Enabling Cross-Border Travel Offers Through National Access Point Federation via Metadata Harmonisation

Alessio Carenini, Andrea Fiano, Mario Scrocca, Marco Comerio, and Irene Celino

Cefriel, Milan, Italy name.surname@cefriel.com

Abstract. Planning cross-border transportation offers requires gathering data from multiple transport operators within and outside a country. The European legislation demands each member state to set up a National Access Point (NAP) for multimodal transport information, nevertheless, interoperability in accessing data from different NAPs is far to be accomplished. In this paper, we describe our approach to consolidate metadata coming from three different European National Access Points in a single metadata catalog using Semantic Web technologies.

Keywords: Metadata Harmonisation · Cross-border Travel Offers · National Access Points

1 Introduction

In the transportation domain, several data and metadata catalogs coexist, each one being maintained by a different initiative or mandated by a specific EU directive or country law. According to the EU Delegated Regulations 2017/1926, 885/2013, 886/2013 and 2015/962, each EU member state has to implement a National Access Points (NAP) to make national transport data discoverable. A NAP is an intermediary digital platform allowing access to traffic and mobility data, and playing a crucial role in data exchange in the field of mobility in Europe. From the point of view of the transport operator looking for mobility-related information, NAPs represent trusted sources of data and metadata, and their content can be reliably used inside their own information systems. A NAP is a web-based portal handling data concerning Safe and Secure Truck Parking's (SSTP), Real-Time Traffic Information (road) (RTTI), Safety Related Transport Information (road) (SRTI) and Multimodal Travel Information (MMTIS) (all modes like train, busses, metro, cycling etc.). EU regulations mandate the usage of Transmodel-based specifications for the data exchange between transport operators and their own reference NAP, therefore aiming for data interoperability. Nevertheless, the regulations don't specify which metadata should be used to describe datasets, and how a NAP should be implemented. As a result, each member State is implementing its own National Access Point using different metadata schemas and exposing its functionalities via custom APIs [1,4].

In the following, we will describe how we extended a metadata catalog, named Asset Manager, to seamlessly support accessing both *local* digital assets, directly added by users of the Asset Manager, and *remote* digital assets from multiple National Access Points. Our scenario is based upon the real requirements coming from Trenitalia[1], which wants to create mobility packages to be sold to tour operators bringing tourists to the Milano-Cortina Winter Olympics in 2026[2]. Creating such mobility packages means locating and accessing timetables of multiple transport operators. Performing this task, even in the case National Access Points are available, is time-consuming and requires checking multiple sources. Consolidating in a single catalog metadata coming from multiple NAPs, together with metadata provided by Trenitalia, means being able to perform more efficiently the task and better fulfil the mobility needs of tourists heading to the Winter Olympics. Such a scenario requires mapping different NAP metadata schemas onto the selected one (DCAT-AP 2.0.1 [7]) and creating multiple metadata ingestion pipelines.

The steps which we implemented were the following:

i) *Metadata schema mapping*: different metadata schemas were mapped onto a single schema, which is also used to describe the *local* assets.

ii) *RML transformation rules*: the conceptual mappings from the specific NAP schemas to the Asset Manager metadata schema were implemented in RML [2].

iii) *(Meta)Data ingestion pipelines*: the RML mappings were integrated in data ingestion and transformation pipelines using the Chimera tool for their execution[3]. The resulting RDF triples, defining metadata for *remote* assets, are added to the RDF repository used by the Asset Manager.

iv) *Exploration API creation*: to ease integration in the user interface, Exploration APIs were created to wrap the execution of SPARQL queries as APIs. Such Exploration APIs allow obtaining the lists of assets belonging to a specific type and their metadata. By doing this, we harmonised the access to both *local* and *remote* assets.

v) *User interface*: the Asset Manager web interfaces, showing the integrated list of assets and their metadata, were updated.

To show the implemented approach for NAP metadata harmonisation, we selected three different NAPs from France, Belgium and the Netherlands. Since the approach is completely generic, this implementation opens the possibility to use the Asset Manager as an aggregator of multiple trusted metadata sources, like open data portals, multimodal National Access Points, or other instances of the Asset Manager. In the following sections, we provide details on each step of the approach and meaningful insights about the implemented solution.

[1] `https://www.trenitalia.com/`

[2] A video describing the scenario and the implemented solution is available at `https://www.youtube.com/watch?v=SoOLheMv1wQ`

[3] `https://github.com/cefriel/chimera`

2 Metadata schemas for National Access Points

The main focus of the NAP regulations is to promote the usage of a specific set of standards, based on Transmodel, across all Europe to improve transport data interoperability. Even though the role of the NAP as a dataset catalog is well-defined by the regulation, each member state is then free to define its own implementation. Such principle led to the appearance of different metadata vocabularies, and the need for interoperability between the metadata schemas adopted by different NAPs. In our scenario, we analysed in details the National Access Points provided by France, Belgium and the Netherlands.

Belgian NAP is built upon CKAN, therefore its API[4] allows for searching for datasets according to specific types or features. The metadata schema is quite rich and contains multi-lingual documentation, geographical coverage, and both contact person and responsible transport operator.

French NAP features a rich API[5] and metadata schema, containing many details about datasets. This National Access Point supports NeTEx representation of static transport data (leveraging on Chouette [3] features), which are made available as *Community resource*, which are alternative representations of the same main information described in the asset. Also, spatial information about the covered area is provided, allowing for geographical queries. As a last detail, an asset has only a responsible organization and the metadata does not mandate for a contact person.

Netherlands NAP has no clear API to obtain metadata, and the actual endpoint[6] has been found by analyzing the JavaScript sources of the NAP website. The metadata schema mandates both a responsible person for the dataset publication and an owner transport operator company. The referenced dataset is listed with the attribute *publicationURL*, and no geographical coverage is present (as opposed to France metadata schema).

Summarising the analysis of the selected NAPs, they all feature different metadata schemas, and even basic information describing who is responsible for the asset is not represented in the same way.

A working group composed of representatives from the Netherlands, Germany, Austria and Sweden started to work on common metadata definitions to be applied to the various NAPs in Europe to increase interoperability and ease the creation of multi-country solutions. The outcome of such group is called *Coordinated Metadata Catalogue*[7] and defines a minimum set of metadata which, according to its authors, should be supported in all the NAP implementations. Using the Coordinated Metadata Catalogue schema allows harmonising those NAP schemas onto a unified schema, as all the basic information contained in the assets coming from the three different NAPs can be represented.

[4] `https://www.transportdata.be/api/3`
[5] `https://transport.data.gouv.fr/swaggerui`
[6] `https://nt.ndw.nu/services-spoa/rest/v1/ui/multimodaal`
[7] `https://www.its-platform.eu/highlights/harmonised-metadata-national-access-points`

3 Automating Metadata Aggregation from National Access Points

The Asset Manager is an RDF-based metadata catalog developed in the context of the Shift2Rail Innovation Programme 4[8]. We show the possibility to use it as an aggregator of metadata coming from multiple trusted sources. The objective is to let companies accessing domain-specific knowledge in a coherent way using a single tool. We defined and validated an approach based on Semantic Web technologies to perform metadata ingestion, to define and execute mappings to a single metadata schema.

Our solution leverages on the Asset Manager and the Chimera tool [5] to: (i) connect to each NAP, (ii) fetch the metadata of its assets, (iii) convert such metadata into a coherent RDF representation to be easily queried via SPARQL, (iv) store the resulting triples inside the RDF repositories, (v) show that the Asset Manager can visualise both *local* and *remote* assets.

3.1 Metadata ingestion

The first and most important part of the NAP metadata ingestion process is understanding the metadata schemas and identify which attributes and data structures can be found in all the different NAPs. We decided to use the Coordinated Metadata Catalogue as a first step in understanding metadata schemas for two reasons. First of all, it embodies the will of the EU member states to converge upon a common metadata schema for NAPs to ease the implementation of ITS solutions. From a practical point of view, the Coordinated Metadata Catalogue specifications acknowledge the existence of other vocabularies, and already provide an alignment to DCAT-AP. We could therefore exploit such alignment, obtaining a two-step conceptual mapping, as depicted in Figure 1: first from the specific NAP metadata schema onto Coordinated Metadata Catalogue, and then from that schema onto DCAT-AP, which is also the same metadata schema used by the Asset Manager.

[8] cf. `https://shift2rail.org/research-development/ip4/`

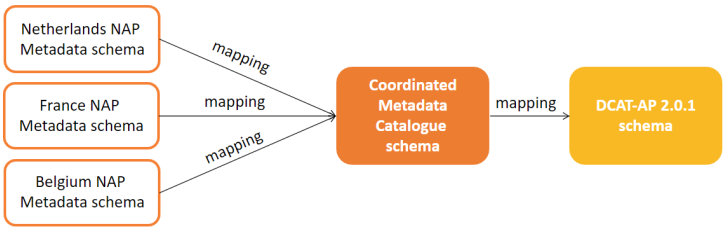

Fig. 1: Conceptual mappings defined for the harmonisation of the different National Access Point metadata schema.

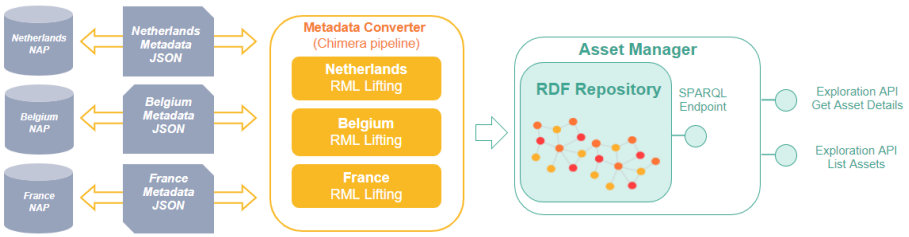

Fig. 2: Overall architecture of the implemented prototype to integrate metadata from the NAPs (France, Belgium and Netherlands).

After creating the RML mapping rules for the selected countries[9], we assembled a metadata ingestion service exposed through a Chimera pipeline[10]. Calling such service triggers the execution of the following actions for each of the countries (as shown in Figure 2):

- The NAP API endpoint is called to obtain JSON metadata
- Lifting is performed on the resulting JSON metadata using the appropriate RML mapping rules and obtaining an RDF representation compliant with the DCAT-AP profile
- The resulting RDF triples are written in the RDF repository used by the Asset Manager as a separate RDF graph

3.2 Accessing metadata from the Asset Manager

The Asset Manager arranges assets in categories according to so-called *asset types*. In the considered scenario, we mapped the items coming from the NAPs metadata ingestion pipelines to the *journey planning* asset type, which can be used to describe either datasets containing timetables or services providing timetables. Whenever a user asks for viewing the list of *journey planning* assets, the Asset Manager performs a single SPARQL query[11] to retrieve the basic information about each published asset.

As can be noticed in Figure 3, when the NAPs metadata ingestion pipeline is activated, the Asset Manager starts showing both local assets and assets coming from National Access Points. This enables users to browse through the consolidated list of assets and to search for the most interesting ones. As a result, this solution opens up the possibility to create multimodal mobility packages accessing information coming from several metadata sources using a single tool.

[9] The developed RML mappings are available at `https://github.com/cefriel/nap-harmonisation/tree/main/rml`

[10] The configuration of the ingestion service is available at `https://github.com/cefriel/nap-harmonisation/blob/main/chimera-route/camel-context.xml`

[11] The query is available at `https://github.com/cefriel/nap-harmonisation/blob/main/asset-manager/query-visualise-assets.sparql`

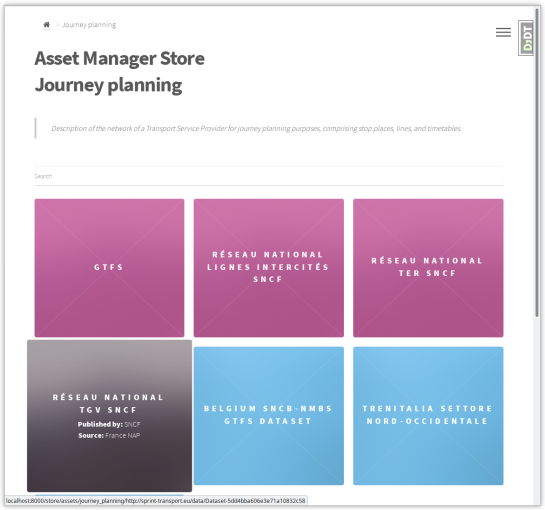

Fig. 3: Visualisation of both *local* and *remote* NAP assets in the Asset Manager.

4 Conclusions and Future Works

The general availability of the National Access Points throughout Europe will improve interoperability in the transportation domain, as it will force all actors to provide data according to the Transmodel-based specifications dictated by the regulators. We demonstrated that converging to a common set of metadata (such as the one proposed in the Coordinated Metadata Catalogue initiative) enables the possibility to treat the entire network of National Access Points as a source of trusted data and metadata which can be exploited to improve cross-border services and plan better travel offers.

The integration of remote metadata providers (such as the NAPs) in the IT systems of a transport operator is an operation which must carefully follow the data quality assurance and the information lifecycle processes defined inside the company. Future works related to the Asset Manager will investigate how to integrate the detection of changes in the metadata acquired from NAPs inside the lifecycle processes of other assets managed by the tool. Since NAPs will become the authoritative source of information in the transportation domain, it is important to promptly detect the availability of new versions of a remote asset used internally by the company (through the Asset Manager) notifying the owners of the dependant applications to check their functionalities and prevent errors.

Although based on a declarative approach, our prototype exploits an external integration engine to perform the actual calling of the API provided by the NAPs. We will investigate the recent support introduced in RML for Web APIs [6] as an alternative solution to define mappings for different NAPs endpoints in a fully declarative way.

Acknowledgments

The presented research was supported by the SPRINT project (Grant Agreement 826172), co-funded by the European Commission under the Horizon 2020 Framework Programme.

References

1. Carenini, A., et al.: SPRINT project Deliverable D2.3 – Requirements for an IF architectural design (F-REL) (2020), `http://sprint-transport.eu/`
2. Dimou, A., Vander Sande, M., Colpaert, P., Verborgh, R., Mannens, E., Van de Walle, R.: RML: a generic language for integrated RDF mappings of heterogeneous data (2014)
3. Gendre, P., Denis, Y., Duquesne, C., Bouziane, Z., Bouree, K., Dezou, L., Lemettais, O.: Chouette an open source software for pt reference data exchange. In: ITS Europe. Technical Session (2011)
4. Mylonas, C., Mitsakis, E., Dolianitis, A., Aifadopoulou, G.: A review of european national access points for intelligent transport systems data. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). pp. 1–8. IEEE (2020)
5. Scrocca, M., Comerio, M., Carenini, A., Celino, I.: Turning transport data to comply with eu standards while enabling a multimodal transport knowledge graph. In: International Semantic Web Conference. pp. 411–429. Springer (2020)
6. Van Assche, D., Haesendonck, G., De Mulder, G., Delva, T., Heyvaert, P., De Meester, B., Dimou, A.: Leveraging web of things w3c recommendations for knowledge graphs generation. In: Web Engineering. pp. 337–352 (2021)
7. Van Nuffelen, B.: Dcat application profile for data portals in europe. Tech. rep. (2020), https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0