vocab Team* subgroup is responsible for maintaining controlled vocabularies and regularly meets to discuss emerging issues and proposals[4].

[2] mobilityDCAT-AP Controlled Vocabularies. Repository at https://github.com/mobilityDCAT-AP/controlled-vocabularies
[3] National Access Point Coordination Organisation for Europe (NAPCORE). https://napcore.eu/
[4] GitHub issues are used to keep track discussions and are publicly accessible at https://github.com/mobilityDCAT-AP/controlled-vocabularies/issues

3. Purpose and role of controlled vocabularies

Controlled vocabularies are used to provide a specific set of values for certain properties in the mobilityDCAT-AP specification, ensuring that metadata owners use consistent, predefined values rather than arbitrary keywords [5].

This consistency is essential for :

- Dataset findability, by enabling predictable search and filtering
- Interoperability, by ensuring that different publishers use identical terms
- Semantic clarity, by avoiding ambiguous or overlapping labels

However, studies have revealed a significant number of shortcomings in metadata at the European level. Of the 39 catalog records from 31 NAPs studied, when graded on a scale from A to F, none received an A [10]. 36 % of them have invalid XML distributions, and a lack of XML schemas is observed in 52 % of the publishers.

The authors highlight shortcomings in semantic interoperability (for example, only 1 out of 5 NAPs returns results for standard URI queries) and gaps in distribution (missing access URLs, ambiguous formats, 0 % coverage of recommended properties) [10].

Finally, inconsistent implementation across countries (different ways of referencing vocabularies), such as the use of free text instead of URIs in some countries, or gaps in certain key properties, significantly complicates automatic interoperability and hinders the integration of mobility data on a European scale [11].

These findings directly illustrate the idea that controlled vocabularies are not just a matter of semantics, but a prerequisite for machine-readable cross-border interoperability. However, they face systemic issues related to metadata that are being addressed within the Vocab Team. The harmonization of vocabularies therefore has a fundamental impact.

4. mobilityDCAT-AP controlled vocabularies

For the definition of mobilityDCAT-AP controlled vocabularies, the team, wherever possible, reuses existing vocabularies provided by the EU Publications Office[5], ensuring alignment with broader European interoperability frameworks. However, mobility specific needs often require extensions or modifications. The team follows a collaborative process with the Publications Office to propose and implement modifications and extensions.

When no suitable external vocabulary exists, the team creates custom vocabularies hosted under the mobilityDCAT AP namespace. Today 11 mobility specific vocabularies have been implemented using RDF serialization. These vocabularies are also accompanied by human-readable HTML documentation. These vocabularies cover specialized mobility concepts such as mobility theme categories, georeferencing methods, and application layer protocols. The full list of vocabularies with the corresponding namespace is available on the dedicated GitHub repository[2].

The Vocab team ensures that the vocabularies remain aligned with EU delegated regulations (e.g., RTTI), sector-specific standards, and NAP uses cases. This alignment requires adding missing regulatory terms, clarifying naming conventions, avoiding breaking changes, and deprecating outdated terms.

5. Implementation and publication workflow

Controlled vocabularies in mobilityDCAT-AP are represented using SKOS, the W3C recommendation for publishing and sharing knowledge organization systems such as taxonomies, thesauri, and authority lists on the Web [6]. SKOS is particularly relevant in this context because it provides a lightweight and interoperable model for identifying concepts through URIs, associating multilingual lexical labels

[5] https://op.europa.eu/en/web/eu-vocabularies/controlled-vocabularies

and definitions, and organising them through semantic relations such as hierarchical and associative links [6]. This makes it well suited for controlled vocabularies that must be both human-manageable and machine-actionable, and that need to be reused across metadata records, portals, and software components.

To support the collaborative maintenance of such vocabularies, we adopted an Excel-based workflow that lowers the entry barrier for domain experts who may not be familiar with RDF or ontology engineering tools. Each controlled vocabulary is maintained in a dedicated Excel file, where contributors can edit the list of terms together with their preferred labels, definitions, notes, and references to the relevant delegated regulations or domain sources. This approach reflects one of the key lessons also observed in the broader mobilityDCAT-AP work, namely that domain experts must be enabled to contribute to semantic artefacts through tools that match their skills and working practices [5].

The Excel file is then used as the single editorial source for an automated publication pipeline. Following the template repository adopted for mobilityDCAT-AP controlled vocabularies[6], each vocabulary is managed separately and published independently from the main specification, so that its lifecycle can evolve autonomously in terms of issues, releases, and versioning. The pipeline relies on `xls2rdf` to transform the Excel content into RDF, including SKOS serialisations of the controlled vocabulary [12].

Once the RDF representation has been generated, the workflow automatically produces human-readable documentation pages and prepares the artefacts for publication. For this purpose, the pipeline uses WIDOCO, a well-established tool for the automatic generation of enriched HTML documentation from RDF vocabularies and ontologies [13]. In addition, when the vocabulary contains hierarchical structures with more than one level, `SKOS Play` can be used to generate a navigable hierarchical visualisation of the taxonomy useful both for internal validation and for external dissemination [14].

Controlled vocabularies are hosted in the dedicated GitHub repository and published online via GitHub Pages, following the same publication strategy adopted for mobilityDCAT-AP [5]. Persistent `w3id.org` URIs are used to decouple identifiers from the hosting platform, while redirection rules support content negotiation so that the same resource can provide either human-readable HTML documentation or machine-readable RDF serialisations such as Turtle, JSON-LD, and RDF/XML. This approach ensures stable access, interoperability, and independent versioning of each controlled vocabulary within the overall mobilityDCAT-AP ecosystem.

6. Governance, voting and approval

Controlled vocabularies follow a separate publication cycle from the mobilityDCAT-AP specification. This autonomy makes it possible to quickly release urgent updates, respond flexibly to regulatory changes, and implement incremental improvements without waiting for major releases.

The process for proposing changes involves editing the Excel file, and then opening a corresponding issue on GitHub to document and discuss the proposal. This makes the update process collaborative and transparent. Decisions are then discussed and voted on during an online meeting. Wherever applicable, the Vocab team tries to involve other groups in NAPCORE that have domain expertise on the topic of a certain vocabulary.

Once consensus is reached, major changes with modifications/deletions of values are voted within T4.4 meetings. Regulatory mandated additions may bypass voting. Changes are finally approved by the WG4.4. and published. Multiple changes can be consolidated into a single voting phase to reduce the administrative burden and limit the number of consultations. Additions directly mandated by European delegated regulations can be incorporated without a full voting procedure when they are considered mandatory to ensure legal compliance. The system thus combines participatory governance with the ability to adapt quickly to regulatory requirements.

In the case of renaming or restructuring, the recommended practice is to retain the old terms as "deprecated" and to provide an explicit reference to the new value ("replaced by"). This approach minimizes compatibility issues and facilitates the transition for users.

[6] https://github.com/cefriel/controlled-vocabulary-template

7. Case studies: specific vocabulary updates

For example, the methodological framework has made it possible to develop a new controlled vocabulary, motivated by the entry into force of the European AFIR (Alternative Fuels Infrastructure Regulation), which requires the provision of standardized data on charging and refuelling infrastructure for alternative fuels. Existing terms do not sufficiently cover the distinctions introduced by AFIR, particularly between fuel types, station types, and data types (static, dynamic).

The development process resulted in the creation of a new controlled vocabulary rather than reorganizing existing terms, particularly to reflect the AFIR distinctions. The name and content of the vocabulary were refined through ongoing discussions: adding or removing terms, clarifying the hierarchy, choosing to avoid acronyms where possible, and aligning with other popular mmobility-related standards such as DATEX II.

These issues highlight the difficulty of balancing regulatory precision, ease of use, and consistency with existing vocabularies. The final proposal remains open to adjustments, but it establishes a structure more aligned with AFIR, with terms covering both static and dynamic data, and a clearer differentiation between station types and energy types.

8. Conclusion

Controlled vocabularies in mobilityDCAT-AP ensure semantic consistency and interoperability by providing standardized values that support harmonized metadata publication across European mobility stakeholders. The work of the NAPCORE Vocab Team is the result of the pragmatic approach of member states. The experience gained demonstrates that the successful development of controlled vocabularies requires not only technical modelling choices, but also a structured governance framework capable of supporting collaboration. The approach of giving preference to the reuse of existing semantic resources whenever possible, while introducing mobility-specific vocabularies only when no suitable alternative is available, helps promote interoperability beyond the mobility sector and strengthens alignment with broader European initiatives on semantic interoperability. At the same time, evolving regulations and emerging mobility services are constantly generating new semantic requirements that cannot always be met by existing vocabularies alone, making the governance of controlled vocabularies an ongoing rather than a one-time activity.

From an implementation perspective, the adoption of SKOS representations, persistent URIs, automated publication pipelines, and independently versioned repositories has enabled the publication of vocabularies that are both machine-readable and easily accessible to human users. Equally important, the Excel-based editorial workflow has lowered the barrier to participation for domain experts, allowing semantic resources to be maintained through processes that remain accessible to non-specialists while preserving semantic quality and technical consistency.

Beyond the specific context of mobilityDCAT-AP, the methodological framework presented in this paper provides a practical example of how controlled vocabularies can be managed as first-class semantic artefacts throughout their entire lifecycle. The combination of collaborative governance, standards-based implementation, transparent publication, and incremental evolution offers a reusable model for other domain-specific metadata initiatives facing similar challenges of interoperability, regulatory compliance, and long-term sustainability. As European mobility data ecosystems continue to expand and integrate, the role of well-governed controlled vocabularies will become increasingly important in enabling reliable data exchange and interoperable digital infrastructures.

Acknowledgments

This work has been carried out in the context of Sub-working Group 4.4 of the NAPCORE project. The authors want to thank the members of the NAPCORE T4.4 Vocab Team who actively contribute to the maintenance of the controlled vocabularies.

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] R. Albertoni, D. Browning, S. Cox, A. G. Beltran, A. Perego, P. Winstanley, Data catalog vocabulary (dcat) – version 2, https://www.w3.org/TR/vocab-dcat-2/, 2020. W3C Recommendation.

[2] M. Scrocca, A. Azzini, P. Bureš, M. Comerio, P. Lubrich, Towards napDCAT-AP: Roadmap and Requirements for a Transportation Metadata Specification, in: Proceedings of Poster and Demo Track and Workshop Track of the 18th International Conference on Semantic Systems, volume 3235 of *CEUR Workshop Proceedings*, CEUR, Vienna, Austria, 2022. URL: https://ceur-ws.org/Vol-3235/paper22.pdf, iSSN: 1613-0073.

[3] B. V. Nuffelen, DCAT Application Profile for Data Portals in Europe (DCAT-AP) v2.0.1, Technical Report, SEMIC, 2020. URL: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0.

[4] GeoDCAT-AP Working Group, Geodcat-ap, 2020. URL: https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/, version 2.0.0.

[5] M. Scrocca, L. Molinas Comet, B. Witsch, D. M. Mustafa, C. Lange, M. Comerio, P. Lubrich, mobilitydcat-ap: A metadata specification for enhanced cross-border mobility data sharing, in: E. Curry, M. Acosta, M. Poveda-Villalón, M. van Erp, A. Ojo, K. Hose, C. Shimizu, P. Lisena (Eds.), The Semantic Web, Springer Nature Switzerland, Cham, 2025, pp. 226–243.

[6] A. Miles, S. Bechhofer, SKOS simple knowledge organization system reference, W3C recommendation, W3C, 2009. URL: http://www.w3.org/TR/skos-reference.

[7] European Commission, Directive 2010/40/eu on intelligent transport systems, 2010. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010L0040&from=EN&ct=1781084625563, official Journal of the European Union.

[8] P. Lubrich, J. Ansorge, M. Böhm, L. Hendriks, T. Hoffmann, J. Jaderberg, A. Kochs, C. Lüpges, EU EIP SA46 Coordinated Metadata Catalogue: Monitoring and Harmonisation of National Access Points in Europe, Technical Report 2.0, EU EIP, 2019. URL: https://www.its-platform.eu/wp-content/uploads/ITS-Platform/AchievementsDocuments/NAP/EU%20EIP_Coord.%20Metadata%20Catalogue_v2.0_191115.pdf.

[9] B. Van Nuffelen, DCAT Application Profile for data portals in Europe (DCAT-AP) v2.0.1, Technical Report, SEMIC, 2020. URL: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0.

[10] C. Mylonas, P. Bureš, M. Stavara, J. Vlčinský, A. Vouitsis, E. Mitsakis, Evaluating barriers to pan-european traffic data exchange: Metadata, structural quality, and consumer usability and accessibility, in: 12th International Congress on Transportation Research (ICTR 2025), Selected Papers, Thessaloniki, Greece, 2025. October 16–18, 2025.

[11] N. Vandekeybus, J. Beernaerts, M. Dierick, K. Le Thien Marie Hoang, mobilityDCAT-AP in practice: Compliance analysis of european NAPs and the belgian reference implementation, in: 17th ITS European Congress, Selected Papers, Istanbul, Turkey, 2026. URL: https://github.com/user-attachments/files/27516897/mobilityDCAT-AP.in.Practice.Vandekeybus.et.al.ITS.Istanbul.2026.pdf, april 27–29, 2026.

[12] Sparna, xls2rdf: Create rdf data from excel spreadsheets, 2026. URL: https://github.com/sparna-git/xls2rdf, software and documentation for converting Excel spreadsheets to RDF, including SKOS vocabularies.

[13] D. Garijo, WIDOCO: a wizard for documenting ontologies, in: International Semantic Web Conference, Springer, Cham, 2017, pp. 94–102. doi:10.1007/978-3-319-68204-4_9.

[14] T. Francart, Skos play!, 2026. URL: https://skos-play.sparna.fr/play/home?lang=en, software for rendering, visualising, validating, and converting SKOS vocabularies, including Excel-to-SKOS support.