Turn Transportation Data into EU Compliance through Semantic Web-based Solutions

Marco Comerio, Alessio Carenini, Mario Scrocca, and Irene Celino
Cefriel – Politecnico di Milano
viale Sarca, 226 – Milano, Italy
E-mail: {marco.comerio,alessio.carenini,mario.scrocca,irene.celino}@cefriel.com

Abstract. The establishment of National Access Points (NAPs) to transportation data, as requested by the EU Regulation 2017/1926, represents a first step towards the realization of semantic interoperable travel information services. This paper presents challenges and opportunities introduced by the Regulation and its impact on transport stakeholders. The design and development of solutions based on Semantic Web technologies for supporting the establishment of NAPs represent research and business opportunities for the Semantic Web community.

1 Introduction

Semantic interoperability in the transportation sector is one of the European Commission challenges: establishing an interoperability framework enables European travel and transport industry players to make their business applications 'interoperate' and provides the travelers with a new seamless travel experience.

The EU-report [1] on the provision of EU-wide multimodal travel information services highlighted several barriers for the realization of comprehensive travel information services in the EU, such as insufficient accessibility of travel and traffic data and the lack of travel and traffic data interoperability. Key enablers to address these barriers are (i) ensuring that users have *access* to the right scope of data and information with the appropriate data sharing mechanism, and (ii) making travel and traffic data *interoperable* with a common set of data exchange standards.

A first step towards the realization of multimodal travel information services is the establishment of National Access Points (NAPs) to multimodal transportation data as defined in the recent Commission Delegated Regulation (EU) 2017/1926 [2]. This Regulation establishes the specifications necessary to ensure the accessibility, exchange, and update of transportation data for the provision of multimodal information services in the European Union. In the Regulation, each EU Member State is required to set up a NAP accessible through APIs that allows access to static data (e.g., timetables, network topology, list of services offered at a station / airport) and dynamic data (e.g., delays, cancellations) relating to different transportation modes (air, train, road vehicle, bus, ferry, metro, tram, shuttle bus, car-sharing, car-pooling, bike-sharing, etc.). In order to foster interoperability across Europe, the Regulation requires to provide data to the NAP defined according to specific standard data formats such as *NeTEx CEN / TS*

16614[1] and *SIRI CEN / TS 15531*[2] and described using national application profiles (e.g., *DCAT-AP* [3]). This EU Regulation 2017/1926 has been already adopted (October 2017), and the first deadline for the provisioning of specific types of static transportation data through NAPs is December 1st, 2019.

Different transport stakeholders (i.e., transport authorities, transport operators, and infrastructure managers) shall be able to continue to use their current systems at the national level (often based on languages / formats different from *NeTEx* and *SIRI*), but they will have to provide their data to the NAP after translation and conversion. Unfortunately, as emerged from the EU EIP SA46 Annual NAP Report 2018 [3], very few transport stakeholders are ready to provide their data and services in compliance with the standards requested by the EU Regulation. To enable a data conversion process, they rely on their in-house support (which may lack knowledge and skills related to the Regulation) or turn to external technology providers that provide custom and often expensive solutions.

The challenges in the transportation sector, as well as the full compliance with the EU Regulation, clearly call for Semantic Web-based solutions: in this paper, we summarize the opportunities for semantic interoperability, and we give some hints on our work to achieve such goal.

This paper is organized as follows. Section 2 describes the impact of the Regulation on transport stakeholders in terms of obligations, challenges, and benefits. Section 3 highlights research and business scenarios for the Semantic Web Community. Section 4 outlines ongoing and future work.

2 Requirements from and Impact on Transport Stakeholders

The establishment of NAPs for multimodal transportation data impacts different categories of transport stakeholders. In this section, we describe obligations, challenges, and benefits for each category in order to provide a comprehensive analysis of the impact and to facilitate the elicitation of requirements for Semantic Web-based solutions.

- **NAP Management Board**: (i.e., the entity that is in charge of defining and managing the NAP for a specific EU country)
  - *Obligation*:
    - Establish the NAP and defining its governance framework;
    - Define a national metadata profile for multimodal transportation datasets.
  - *Challenge*:
    - Regrouping the existing public and private access points in a single point enabling access to available data that fall within the scope of the Regulation, solving interoperability and compliance issues;

[1] http://netex-cen.eu/

[2] www.transmodel-cen.eu/standards/siri/

        - Provide a metadata profile aligned with the one adopted by other EU NAPs in order to promote interoperability.
    - *Benefit*:
        - The NAP may take various forms such as a database, data warehouse, data marketplace, repository, web portal, or similar depending on the type of data. The NAP Management Board could define a business model covering potential commercial aspects of multimodal data transport sharing, brokerage, and trading.

- **Transport authorities** (i.e., any public authority responsible for the traffic management or the planning, control or management of a given transport network or modes of transport, or both, falling within its territorial competence) and **Transport operators** (i.e., any public or private entity that is responsible for the maintenance and management of the transport service, including transport on-demand service):
    - *Obligation*:
        - Provide static data (e.g., timetables, network topology, fares) and dynamic data (e.g., delays, cancellations) to the NAP compliant to the requested data formats. The full list of data categories to made accessible through the NAP is in Annex I of the EU Regulation [2];
        - Provide the metadata in order to allow users to discover and use the datasets made accessible through the NAPs.
    - *Challenge*:
        - Turn available data into the requested data formats and, in case, complement/enrich them with additional data sources.
    - *Benefit*:
        - Published datasets are interoperable and could be easily used by Intelligent Transport System (ITS) applications. This, of course, potentially augments the knowledge and the adoption of their transport offerings;
        - In case of a NAP realized as data marketplace, transport authorities and operators could receive revenues from the trading of their data.

- **Transport infrastructure managers** (i.e., any public or private body or undertaking that is responsible in particular for establishing and maintaining transport infrastructure, or part thereof):
    - *Obligation*:
        - Provide data (e.g., list of stop facilities access nodes including platform information, help desks/information points, ticket booths, lifts/stairs, entrances, and exit locations) and the related metadata to the NAP compliant to the requested data formats.
    - *Challenge*:
        - Turn available data into the requested data formats and, in case, complement/enrich them with additional data sources.
    - *Benefit*:

        - Published datasets are interoperable and could be easily integrated into ITS applications for e.g., indoor navigation and location-based experiences.

- **Transport Standardization Bodies** (i.e., European committees whose mission is to foster the economy of the EU providing an efficient infrastructure for the development, maintenance, and distribution of coherent sets of standards and specifications):
    - *Obligation*:
        - Notify the NAP Management Board on the release of new versions of the data formats/standards requested by the EU Regulation.
    - *Challenge*:
        - Support the definition of metadata and also the data conversion process providing a reference vocabulary for the transportation sector.
    - *Benefit*:
        - Improve and enhance the adoption of developed standards.

- **Intelligent Transport System (ITS) providers** (i.e., any provider of advanced application which aims to provide innovative services relating to different transportation modes enabling users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks) and **Mobility-as-a-Service (MaaS) providers** (i.e., any mobility provider that aggregates multiple mobility service providers' offerings and supplies them to users through a single digital platform, providing real-time information and online payment options):
    - *Obligation*:
        - Follow terms and conditions established by the NAP Management Board concerning the NAP data usage.
    - *Challenge*:
        - Use datasets and APIs offered by NAPs to enhance their mobility service offerings and to improve their businesses.
    - *Benefit*:
        - The NAP represents a structured solution to obtain interoperable transport data to be elaborated for providing innovative mobility services to users/travelers.

3 The EU Regulation 2017/1926: challenges, opportunities for semantic interoperability

From the analysis of the impact of the EU Regulation 2017/1926 to transport stakeholders, three different opportunities for semantic interoperability emerge:

- ***Definition of reference ontologies for the transportation sector***: the first opportunity for semantic interoperability comes from the need for reference ontologies to unambiguously describe the operational aspects of the transportation domain.

- ***Definition of metadata profiles for transportation datasets***: as described in Section 2, metadata descriptions of datasets according to a specific application profile are required by the EU Regulation. The second opportunity for Semantic Web technologies is to harmonize the metadata description of transportation datasets through the definition of a semantically-enhanced metadata profile. This application profile could be based on DCAT-AP [3], extended with the new concept of *Data Service* introduced by DCAT Version 2 [10] as a collection of operations that provide access to one or more datasets or data processing functions.
- ***Design and development of Semantic Web-based transportation data converters***: the EU Regulation asks to turn available transportation data into specific data formats and, in case, complement/enrich them with additional data sources. Due to the heterogeneity of the used data models, as of today, there are no holistic solutions on the market covering all the needed data conversion and enrichment processes. Therefore, the design and development of transportation data converters based on Semantic Web technologies represent another research and business opportunity for the Semantic Web community. Transportation data converters should be software components supporting the conversion between transportation standards by means of an intermediate step where data are mapped to a common semantic model. As an example, a specific (semantic) converter enables the translation of transportation schedule, geographic and fare information expressed in GTFS to a NeTEx specification preserving the original meaning.

4 Ongoing and Future Work

Due to the delay of several EU countries in establishing the NAP to multimodal transport data according to EU Regulation 2017/1926, transport stakeholders do not perceive the need for data conversion as an urgent task. This effect is increased by the lack of understanding of how the NAP could positively affect their businesses. However, the authors of this paper believe that the needs for reference transportation ontologies, metadata profiles, and transportation data converters will emerge very soon. In this section, we give some hints on our work aiming at moving towards Semantic Web-based solutions for these needs.

The SNAP project[3] is addressing the challenge of defining a reference ontology for the transportation sector. The ongoing activity consists in deriving a reference ontology from a sub-part of the CEN Transmodel[4] (i.e., an abstract model of common public transport concepts and data structures that can be used to build many different kinds of public transport information system). The resulting ontology will represent a starting point for a potential collaboration with the CEN TC278- WG3-SG4[5] that has formally defined the Transmodel.

Concerning the definition of metadata profiles, we proposed the Semantic Asset Manager [7], i.e., an organized collection of asset types (e.g., datasets, APIs) enhanced

[3] www.snap-project.eu
[4] www.transmodel-cen.eu/
[5] www.transmodel-cen.eu/standards-context/

by tools for their publication, governance, and discovery. An enhanced version of the Semantic Asset Manager is under development in the context of the SPRINT project[6] where the description of each asset type is made through a DCAT-AP metadata profile. This enhanced version will be tested in the context of the Shift2MaaS project[7] as an enabling component for MaaS solution providers.

For what concerns the design and development of transportation data converters, a potential solution is based on the any-to-one approach to semantic interoperability described in [8]: legacy data schemas used by transport stakeholders are mapped to a reference ontology, and this ontology is mapped, e.g., to formats/standards requested by the EU Regulation. This approach simplifies the mapping activities and their reusability. We adopted this approach in a different domain [6], and we are bringing and adopting such experience in the transportation domain. An annotation-based solution to data transportation conversion is described in [5], and a new RML-based solution [9] has been recently developed in the context of the SNAP project.

Acknowledgments
This work has been supported by SNAP (EIT Digital-19281), SPRINT (H2020-826172) and Shift2MaaS (H2020-826252).

References

1. EU Commission (2016). Study on ITS Directive, Priority Action A: the Provision of EU-wide Multimodal Travel Information Services. Available at: http://ec.europa.eu/transport/themes/its/consultations/doc/2015-its-mmtips/consultation-report.pdf
2. EU Commission (2017). Commission Delegated Regulation (EU) 2017/1926. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R1926
3. ISA working group (2015). DCAT application profile for data portals in Europe. Available at: https://joinup.ec.europa.eu/system/files/project/dcat-ap_final_v1.00_0.html.
4. Hendriks L., Jorna R., et Al. (2018). Monitoring and Harmonisation of National Access Points in Europe. Available at: www.its-platform.eu/filedepot_download/1971/6491
5. Carenini, A., Dell'Arciprete, U., Gogos, S., Kallehbasti, M., Rossi, M., Santoro, R. ST4RT – Semantic Transformations for Rail Transportation. In Proc. TRA 2018.
6. Della Valle, E., Cerizza, D., Celino, I., ... Monteleone, G. (2007). SEEMP: a semantic interoperability infrastructure for e-government services in the employment sector. *The Semantic Web: Research and Applications*, 220-234.
7. Carenini A., Comerio M., Celino I. (2018). Semantic-Enhanced National Access Points to Multimodal Transportation Data. In Proc. of the Int. Semantic Web Conference 2018.
8. Vetere, G., Lenzerini, M. (2005). Models for semantic interoperability in service-oriented architectures. *IBM Systems Journal*, *44*(4), 887-903
9. Comerio M. An Innovative Semantic Solution to Turn Transport Data into EU Compliance. In Proc. of SEMANTiCS 2019 – Industry & Use Case Presentations.
10. Albertoni R., Browning D., Cox S., Perego A., Winstanley P. (2019). Data Catalog Vocabulary (DCAT) V2 - W3C Working Draft. Available at: www.w3.org/TR/vocab-dcat-2

[6] www.sprint-transport.eu/
[7] www.shift2maas.eu/