Contract No. H2020 – 730844

GOVERNANCE OF THE INTEROPERABILITY FRAMEWORK FOR RAIL AND INTERMODAL MOBILITY

D4.1 Initial Semantic Interoperability Technology Market Watch

Due date of deliverable: 31/12/2017

Actual submission date: 17/01/2018

Leader/Responsible of this Deliverable: Marco Comerio (Cefriel)

Reviewed: Y

Document status		
Revision	Date	Description
1	30/11/17	First issue
2	10/12/17	New version after CNC, UITP, UPM and POLIMI review of chapters 1-4.
3	20/12/17	New contributions from UITP.
4	10/01/18	Correction of typos and missing references.

Project funded from the European Union's Horizon 2020 research and innovation programme		
Dissemination Level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Start date of project: 01/11/2016 Duration: 24 months

EXECUTIVE SUMMARY

Different transport operators (of different modes) use different and often heterogeneous systems to support/implement their transport services. Specifically referring to the digital aspect of transport systems, a number of standards and conventions indeed exist, but they lack a broad and general interoperability; in some cases, interoperability is solved at syntactic level (by agreeing on the use of a common standard or by developing ad hoc adapters for bilateral solutions).

Interoperability is one of the IP4 Shift2Rail challenges. It consists in the establishment of a framework which will enable the business applications belonging to a critical mass of European travel and transport industry players to 'interoperate' so as to provide the customer with: comprehensive information on available transport options and the corresponding processes for their booking, payment, ticketing, consumption, modification, and, more exclusively for the business partners (e.g. transport service providers, distributors, retailers), their financial settlement.

Shift2Rail would like to embrace the opportunity to achieve interoperability at **semantic level** by developing a shared model of the meaning of the exchanged information in the transport domain and by fostering the adoption of multilateral solutions.

The WP4 of the GoF4R project has the global objective to analyse the technology market in the field of semantic interoperability solutions, in order to understand the feasibility of a large-scale adoption of such solutions within the rail and transport domain. In particular, with this deliverable, the GoF4R project:

- Identify the main challenges and barriers for the semantic interoperability of data formats and structures, which is required in order to allow proper data sharing and exchange between different systems, subsystems and applications.
- Analyse the available semantic interoperability tools on the market to assess their maturity and the feasibility to adopt them in the rail and transport sector.
- Survey the degree of adoption of semantic interoperability solutions in different domains, highlighting the success stories, as well as the obstacles and resistance;
- Evaluate the supply market in terms of skills and professional services, in order to understand the potential difficulties to ensure the evolution and maintenance of semantic solutions.

The rest of the deliverable is structured as follows. Section 1 reports the initial analysis of the maturity of the semantic interoperability solutions supply market. Section 2 describes the results of the initial analysis of the adoption of semantic interoperability solutions in the transportation and Public Administration domains. Section 3 reports on the initial analysis of the availability of semantic interoperability skills and professional services supply market. Finally, Section 4 draws some conclusions, identifies input for WP5 and describes the next steps of the final extended analysis that will be described in D4.2 (planned at M24).

ABBREVIATIONS AND ACRONYMS

Abbreviation	Description
ADMS	Asset Description Metadata Schema
API	Application Programming Interface
ATTRACkTIVE	Advanced Travel Companion and Tracking Services
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
Co-Active	CO-modal journey re-ACcommodation on associated Travel serVices
COHESIVE	COHErent Setup and Demonstration of Integrated Travel SerVices
CONNECTIVE	Connecting and Analysing the Digital Transport Ecosystem
CPSV-AP	Core public Service Vocabulary – Application Profile
DCAT-AP	Data Catalog – Application Profile
DW	Data Warehouse
DWO	Data Warehouse Ontology
EC	European Commission
EIF	European Interoperability Framework
ERA	European Union Agency for Railways
ESO	European Standardisation Organization
ETL	Extract, Transform and Load
ETSI	European Telecommunications Standards Institute
EU	European Union
FSM	Full Service Model
GA	Grant Agreement
GO	Global Ontology
GoF4R	Governance of the Interoperability Framework for Rail and Intermodal Mobility
GTFS	General Transit Feed Specification
H2020	Horizon 2020 framework programme
IATA	International Air Transport Association

IEC	International Electrotechnical Commission
IF	Interoperability Framework
InteGRail	Intelligent Integration of Railway Systems
IP4	Innovation Programme 4
ISO	International Organization for Standardization
IT2Rail	Information Technologies for Shift2Rail
ITS	Intelligent Transport Systems
ITU	International Telecommunication Union
JSON	JavaScript Object Notation
JU	Shift2Rail Joint Undertaking
LO	Local Ontology
MAAP	Multi Annual Action Plan
NAP	National Access Point
NaPTAN	National Public Access Node
NeTEx	Network Timetable Exchange
NGO	Non-Governmental Organization
NLP	Natural Language Processing
NSB	National Standardisation Bodies
ODP	Ontology Design Pattern
OWL	Web Ontology Language
PSI	Public Service Information
RaCoOn	Rail Core Ontology
RailML	Railway Markup Language
RDF	Resource Description Framework
REST	REpresentational State Transfer
RTTP	Real-Time Traffic Plan
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organisation System
SPARQL	SPARQL Protocol and RDF Query Language

SPIN	SPARQL Inferencing Notation
SOAP	Simple Object Access Protocol
ST4RT	Semantic Transformations for Rail Transportation
SWRL	Semantic Web Rule Language
TAP TSI	Telematics Applications for Passenger service – Technical Specification for Interoperability
UIC	International Union of Railways
UITP	International Association of Public Transport
UML	Unified Modeling Language
WSDL	Web Service Description Language
XML	eXtensible Markup Language

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INITIAL ANALYSIS OF THE MATURITY OF THE SEMANTIC INTEROPERABILITY SOLUTIONS SUPPLY MARKET

This initial analysis of the available semantic interoperability solutions starts from the scouting of specificities of the data formats and structures that characterize current rail and transport data sources, ranging from those that have been born to be open (e.g., GTFS[1]) to those that are used internally and for exchange between systems. The analysis proceeds with the identification of the main challenges in providing integrated data to multimodal travel information services. Finally, the analysis focuses on tools and products to support semantic data integration with the aim to evaluate the availability and maturity of such technologies to be adopted in the rail and transport sector.

1.1 SCOUTING OF INTEROPERABILITY STANDARDS IN THE RAIL AND TRANSPORT SECTOR

The GoF4R project made a back-office analysis of the interoperability standards already existing in the rail and transport sector. The analysis intends to identify existing standard solutions, which can be reference cases for the communication needs in Shift2Rail Innovation Programme 4 (IP4) and can guide GoF4R upcoming subtasks and tasks. For such purpose, solutions which adopt semantic technologies have been highlighted.

In this section, we report a summary of the performed analysis. An extended report of the analysis is available in Annex I. Moreover, metadata descriptions of the analysed standards have been produced and are available in Annex II.

We notice that the common meaning of the word “standards” is often used for referring to a variety of documents which are shared for common applications by the stakeholders of a given sector. Since the word is also formally making reference to “standards” produced by legal official European and International bodies, namely CEN, CENELEC and ETSI at EU level (the ESOs, European Standardisation Organisations) and ISO, IEC and ITU at international level, it would be better to use the term “specifications” and to make a distinction between the three following categories:

- **Regulation documents (Specifications which are legal requirements and legally binding):** these *legal requirements for passenger travel* are resulting from Directives or Regulations produced by the European Union or the European Commission (in case of delegated acts), some of which are directly European railways related like the TAP TSI and PRM TSI. In the rail sector[2], the European Commission is supported by the European Union Agency for Railways (ERA[3]).

[1] https://developers.google.com/transit/gtfs/

[2] Urban rail – metro, tram, Light Rail as well as some suburban/regional rail lines – has been excluded from the scope of the ERA technical specifications for interoperability and safety. ERA’s responsibility covers the (mainline) interoperable European railway system.

[3] http://www.era.europa.eu/Pages/Home.aspx

- **Standardisation documents (Standards established by the European and International standardisation bodies):** the European and International standardisation bodies are producing a very big amount of specifications which are categorised as “Standards” with the prefix “EN” or “ISO” or “IEC” followed by a number, or as other documents such as Technical Report and Technical Specification. The production of these documents is following rather complex procedures implying National Standardisation Bodies (the “NSBs”) like AFNOR, in France, DIN in Germany etc., which produce as well their own national specifications (which have to comply with upper level specifications or standards from NSBs or ISO and IEC).

  These documents are produced (on their own budget) by volunteering experts from the relevant market stakeholders who are registered in Technical Committees or Working Groups managed by the standardisation bodies (for ESOs, usually on the basis of proposals made by NSBs). These standards are of voluntary use by the market actors, who have usually to pay for getting the standards (only a few of them are free). Some standards are produced as a complement of EU legislation, on request from the European Commission or from the Agency: they are called “harmonised” standards. They are also of voluntary use, except when they are referenced in EU legal text: in such a case they become mandatory.

  The “standardisation” as an official process is extensively described in GoF4R D3.1 “Interim Report on Regulatory Environment”.

- **Specifications documents (Specifications produced by non-State related international bodies or Platforms):** numerous companies (e.g., International Union of Railways (UIC) in the railway sector and IATA in the aviation sector) are grouping together at international level in organisations or associations or platforms to produce their own specifications which are of voluntary application. An example of the specifications produced by such bodies is the Full Service Model (FSM), produced by the mainline rail sector.

The back-office analysis made by the GoF4R project considered the three categories of “specifications” with reference to the four considered transportation modes: *railways*, *public transport*, *aviation* and *automotive*. Railways and public transport have been widely analysed, whereas aviation and automotive have been only partially investigated and their analysis will be extended in the second year of the GoF4R project. The objective of the back-office analysis is to check interoperability solutions in the transport domain.

Interoperability in transport is a wide and complex issue, which arises whenever there are interfaces between different parts of a system, which need to be compatible, in order for the whole system to work properly. This problem is even more complex in railways, due to the large number of subsystems and interfaces which the system comprises. However, in the context of the GoF4R project, focus is on the interoperability of data formats and structures, which is required in order to allow proper data sharing and exchange between different systems, subsystems or applications. An important goal is also to clarify when interoperable interfaces are based, at least to some extent, on semantic technologies.

The four analysed transportation modes present the following approaches to address interoperability by means of specifications:

- **Railways**. As interoperability is a vital issue in the European railway sector, European regulations have widely addressed the problem by means of the Interoperability Directive 2008/57 - which has been repealed with effect from 16 June 2016 by Directive 2016/797 - and of a number of related Regulations addressing more in detail the individual subsystems (e.g. Locomotives and Passenger Coaches, Freight Wagons, Energy, Control-Command and Signalling, Infrastructure), also known as Technical Specifications for Interoperability (TSI). In order to support such Regulations, a system of harmonised standards has been defined, which when applied, on a voluntary basis, give, within the limits of their scope, a presumption of conformity to the requirements in the related regulation documents. In some cases, specific standards or parts of them are referenced within the Regulation documents themselves, so becoming part of the regulation and therefore mandatory in Europe.
- **Public transport**. The public transport sector has already widely adopted the use of open interfaces and common data structures for providing and building passenger information systems. There are still many local and proprietary implementations, but those are all based on commonly known open data formats and structures. The trend seems to be that the cities have locally developed their own systems and are now gradually moving to unified nationwide systems.
- **Aviation and Automotive**. The aviation and automotive sectors have a different approach to standardisation as they mainly refer directly to central international standard organisations, where standardisation proposals prepared by industries are proposed and discussed. Basically, these sectors bypass the intermediation role of official standard organisations, by producing “de facto” standards.

From the back-end analysis of main standards in the railways and transport sector, it appears that a large range of solutions related to the interoperability of data formats and structures already exist. Solutions which include semantic technologies exist but seem to be not completely developed or focusing on one transport sector. In some cases, the interoperability solutions are supported by stakeholder organisations, which agree to cooperate in the creation and maintenance of common platforms. These are inputs for the analysis of currently existing examples of governance structures in the field.

In the railways sector interoperability solutions finds obstacles due to (i) the conservative attitude of main stakeholders; (ii) the complexity of the system, (iii) the quick and continuous evolution of the ICT technologies, and (iv) the investment cost of adopting new technologies when the existing ones have not ended their lifetime. Even if the situation is changing now, these can be reasons for not having yet established standards for semantic based interoperability in the railways sector.

In the public transport sector, a higher use of open interfaces and common data structures for providing and building passenger information systems has already been achieved. Similarly, sectors like automotive and aviation reached a high level of interoperability and in some cases even interchangeability between their subsystems.

It is to be noted that the different transportation modes largely differ from the point of view of the number of interfaces relevant for interoperability, their level of interaction and in general in their system architecture. However, all of them are anyway similarly involved in the objective of providing unified and seamless services to travellers. The heterogeneity of information systems and interfaces will require a common convergence towards semantic interoperability, as under development within Shift2Rail IP4.

As mentioned above, the GoF4R project focuses on the interoperability of data formats and structures in the railways and transport sectors and the performed back-office analysis reveals that the interoperability standards deal with different types of data characterized by different interoperability challenges:

- **Infrastructure data:** topological information on transportation networks (lines, routes, connections).
    - Interoperability challenges: the management of different data models often characterized by different granularities and frequent conflicts.

- **Static transportation data:** information on timetables, stations (including transfers and navigation paths), fares and electronic tickets.
    - Interoperability challenges: the management of different data models often characterized by different levels of Quality of Data (e.g., timeliness, accuracy); the management of sensible data.

- **Real-time transportation data:** real-time data on the status of infrastructures and vehicles.
    - Interoperability challenges: the management of different data models often characterized by different granularities and different level of Quality of Data; the management of sensible data.

One of the Shift2Rail IP4 challenges consists in the establishment of a framework which will enable the business applications belonging to a critical mass of European travel and transport industry players to 'interoperate' in order to support the development of **multimodal travel information services** that provide the customers with comprehensive information on the available transport options, along with the corresponding processes for their booking, payment, ticketing, consumption and modification. Basically, among others, Shift2Rail IP4 has to deal with infrastructure data, static transport data and real-time transportation data and has to tackle the interoperability challenges associated with the management of these data.

The EU-report on the provision of EU-wide multimodal travel information services [1] highlights that one of the main **barriers of comprehensive travel information services** in the EU is the lack of travel and traffic **data interoperability**. Moreover, the EU-report reveals that a key enabler to address this barrier is **making travel and traffic data interoperable with a common set of data exchange standards.** Basically, even if a large range of solutions related to the interoperability of data formats and structures already exist, they appear to be not enough to support multimodal travel information services. As an example, different specifications for static transportation data exist (e.g., GTFS [2], NeTEx [3]) but since different systems adopt different specifications, the data interoperability problem remains unresolved.

The GoF4R project clusters the above mentioned data interoperability challenges in the following two common data management activities:

- **Data Mapping**: mapping between data models characterized by different levels of granularities;
- **Data Integration**: integration of data specified according to different models and with different levels of Quality of Data (e.g., completeness, accuracy, timeliness and consistency).

Two additional coordinates make the mentioned activities harder:

- **The management of sensible data** (e.g., gender and PRM conditions of passengers);
- **The management and integration of real-time data**.

In the next sub-section, a case study is proposed to clarify the data mapping and integration activities and to verify if and how they can be solved using semantic technologies.

1.2 DEFINITION OF REQUIREMENTS FOR SEMANTIC DATA INTEGRATION IN THE RAIL AND TRANSPORT SECTOR

Usually, interoperability challenges are solved with syntactic interoperability solutions (i.e., two systems are syntactically interoperable if they are able to communicate between each other. A syntactic interoperability solution consists in creating ad hoc "adapters" that convert/translate/integrate the exchanged information between different data models).

The GoF4R project aims to understand how we can solve these interoperability challenges using semantic Web technologies. Two systems are semantically interoperable if they are interoperable and if their information exchange is based on the meaning of the exchanged information.

The following case study is provided to evaluate the feasibility of semantic interoperability solutions for data mapping and integration and to guide the definition of requirements for such solutions in the railways and transport sectors.

CASE STUDY: Semantic data integration and quality improvement to support multimodal transportation.

BRIEF DESCRIPTION: Multimodal transport is characterized by the use of more than one transportation mode within the scope of a single end-to-end journey. The IT2Rail Travel Companion acts, among others, as a multimodal journey planner promoting a more effective use of existing network capacity both within rail and in the wider multimodal transportation system. The IT2Rail Travel Companion must be supplied with data from transport operators concerning routes, stations, timetables, and fares. Such data are in various data formats (e.g., conformant to NeTEx or GTFS specifications) and, currently, the interoperability between them is only partially covered. Moreover, it can happen that the comparison of data from different sources may also lead to the detection of inconsistencies and, more in general, different levels of data quality.

CHALLENGE: Identify a semantic interoperability solution for multimodal journey planners enabling:

- batch integration of data concerning routes, stations, timetables, and fares defined in various data formats;
- data quality profiling and improvement.

Two different approaches to solve the case study are investigated in what follows:

- **semantic data warehousing**: all necessary data is collected in a central repository before a user query is issued;
- **on-demand retrieval**: data is collected from the integrated sources dynamically during query evaluation.

1.2.1 Semantic Data Warehousing

The EC has defined a lot of initiatives for developing Interoperability Solutions for European Public Administration (see Section 2.1.2), which are embedded in the European Interoperability Framework - EIF (see Section 2.1.1). The EIF proposes actions like the creation of "Base registries" where a base registry is a trusted and authoritative source of information which can and should be digitally reused by others, where one organisation is responsible and accountable for the collection, use, updating and preservation of information. Base registries are reliable sources of basic information on data items. Here, we aim to evaluate the feasibility of developing base registries as semantic data warehouses.

Examples of semantic data warehousing approaches to data integration and quality improvement are described in [4][5]. Starting from the analysis of these works, a possible solution for the design and implementation of a semantic data warehouse to support multimodal transportation is proposed and requirements in terms of required features and components are extracted.

As it emerges from the back-office analysis described in Section 1.1, in the rail and transport domain, data are stored in various distributed sources often characterized by heterogeneous data schemas. As an example, Google's General Transport Feed Specification (GTFS) and Network and Timetable Exchange (NeTEx) are widely used formats for distributing timetables to third parties. GTFS and NeTEx should be considered as complementary, since their data schemas have been conceived to cover different stages in the data management process. GTFS covers stops, lines, and timetabled journey information sufficient to answer basic journey planning queries. NeTEx covers many other aspects of Public Transport Information apart from timetables (e.g., network descriptions and interchanges) as well as supporting a richer timetable model that can include journey patterns, timing patterns, joined journeys, train makeup, connection timings, etc.

Due to the heterogeneity and complementarity of data in the rail and transport domain, the definition of data warehouses appears to be a proper solution to data integration and quality improvement. Data warehouse (DW) technology aims at materializing data and organizing them in order to facilitate their analysis. A DW can be seen as a **materialized data integration system**, where data are viewed in a multidimensional way [6] (i.e., data are organized into facts describing subjects of analysis. These data are analysed according to different dimensions). A data integration system offers a global schema representing a unified view of data sources. Formally, it may be defined by a triple: <G, S, M>, where G is the global schema, S is a set of local schemas that describe the structure of each source participating in the integration process, and M is a set of assertions relating elements of the global schema G with elements of the local schemas S. In the DW context, the integration aspect is performed through an ETL (Extract-Transform-Load) process, where data are extracted from sources, pre-processed and stored in the target DW.

A semantic data warehouse can be seen as a materialized data integration system where the mapping between source schemas and the global schemas as well as the ETL process are supported by the availability of ontologies and semantic web technologies. As an example, the availability of ontologies may allow defining semantic mappings between schemas of sources and the target DW schema. This leverages the integration process to the ontological level, and frees it from all implementation issues.

The proposed solution consists in a framework supporting the following tasks:

- source schema mappings;
- requirements analysis and conceptual data warehouse design;
- logical data warehouse design and ETL process definition;
- deployment and physical design.

In the following, a description of each task is provided and requirements are extracted.

Source schema mappings

Description: each data source is characterized by its local data schema. As a first step for bridging across the heterogeneities of diverse local sources, a **local ontology** must be generated from each source schema (e.g., relational, XML, or RDF). A local ontology is a conceptualization of the elements and relationships between elements in each source schema. When local ontologies are defined, **reasoning capabilities** are exploited for checking the consistency of each ontology (classes and instances) and inferring subsumption relationships. This is a first attempt to data quality improvement.

Moreover, in order to support data integration, a **global ontology** is needed. A global ontology is a conceptualization of the elements and relationships between elements that define the analysed domain. For what concerns the Transportation domain, the global ontology can be represented by, e.g., the IT2Rail ontology, the Rail Core Ontology (RaCoOn) [7] or the Public Transport Domain ontology [8].

The integration process starts with the definition of **mappings** between each local ontology and the global ontology. Different types of mappings (Equivalence, Containment or Overlap) can be defined between elements. The (semi-) automatic discovering of such mappings can be supported by techniques for ontology matching/alignment [9].

Requirements:

1. *Generation of local ontologies from heterogeneous source schemas (e.g., relational, XML, or RDF)*;

2. *Editing/visualization of ontologies*;
3. *Reasoning capability for checking consistency of an ontology and inferring subsumption relationships*;
4. *Support for ontology matching/alignment*.

Requirements analysis and conceptual data warehouse design

Description: A goal is an objective that the system under consideration should achieve. Identifying goals of users is a crucial task for DW development. Concerning our case study, an example of goal could be the retrieval of all the available transportation links between specified locations that enables the development of multimodal travel planning applications.

The requirement analysis aims at defining a connection between **users' goals** and the global ontology. The connection consists in the mapping between coordinates of each goal (e.g., transportation links and locations) and the elements in the global ontology. The connection can be established by means of e.g., ontology annotations. Knowing that the global ontology is linked to the data sources, these concepts chosen to express goals inform the designer about the most relevant data to store in the DW.

The conceptual data warehouse design aims at producing the **DW ontology** (DWO) defined from the global ontology (GO) by extracting all concepts and properties used by user goals. Three scenarios materialize this definition:

1. DWO = GO: the GO corresponds exactly to users' requirements,

2. DWO ⊂ GO: the DWO is extracted from the GO,

3. DWO ⊃ GO: the GO does not fulfil all users' requirements.

The designer may extend the DWO by adding new concepts and properties when the GO does not satisfy all users' requirements. These new concepts are annotated and are set by null values in the target warehouse. In case, the availability of additional sources that cover the new concepts and properties can be investigated.

When DWO is defined, reasoning capabilities are exploited again for checking the consistency of the ontology and inferring subsumption relationships. This is a second attempt to data quality improvement.

Since the DW conceptual model is characterized by a multidimensional structure where *fact* concepts are central concepts and *dimensions* are concepts influencing them, these roles of concepts and properties are stored as **ontological annotations**.

Requirements:

5. *Annotation of ontologies*;

Logical data warehouse design and ETL process definition

Description: The logical data warehouse design consists in transforming the DWO to a relational database model. The result is the production of the target DW schema. After that, the ETL process aims at populating the target DW schema with data extracted from the data sources. The ETL process must support several different operators to perform the extraction of data from data sources, its transformation and quality improvement and its final storage in the target DW. Examples of such operators are:

- RETRIEVE(S,C): extract instances of the class C from the source S;
- UNION(C,C'): unify instances whose corresponding classes C and C' belong to different sources;
- JOIN(C,C'): join instances whose corresponding classes C and C' are related by a property;
- DD(I): detect duplicate values from the incoming data
- CONVERT(C,C'): convert incoming data from the format of C to the format of C'.

These operators have to be leveraged to deal with the semantic aspects of sources. Basically, the ETL process depends on the semantics of mappings between the global ontology (GO) and each local ontology (LO). Four semantic mappings should be considered:

1. Equivalent (CGO ≡ C'LO): the class C' in the LO is equivalent to the class C in the GO. No transformation is needed. Instances are extracted from sources, merged, united or joined then loaded in the target data store.
2. Containment sound (CGO ⊃ C'LO): the class C' in the LO owns all the properties of the class C in the GO. No transformation is needed. Instances are extracted from sources, merged, united or joined then loaded in the target data store.
3. Containment complete (CGO ⊂ C'LO): the class C' in the LO owns only a subset of the properties of the class C in the GO. Some instances need to be transformed (converted, filtered and aggregated) then merged, unified or joined and finally loaded to the target data warehouse.
4. Overlap mappings: there is a partial overlap between properties of class C in GO and class C' in LO. In this case, there is the need to identify the constraints required by the GO class and not applied to the local ontology class.

Requirements:

6. *Transformation of ontologies (defined in e.g., OWL, RDF) to rational database models*;
7. *Extraction and transformation of instances from ontologies characterized by semantic mappings*;
8. *Support for the storage of ontologies (defined in e.g., OWL, RDF)*.

Deployment and physical design

Description: In a traditional DW, the deployment followed one-to-one rule, where each warehouse table is stored following one storage layout. In a semantic DW, the deployment may follow a one-to-many rule: the ontology model and the instances may have different storage layout.

Requirements:

9. *Support for the separate storage of ontology model and instances*;

1.2.2 On-demand Retrieval

The main purpose of an on-demand retrieval solution to data integration is to provide a semantically unified interface for querying (selected) heterogeneous information sources. On the contrary to a semantic data warehousing solution, an on-demand retrieval solution does not aim at merging all possible sources together providing a cumulated view of all attributes, but it utilizes mapping between ontologies to perform on-the-fly data integration.

As described in [9], ontologies have been extensively used in data integration systems because they provide an explicit and machine-understandable conceptualization of a domain. They have been used in one of the three following ways:

- **Single ontology approach**. All source schemas are directly related to a shared global ontology that provides a uniform interface to the user. However, this approach requires that all sources have nearly the same view on a domain, with the same level of granularity.
- **Multiple ontology approach**. Each data source is described by its own (local) ontology separately. Instead of using a common ontology, local ontologies are mapped to each other. For this purpose, an additional representation formalism is necessary for defining the inter-ontology mappings.
- **Hybrid ontology approach**. A combination of the two preceding approaches is used. First, a local ontology is built for each source schema, which, however, is not mapped to other local ontologies, but to a global shared ontology. New sources can be easily added with no need for modifying existing mappings.

Due to the heterogeneity of data schemas in the transportation domain and also due to the different level of granularity that characterizes such schemas, a single ontology approach appears to not be the best solution. In the following, a multiple ontology approach to a centralized data integration solution and a hybrid ontology approach to a peer-to-peer data integration solution will be analysed and requirements in terms of necessary features and components will be extracted.

A multiple ontology approach to a centralized data integration solution

Description: starting from the survey in [9] and from the work in [11], the following characteristics of a multiple ontology approach to a centralized data integration solution have been extracted:

- **A method for the generation of local ontologies from source schemas**: similarly to what happens for semantic data warehouses, when different heterogeneous data sources are available, a local ontology must be generated from each source schema.

- **A method for the definition of a reference global ontology**: the method supports the definition of a reference global ontology by merging a set of local ontologies. The process of ontology merging consists of several operations:

    - *Copying a class and/or its properties:* classes and properties that do not exist in the target ontology are copied into it.

    - *Class Merging:* conceptually equivalent classes in the local and target ontologies are combined into one class in the target ontology.

    - *Property Merging:* conceptually equivalent properties of a class in the local and target ontologies are combined into one property in the target ontology.

    - *Relationship Merging:* conceptually equivalent relationships from one class *c*1 to another class *c*2 in the local and target ontologies are combined into a single relationship in the target ontology.

    - *Class Generalization:* related classes in the local and target ontologies can be generalized into a superclass.

  We note that along with the above operations, semantic correspondences are established. For example, for each element *pL* in a local ontology, if there exists a semantically equivalent element *pG* in the global ontology, the two elements will be merged and a mapping between *pL* and *pG* is generated.

- **An approach to high-level queries re-writing and execution**: the approach supports the users in defining query in terms of the global ontology. The defined query is said to be a *high level* query since its formulation does not require awareness of particular source schemas. The query is then parsed, decomposed and reformulated by a rewriting algorithm into subqueries for specific source schemas.

- **A query engine and a federator**: the subqueries defined by the rewriting algorithm are orchestrated and executed by a query engine. Since data is not materialized, a federator is also needed. The results are merged and presented to the user.

Requirements:

10. *support for ontology merging (i.e., build a global ontology by merging a set of local ontologies)*;

11. *support for query parsing, decomposition and reformulation (i.e., from a single query defined in terms of the global ontology to multiple queries defined in terms of local ontologies)*;

12. *support for query orchestration and execution and result merging*.

A hybrid ontology approach to a peer-to-peer data integration solution

Description: starting from the survey in [9] and from the work in [12], the characteristics of a hybrid ontology approach to a peer-to-peer data integration solution have been analysed. Similar to the multiple ontology approach to a centralized data integration solution, this approach is characterized by a method for the generation of local ontologies from source schemas and by a method for the definition of a reference global ontology. Distinguishing characteristics are:

- **An architecture characterized by two types of peers**: a *hybrid* peer-to-peer architecture characterized by *super-peers* containing the global ontology, and *peers* each containing a data source and a local ontology. In our case study, each peer represents an autonomous information system managed by a transportation company and connects to a super-peer via semantic mappings.

- **An architecture characterized by distributed query endpoints**: the architecture should support the following query execution process:

  - the user poses a query (i.e., the *source query*) against the local schema or local ontology in a peer (i.e., the *source peer*). The query will be executed on the local source to get a *local answer*.

  - Meanwhile, the source query is rewritten into *target queries,* one for each connected peer (i.e., the *target peer*). The query rewriting algorithm utilizes mappings between the source peer to the super-peer and mappings between the super-peer to the target peer.

  - By executing the target queries, each target peer returns an answer (i.e., the *remote answer*) to the source peer. The local and remote answers are integrated and returned to the user.

Requirements:

13. *support for query translation (from local query to global query through semantic mapping);*

14. *support for integration of local and remote query answers.*

1.3 State-of-the-art of semantic interoperability tools supporting semantic data integration

This section provides the state-of-the-art of semantic interoperability tools in the market that support semantic data integration. After a selection process, each selected tool has been analysed in order to check if the functionalities it provides cover the 14 requirements for the semantic data integration solutions identified in Section 1.2. The performed analysis aims to evaluate the availability and maturity of semantic data integration tools on the market and the feasibility to adopt them in the rail and transport sector. The results are described in the following sub-sections.

1.3.1 Selection process and descriptions of the analysed tools

The methodology adopted for the selection of the tools consisted in asking to all the GoF4R-WP4 partners to include the semantic interoperability tools that they know or they have used. We believe that, in this way, thanks to the heterogeneity of the partners of the WP4 (research centres, institutions and industries) and their heterogeneous scopes w.r.t. the usage of semantic interoperability tools, we can obtain a wide vision of the market in this field. Following this process, we collected 19 tools that were known to address system and/or data interoperability using semantic technologies. Some of them are well-known tools in the field of semantic technologies (e.g., Virtuoso), while others are less known but were suggested by partners since they had been used for solving interoperability problems in the area of transport/smart cities (e.g., Sofia2). The set of selected tools are included in Table 1 where each tool is associated with the company or organisation that is in charge of providing and maintaining it and its category (Data Integration Solution, Triple Store, Reasoner, Web Service Architecture or Natural Language Processing (NLP) API).

After this selection step, we performed a preliminary analysis of the tools and we decided not to perform further analysis on some of them that do not cover the requirements identified in our semantic data integration use case (Section 1.2) or have not enough public documentation to make a proper analysis.

Company	Solution	Category	Selected
Denodo	Denodo Platform	Data Integration Solution	X
Capsenta	Ultrawap	Data Integration Solution	X
Oracle	Oracle Spatial and Graph	Data Integration Solution	X
Openlink	Virtuoso	Triple Store	X
Ontotext	GraphDB	Data Integration Solution	X
TopQuadrant	TopBraid	Triple Store	X
Open Source Community	Apache Marmota	Triple Store	X
Indra	Sofia2	Data Integration Solution[4]	X
Blazegraph	Blazegraph DB	Triple Store	
Franz Inc.	Allegrograph	Triple Store	X
Stardog Union	Stardog	Data Integration Solution	X
Metreeca	Meetreca Platform	Data Integration Solution	X
Alionscience	Alion Semantic Mediation Bus	Web Service Architecture	
SoftPlant	Living Semantic Platform	Data Integration Solution	X
VIStology Inc.	BaseVISor 2.0	Reasoner	
SpazioDati	Dandelion API	NLP API	
FuildOps	FluidOps Platform	Data Integration Solution	
Derivo	Knowledge Server	Reasoner	
ERTMS Solutions	ODASE	Data Integration Solution	X

[4] In this analysis, we set this category to Sofia2 but the tool provides many more features in other categories.

Table 1: Initial tools selection

Finally, we selected 13 tools from the *Data Integration Solution* and *Triple Store* categories. The reason for this selection is that the nature of these categories is well-adapted to the data-integration use case that has been defined in the Section 1.2. We define as *Triple Store* the tools where you can store and query data following semantic web technologies and as *Data Integration Solutions* the tools that extend features of a *Triple Store*, also allowing to take into account data in different formats, merge and consult them based on a shared vocabulary. We describe each selected tool with a link to its main web page:

- **Denodo Platform[5]:** The main goal of this tool is to provide solutions based on data virtualization for integrating heterogeneous data from multiple sources without replicate them. The semantic web technologies that it supports are SPARQL, RDF and D2R mappings. It is able to integrate data from sources like (NoSQL) databases or cloud applications and multiple formats like HTML, PDF, XML or CSV. It has a free/demo version that calls Denodo Express[6].
- **Ultrawap[7]:** This tool is a semantic interoperability solution focused on integrating heterogeneous data from multiple sources without the need of data warehousing. The semantic web technologies that it supports are OWL, SPARQL, RDF, R2RML and D2RQ. It supports the integration of data from SQL databases. It does not have a free version available.
- **Oracle Spatial and Graph[8]:** The main goal of this tool is to provide a RDF Semantic Graph in the Oracle ecosystem and to support views and queries over relational databases as RDF triples. The semantic web technologies that it supports are RDF, OWL and SPARQL. It supports the integration of data from Oracle databases. Free code and demonstrations are available[9].
- **Virtuoso[10]:** The main goal of this tool is to provide a simple and easy way to query Linked Data. It is one of the most popular triple store in the semantic web world. It supports RDF and SPARQL semantic web technologies. It is able to index 500M of triples per 16 GB of RAM. It has a full open version.
- **GraphDB[11]:** The main goal of this tool is to transform data from common formats and integrate them into a knowledge graph for publishing, sharing or finding connections between the data. It supports SPARQL and RDF semantic web technologies. It integrates data from tabular data like CSV, Excel or JSON. It provides a free version called GraphDB Free.
- **TopBraid[12]:** The main goal of this tool is to provide an enterprise platform for building services and applications to operate with Linked Data. The semantic web technologies that it supports

[5] https://www.denodo.com/en/denodo-platform/overview
[6] https://www.denodo.com/en/denodo-platform/denodo-express
[7] https://capsenta.com/#section-ultrawrap
[8] http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/index.html
[9] http://www.oracle.com/technetwork/database/options/spatialandgraph/downloads/index.html
[10] https://virtuoso.openlinksw.com/rdf/
[11] http://ontotext.com/products/graphdb/
[12] https://www.topquadrant.com/tools/IDE-topbraid-composer-maestro-edition/

are SPARQL, RDF, SPIN, SHACL and OWL. It has some data source adaptors for importing data from XML. It does not have a free version available.

- **Apache Marmota[13]:** The main goal of the tool is to provide an open platform for Linked Data that can be used, extended and deployed by organizations. It supports RDF and SPARQL semantic web technologies. The tool is free for use.

- **Sofia2[14]:** The tool offers a semantic platform to integrate data from multiple systems and devices, allowing interoperability between them. It supports mappings from multiple data sources to a common vocabulary (ontology). It also provides some free demos[15].

- **Allegrograph[16]:** This tool is very similar to Blazegraph DB since it is focused on provide high performance in the query processing. The semantic web technologies that it supports are SPARQL, RDF and SPIN. It is able to store more than 1000 trillion of triples. It does not provide any demo or free version.

- **Stardog[17]:** This tool is focused to allow its clients to query, search and analyse distributed data using Knowledge Graph technologies. The semantic web technologies that it supports are RDF, SPARQL and OWL 2. It is able to integrate data from Documents, RDBMS, LDAP, XML or emails without the needed of a transformation. It has a free 30-day trial of its enterprise version and a Stardog Community version for a non-commercial use.

- **Metreeca Platform[18]:** The main goal of this tool is to provide an open-data platform supporting linked data and W3C standards delivering easy read/write REST APIs to integrate heterogeneous sources. It supports some semantic web technologies like SPARQL, SHACL or JSON-LD REST APIs. An online demo is available[19] and the platform will be released soon as an open source project.

- **Living Semantic Platform[20]:** This solution provides a platform with REST services for querying linked data and integrate multiple data sources. It supports SPARQL and OWL 2 QL semantic web technologies. It is also able to integrate SQL databases and other data sources in formats like Excel, PDF or XML. It has a demo online[21].

- **ODASE[22]:** This tool aims at providing a complete solution in order to model both the knowledge domain and the business logic of the final application, therefore minimising the need for

[13] http://marmotta.apache.org/
[14] http://sofia2.com/
[15] http://sofia2.com/demostradores_en.html
[16] https://allegrograph.com/allegrograph/
[17] http://www.stardog.com/
[18] https://www.metreeca.com/
[19] https://demo.metreeca.com/
[20] http://www.softplant.de/en/portfolio-item/living-semantic-platform-2/
[21] https://demo.softplant.de/LivingSemanticPlatform/
[22] http://www.odaseontologies.com/

additional coding. To do this, it applies rules defined using the SWRL language. In this way, it seems able to cover almost all the identified requirements, with some of them requiring deeper analysis. It is a commercial solution, proven in real life full-scale applications.

1.3.2 Analysis of the selected tools

Each selected tool has been analysed in order to check if the functionalities it provides cover the 14 requirements for the semantic data integration solutions identified in Section 1.2. Table 2 reports the results of the analysis. For each requirement, the table shows if the tool fully supports it (the cell is filled in green), partially supports it (the cell is filled in yellow) or does not support it (the cell is filled in red). In case in which the available documentation was not enough to perform an analysis, the text “*not found in the documentation*” is specified.

From the analysis, the following results emerge:

- No tool covers the entire set of requirements for a semantic data integration solution in the transportation domain. A combination of tools appears to be required.
- The requirements that results to be less covered (cells filled in red or yellow) or not specified in the documentation of the tools are #4 “Support for ontology matching/alignment”, #5 “Annotation of ontology” and #10 “Support for ontology merging”. The lack of coverage of requirement #5 can be solved using an external ontology editor to make annotations. The lack of coverage of requirements #4 and #10 could be solved using a different category of tools more focused on ontology mapping and alignment. An example of tool of this category is AgreementMaker[23].

However, the analysis is performed using only the documentation released by tools providers on public websites. As mentioned above, very often this documentation is not enough to evaluate the coverage of specific requirements. For this reason, an extended analysis which directly involves the tools providers is needed.

The extended analysis will be done during the second year of the project. It will involve the providers or developers of the tool, which will be asked to answer a questionnaire and possibly participate in an interview.

[23] https://github.com/agreementmaker/agreementmaker

Table 2: Coverage of requirements for semantic data integration offered by tools on the market.

REQUIREMENT\TOOL	Denodo Platform	Ultrawap	Oracle Spatial and Graph	Virtuoso	GraphDB	TopBraid	Apache Marmotta	Sofia2	Allegrograph	Stardog	Metreeca	Living Semantic Platform	ODASE
#1. Generation of local ontologies from heterogeneous source schemas	supported	supported	not supported	supported	not supported	supported	not supported	supported	supported	Not found in the documentation	not supported yet. Planned.	supported	supported
#2. Editing/visualization of ontologies	supported	supported	supported	Partially. It supports tabular visualization of triples.	supported	Partially. It supports tabular visualization of triples.	Partially. It supports tabular visualization of triples.	supported	not supported	supported	Partially. It supports linked-data navigation and editing/ visualization of SHACL-based specs.	supported	supported
#3. Reasoning capability for checking consistency of an ontology and inferring subsumption relationships	Not found in the documentation	Not found in the documentation	supported	supported	supported	supported	supported	supported	supported	supported	not supported.	supported	supported
#4. Support for ontology matching/alignment	Not found in the documentation	supported	not supported	not supported	not supported	not supported	Not found in the documentation	Not found in the documentation	Not found in the documentation	not supported	not supported yet. Planned.	Not found in the documentation	Not clear from the documentation
#5. Annotation of ontologies	Not found in the documentation	not supported	not supported	not supported	not supported	supported	not supported	Not found in the documentation	Not found in the documentation	Not found in the documentation	not supported.	Not found in the documentation	supported
#6. Transformation of ontologies to relational database models	supported	not supported	supported	supported	not supported	Not found in the documentation	supported	not supported	not supported	Not found in the documentation	not supported.	not supported	supported
#7. Extraction and transformation of instances from ontologies characterized by semantic mappings	supported	supported	supported	not supported	not supported	supported	not supported	Not found in the documentation	Not found in the documentation	Not found in the documentation	not supported.	not supported	supported
#8. Support for the storage of ontologies	supported	not supported	supported	supported	supported	supported	supported	supported	supported	supported	not supported.	supported	supported
#9. Support for the separate storage of ontology model and instances	not supported	supported	not supported	not supported	not supported	supported	supported	supported	supported	supported	not supported.	Not found in the documentation	supported
#10. Support for ontology merging	Not found in the documentation	Not found in the documentation	not supported	not supported.	not supported.	Not found in the documentation	Not found in the documentation	supported	Not found in the documentation	not supported	not supported.	supported	Not clear from the documentation
#11. Support for query parsing, decomposition and reformulation	Not found in the documentation	supported	supported	Partially. Limited to inferences, no data transformation.	Partially. Limited to inferences, no data transformation.	Not found in the documentation	Not found in the documentation	not supported.	supported federation	supported.	not supported.	Not found in the documentation	Not clear from the documentation
#12. Support for query orchestration and execution and result merging	supported via SPARQL	supported	supported via SPARQL and Java APIs	supported via SPARQL	supported via SPARQL	supported via SPARQL	supported via SPARQL	not supported.	supported via SPARQL	supported	not supported.	supported via SPARQL	supported
#13. Support for query translation	supported via SPARQL	not supported.	supported via SPARQL and Java APIs	supported via SPARQL	supported via SPARQL	supported via SPARQL	supported via SPARQL	not supported.	supported via SPARQL	supported	not supported.	supported via SPARQL	supported
#14. Support for integration of local and remote query answers	supported via SPARQL	supported	supported via SPARQL and Java APIs	supported via SPARQL	supported via SPARQL	supported via SPARQL	supported via SPARQL	not supported.	supported via SPARQL	Not found in the documentation	not supported.	supported via SPARQL	supported

2. INITIAL ANALYSIS OF THE ADOPTION OF SEMANTIC INTEROPERABILITY SOLUTIONS IN INDUSTRY AND GOVERNMENT

In recent years the amount of data has increased in a lot of different fields and domains. Extracting knowledge from such data is often a difficult task because we have to deal with problems such as data heterogeneity and distribution. Semantic interoperability aims at solving these issues by applying Semantic Web technologies with the main goal of creating common and shared vocabularies to represent data from a domain.

In this chapter, we make an initial analysis of the adoption of semantic interoperability solutions in industry (with a special focus on transport) and government. We divide the analysis in three different parts. In the first one we describe the main relevant semantic interoperability solutions in two different domains: transport and public administration. We also define the meaning of (semantic) interoperability and we describe some past relevant projects and initiatives in the fields. In the second part, we analyse the results from a survey on semantic interoperability that we have delivered to relevant stakeholders in the domain of transport. Finally, in the last part, we describe the feedback received from transport companies and institutions on the 1st GoF4R Workshop in Semantic Interoperability, co-located with the 9th IT Combined Meeting at Cologne on 19 September 2017.

2.1 STATE-OF-THE-ART IN SEMANTIC INTEROPERABILITY IN TRANSPORT AND PUBLIC ADMINISTRATION SECTORS

Proposing semantic interoperability solutions is one of the most important issues that the European Union has focused on these last years. The main goal of semantic interoperability is that software, systems or datasets would be able to exchange data using multiple common vocabularies[24] based on common ontologies.

This section presents an analysis of some of the most relevant solutions in the area of semantic interoperability, with a special focus on the transport and public administration domains. It also provides a summary of what is interoperability, a set of relevant project that have had semantic interoperability in their central axis and an analysis of why this field becomes relevant in a domain like transportation.

2.1.1 Interoperability and the European Interoperability Framework (EIF)

Interoperability means the ability of information and communication technology (ICT) systems and of the business processes they support to exchange data and to enable the sharing of information and knowledge. Two systems are interoperable if they are able to communicate to each other through a complete information exchange[25].

The basic version of interoperability between digital systems is often referred to as "syntactic interoperability", to distinguish it from semantic interoperability. Syntactic Interoperability is usually

[24] http://lod-cloud.net/

[25] http://ec.europa.eu/idabc/en/document/5313/5883.html

solved by creating ad hoc "adapters" that convert/translate the exchanged information between different formats; this however, is not a scalable solution since, if there are N non-interoperable systems, the required number of adapters is $O(N^2)$; the European Interoperability Framework (EIF) refers to this as "bilateral solutions".

According to the EU Interoperability Framework (EIF), described below:

“Semantic interoperability ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words ‘what is sent is what is understood’. In the EIF, semantic interoperability covers both semantic and syntactic aspects:

- The semantic aspect refers to the meaning of data elements and the relationship between them. It includes developing vocabularies and schemata to describe data exchanges, and ensures that data elements are understood in the same way by all communicating parties;
- The syntactic aspect refers to describing the exact format of the information to be exchanged in terms of grammar and format.

A starting point for improving semantic interoperability is to perceive data and information as a valuable public asset.”

It is often stated that semantics is able to turn data into information. Moreover, the elaboration capability made possible by semantics (reasoning), allows to infer new hidden information from the existing one, so turning information into knowledge.

European Interoperability framework

What is presented in this section is in a very large part extracted from the recent EC communication COM (2017)134 final[26] adopted on 23 March 2017.

The European interoperability framework is a commonly agreed approach to the delivery of European public services in an interoperable manner. It defines basic interoperability guidelines in the form of common principles, models and recommendations.

In June 2002, European heads of state adopted the eEurope Action Plan 2005 at the Seville summit. It calls on the European Commission “to issue an agreed interoperability framework to support the delivery of pan-European eGovernment services to citizens and enterprises”. This framework addresses information content and recommends technical policies and specifications to help connect public administration information systems across the EU.

A European Interoperability Framework has been adopted in 2010 providing 25 recommendations and guidelines for eGovernment services so that public administrations, enterprises and citizens can interact across borders, in a pan-European context. As the field of information technology is developing by fast speed and new EU policies have emerged, and in order to better react on emerging technological trends like open data and cloud computing a need for a new framework was

[26] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: European Interoperability Framework – Implementation Strategy. 23 March 2017.

confirmed during a stakeholder consultation from April to June 2016. It targeted citizens, businesses, research centres, academic institutions, public administrations, as well as standardisation organisations and businesses supplying services to public administrations.

A new European Interoperability Framework (EIF) has been recently produced[27] as part of the EC Communication (COM(2017)134). The framework gives specific guidance on how to set up interoperable digital public services. It offers public administrations 47 concrete recommendations on how to improve governance of their interoperability activities, establish cross-organisational relationships, streamline processes supporting end-to-end digital services, and ensure that both existing and new legislation do not compromise interoperability efforts. This new framework puts more emphasis on how interoperability principles and models should apply in practice. The updated interoperability recommendations have been made more specific to facilitate their implementation, with a stronger focus on openness and information management, data portability, interoperability governance, and integrated service delivery.

Principles

The EIF is accompanied by the Interoperability Action Plan, which outlines priorities that should support the implementation of the EIF from 2016 to 2020. This plan is comprised of five focus areas as follows:

- Ensure governance, coordination and sharing of interoperability initiatives.
- Develop organisational interoperability solutions.
- Engage stakeholders and raise awareness on interoperability.
- Develop, maintain and promote key interoperability enablers.
- Develop, maintain and promote supporting instruments for interoperability.

Accordingly, the considerations and recommendations of the EIF are based on 12 underlying principles (see Figure 1) grouped into four categories:

1. Principle setting the context for EU actions on interoperability (No. 1).
2. Core interoperability principles (Nos. 2 to 5).
3. Principles related to generic user needs and expectations (Nos. 6 to 9).
4. Foundation principles for cooperation among public administrations (Nos. 10 to 12).

[27] https://ec.europa.eu/isa2/eif_en

1: Subsidiarity and proportionality

2: Openness

3: Transparency

4: Reusability

5: Technological neutrality and data portability

6: User-centricity

7: Inclusion and accessibility

8: Security and privacy

9: Multilingualism

10: Administrative simplification

11: Preservation of information

12: Assessment of Effectiveness and Efficiency

Figure 1: European Interoperability Framework principles[28]

Types of Interoperability

The new EIF defines four different types of interoperability:

a) **Legal interoperability:** Legal interoperability is about ensuring that organisations operating under different legal frameworks, policies and strategies are able to work together. This might require that legislation does not block the establishment of European public services within and between Member States and that there are clear agreements about how to deal with differences in legislation across borders, including the option of putting in place new legislation.

b) **Organisational interoperability:** This layer refers to the way in which public administrations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals. In practice, organisational interoperability means documenting and integrating or aligning business processes and relevant information exchanged. Organisational interoperability also aims to meet the requirements of the user community by making services available, easily identifiable, accessible and user-focused thanks to Business process alignment and organisational relationship.

c) **Semantic interoperability:** Agreements on reference data, in the form of taxonomies, controlled vocabularies, thesauri, code lists[29] and reusable data structures/models[30] are key prerequisites for achieving semantic interoperability. Approaches like data-driven-design, coupled with linked data technologies, are innovative ways of substantially improving semantic interoperability.

[28] EIF - Implementation Strategy - COM(2017)134 - Annex II, Figure 2, page 8.

[29] For example the Eurovoc thesaurus and the European skills, competence and occupations (ESCO) taxonomy.

[30] The core person, core business, core location and core public service developed by the ISA Programme are examples of cross-sector, reusable data models.

Given the different linguistic, cultural, legal, and administrative environments in the Member States, this interoperability layer poses significant challenges. However, unless standardisation efforts mature in the semantic interoperability layer, it is difficult to ensure seamless information exchange, free movement of data, and data portability among Member States to support a digital single market in the EU.

d) **Technical interoperability:** This aspect of interoperability covers the technical issues - applications and infrastructures - of linking computer systems and services. It includes key aspects such as open interfaces specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

The EIF conceptual model for integrated public services provision

The EIF proposes a "conceptual model for integrated public services" to guide their planning, development, operation and maintenance by Member States. It is relevant to all governmental levels, from local to EU. The model is modular and comprises loosely coupled service components[31] interconnected through shared infrastructure. The model promotes the idea of interoperability by design. It means that for European public services to be interoperable, they should be designed in accordance with the proposed model and with certain interoperability and reusability requirements[32] in mind. The model promotes reusability as a driver for interoperability, recognising that the European public services should reuse information and services that already exist and may be available from various sources inside or beyond the organisational boundaries of public administrations. Information and services should be retrievable and be made available in interoperable formats. The components of the conceptual model are presented in the Figure 2 and the most relevant in terms of semantic interoperability are:

- **Base registries:** A base registry is a trusted and authoritative source of information which can and should be digitally reused by others, where one organisation is responsible and accountable for the collection, use, updating and preservation of information. Base registries are reliable sources of basic information and constitute the 'master data' for public administrations and European public service delivery. 'Authoritative' here means that a base registry is considered to be the 'source' of information, i.e. it shows the correct status, is up-to-date and is of the highest possible quality and integrity. An example of a base registry is the one on travel and traffic data to be produced by the National Access Point – NAP - created on 31st May 2017 by the Commission Delegated Regulation C (2017) 3574 final[33] supplementing the ITS Directive 2010/40/EU.

- **Open Data:** The Directive on the reuse of public sector information[34] provides a common legal framework for reuse of public sector data. The focus is on releasing machine-readable

[31] Service oriented architecture (SOA) is an implementation of that concept.

[32] The interoperability maturity model (IMM) developed in the context of the ISA programme can be used to assess a service's readiness for interoperability.

[33] Commission Delegated Regulation C (2017) 3574 final supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services.

[34] The directive on the re-use of public sector information (Directive 2003/98/EC, known as the 'PSI Directive') was revised by Directive 2013/37/EU which entered into force on 17 July 2013. In 2017, fulfilling the periodic review obligation prescribed by the Directive, the EC has launched a public online consultation closed on 13/12/2017.

data for use by others to stimulate transparency, fair competition, innovation and a data-driven economy. To ensure a level playing field, the opening and reuse of data must be non-discriminatory, meaning that data must be interoperable so that can be found, discovered and processed. Ideally basic metadata[35] and the semantics of open datasets should be described in a standard format readable by machines.

- **Catalogues:** Catalogues help others to find reusable resources (e.g. services, data, software, data models). Various types of catalogue exist, e.g. directories of services, libraries of software components, open data portals, registries of base registries, metadata catalogues, catalogues of standards, specifications and guidelines. Commonly agreed descriptions of the services, data, registries and interoperable solutions published in catalogues are needed to enable interoperability between catalogues[36]. A specific kind of catalogue is the European interoperability cartography (EIC)[37] of interoperability solutions available for reuse and sharing.

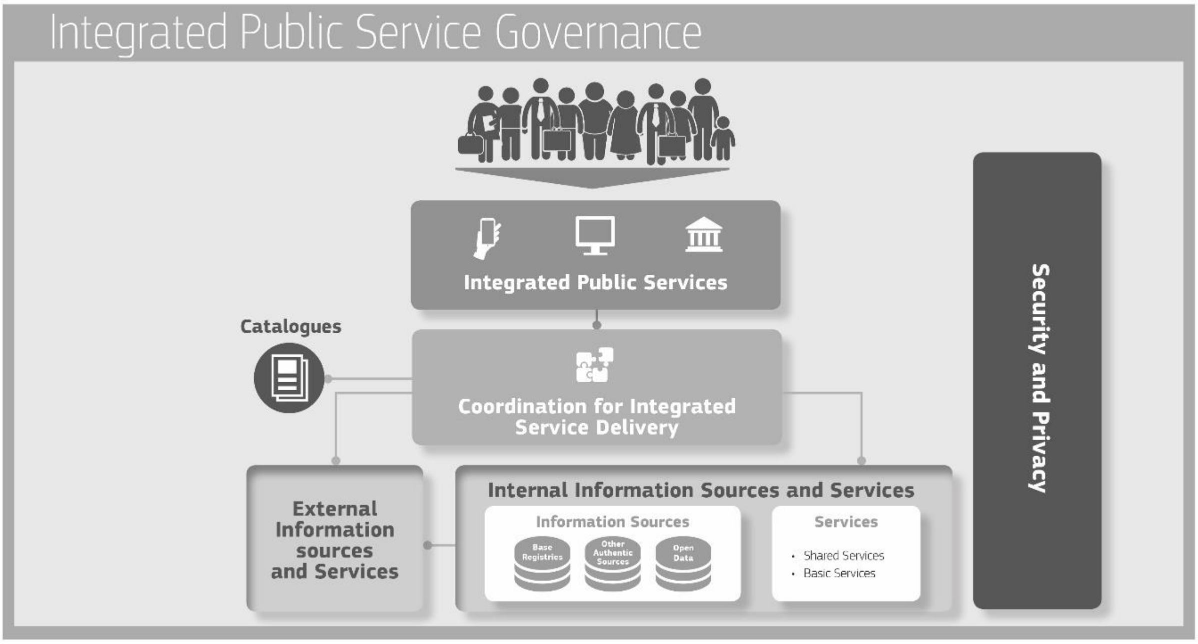

Figure 2: EIF conceptual model for integrated public services[38]

2.1.2 Recent projects in the area of semantic interoperability

In this section, we describe some of the most relevant and recent projects that have been done in the area of semantic interoperability, focused on projects in public administration and transport domains.

[35] For example those included in the DCAT-AP specification developed in the context of the ISA programme.
[36] The DCAT-AP, the Core Public Service Vocabulary and the Asset Description Metadata Schema are examples of specifications used to describe open data, public services and interoperability solutions respectively.

[37] https://ec.europa.eu/isa2/solutions/eira_en

[38] EIF - Implementation Strategy - COM(2017)134 - Annex II, Figure 4, page 29.

R&I projects in the Public Administration domain

The European Commission has made big efforts to promote semantic interoperability among organizations by, for example, funding EU projects, creating the EU interoperability framework or developing a generic data portal for storing datasets. That is why GoF4R intends to be focused on three relevant cases related to public administration and promoted by the European Union: DCAT-AP, ADMS, and CPSV-AP.

In addition, some other initiatives have also been carried out by the European Union. One of the most relevant has been a course focused on teaching and creating awareness about the usability and usefulness of semantic interoperability in organizations, companies and public bodies. These courses are divided into a set of modules that were given by the ISA[39] program (Interoperability Solutions for European Public Administrations).

- **DCAT-AP: Data Catalog Vocabulary – Application Profile**

  Metadata are data that describe other data, needed to be able to discover datasets within a certain domain. Simplifying, metadata describe the available information about or within a dataset(s). For example, in the transport domain, the metadata files may contain general information like the name of the distribution file or the company web page to whom data belongs. They may also contain more specific information about public transport, like the geographic area covered by a specific service.

  On January 2014, the DCAT[40] W3C recommendation for metadata generation was published. DCAT is a RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. Using DCAT, publishers increase discoverability and applications easily consume metadata from multiple catalogues. Besides, it facilitates federated queries across sites. In Spain, for example, this recommendation is the basis for the Interoperability Technical Norm[41] (national recommendation to publish metadata), followed by most public administrations, and especially those that want to be federated at http://datos.gob.es.

  DCAT Application Profile (DCAT-AP), an extension of DCAT, has been developed by a joint initiative of DG CONNECT, DG DIGIT and Publications Office of the EU. This vocabulary includes classes and properties that the initial vocabulary had not taken into account. DCAT-AP is an ontology that contains mandatory, recommended and optional classes and properties, so it's easy to adapt it to different domains. The European Data Portal[42] implements the DCAT-AP as the common vocabulary for harmonising descriptions of over 258,000 datasets harvested from 67 data portals of 34 countries.

  The extension is created to find a solution to the problems with the reuse and localization of open data at European level. The DCAT-AP[43] team identified two different scenarios, one for data re-users, who find it sometimes difficult to get an overview of which datasets exist and which public

[39] https://ec.europa.eu/isa2/home_en

[40] https://www.w3.org/TR/vocab-dcat/

[41] https://www.boe.es/boe/dias/2013/03/04/pdfs/BOE-A-2013-2380.pdf

[42] https://www.europeandataportal.eu/

[43] https://joinup.ec.europa.eu/asset/dcat_application_profile/description

administrations are maintaining and another for data providers, who want to encourage reuse of their datasets by making them searchable and accessible. DCAT-AP provides an easy way to develop profiles for a specific domain. This approach has been followed by other profiles, such as GeoDCAT-AP[44] or StatDCAT-AP[45], which are now standards in the European Data Portal for representing geospatial and statistical metadata[46].

GeoDCAT-AP provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 and those defined in the framework of the INSPIRE Directive, which are focused on describing geospatial datasets, dataset series, and services. The idea of this extension is to provide a way of transforming from these standards, which are widely used, to DCAT-AP. The main goal is to offer to owners of geospatial data the possibility of representing this information in RDF. Note that this extension does not make any changes, nor does it add anything new to the DCAT-AP standard model, as it is only a transformation guideline.

The main goal of StatDCAT-AP is to enhance interoperability between the metadata of statistical datasets within the statistical domain and between statistical data and open data portal.

- **ADMS: Asset Description Metadata Schema**

  ADMS[47], the Asset Description Metadata Schema, is a profile of DCAT for describing so-called Semantic Assets (or just 'Assets'), that is, highly reusable metadata (e.g. xml schemata, generic data models) and reference data (e.g. code lists, taxonomies, dictionaries, vocabularies) that are used for eGovernment system development. Someone searching for an asset is likely to have different needs, priorities and expectations from someone searching for a dataset in a data catalogue and these differences are reflected in ADMS. In particular, users seeking an Asset are likely to be searching for a document — something they can open and read using familiar desktop software, as opposed to something that needs to be processed. Of course, this is a very broad generalization. If a code list is published as a SKOS Concept scheme then it is both an Asset and a dataset and it can be argued that all Assets are datasets. Therefore, the difference in user expectation is at the heart of what distinguishes ADMS as a profile of DCAT. A further distinction between DCAT and ADMS can be made in that DCAT is designed to facilitate interoperability between data catalogues, i.e. the catalogue itself is at the heart of the vocabulary. ADMS is focused on the assets within a catalogue.

- **CPSV-AP: Core public Service Vocabulary – Application Profile**

  The Core Public Service Vocabulary[48] is a simplified, reusable and extensible data model that captures the fundamental characteristics of a service offered by public administrations. It was created by the ISA program as a baseline vocabulary for European public administrations that manage public services. CPSV does not provide a strict standard or standard format for

[44] https://joinup.ec.europa.eu/node/154143/
[45] https://joinup.ec.europa.eu/node/147940
[46] An application profile for the transport domain has been developed, the TransporDCAT-AP, which is presented in the Section 2.1.3
[47] https://www.w3.org/TR/vocab-adms/
[48] https://joinup.ec.europa.eu/solution/core-public-service-vocabulary

governments to adhere to. Instead, it can be perceived as a set of “terms” (Concepts) out of which administrations can pick and choose when describing their services.

CPSV-AP[49] is an extension of CPSV that provides a schema in which to connect the terms together in the scope of a specific application. The latest specification of this vocabulary mentions three concrete use cases for which it may be used: (i) Finding information about public services more easily, (ii) Building user-centric catalogues of public services at all levels from regional to a European federated catalogue, (iii) Managing portfolios of public services. To realise these use case, CPSV-AP provides a schema which relates the following classes: A *Public Service*, an *Event*, an *Output*, a *Rule*, a *Criterion Requirement* and a *Channel*.

The CPSV-AP Schema is available as Linked Open Data in many formats through the Joinup portal and some examples following the schema are available[50,51]. It is important to remark that CPSV-AP is just a schema, and it is of little use if any administration adopts it. So every new version of CPSV-AP remains simple but flexible so that even the smaller and less technologically apt administrations will able to apply it.

R&I projects in the transport domain

Semantic technologies are quickly spreading and are used in many sectors. Here quick summaries of experiences coming from projects in the sector of (railway) transportation are presented, where semantic technologies were included in the work programme and represented an important aspect of the investigated solutions. They give an overview on research about interoperability aspects related to data exchange in the last 12 years, presenting the past, present and future work in the field. Other relevant projects for multi-modal transportation, such as ALL WAYS TRAVELLING[52], IFG-Project[53] and the Full Service Model Initiative[54], are not analyzed in this section since they do not provide semantic interoperability solutions.

The current projects in the transport domain are connected via the idea of attractive transport ecosystem. In particularly, the idea is that the European transportation ecosystem must meet travelers needs to support anytime, anywhere, door-to-door, multimodal journeys encompassing all modes of transportation within a Single European Transport Area.

It requires interoperability between the business applications that need to dialogue with each other for that purpose, amongst the relevant transport supply-chain entities, regardless of their geographical situation (country, regions, city), the type of mobility service (journey planning, shopping facility, booking engine, ticketing engine etc.) or transport mode, and associated format(s)/protocol(s) adopted for such dialogues by the existing legacy systems[55].

[49] https://joinup.ec.europa.eu/solution/core-public-service-vocabulary-application-profile

[50] https://stad.gent/data/products/

[51] https://github.com/opencitydata/sector-publico-servicio

[52] http://www.allwaystravelling.eu/

[53] http://www.smart-ticketing.org/category/publications/ifm_project/

[54] https://tsga.eu/fsm

[55] http://projects.shift2rail.org/s2r_ip4_n.aspx?p=CO-ACTIVE

- **INTEGRAIL: Intelligent Integration of Railway Systems**

  The aim of InteGRail[56] (FP5 – 2005-2009) was to create a holistic, coherent information system, integrating the major railway sub-systems[57], in order to achieve higher levels of performance of the railway system in terms of capacity, average speed and punctuality, safety and the optimized usage of resources.

  Building on results achieved by previous projects, InteGRail proposed new intelligent procedures and contributed to the definition of a new approach in dealing with information exchange and data sharing in the mainline railway sector. In fact, in order to increase the performance and efficiency of existing systems, the project envisaged solutions based on a railway Ontology which allowed to convert dispersed and incompatible data into a common language and therefore to enable the possibility of new collaborating services. The railway Core Ontology was developed and then extended in different directions in order to allow the setup of specific demonstrators in railway domains like monitoring and diagnostics, traffic management and maintenance.

  The development of a complete railway Ontology was beyond the scope of the project; InteGRail developed a Core Ontology which was then extended as much as needed in order to support the envisaged demonstrators. Such Ontology was shared between a number of application servers distributed in Europe, creating a distributed peer-to-peer network within the Internet (InteGRail grid). However, sometime after the end of the project the servers were deactivated.

  The project could show that the holistic approach was needed, as otherwise optimising a subsystem could have a negative effect on another subsystem, and practicable, as the demonstrators were showing the possibility of real-time collaboration between a number of different Web services.

  The basic role of semantic technology in achieving such result was also clearly showed, in order to achieve general, low-cost and flexible solutions for information exchange, applicable (potentially) to any existing information system. As a matter of fact, servers using different technologies (Windows, Linux) were part of the above-mentioned grid.

  Despite a huge dissemination effort, the project however could not overcome the barriers present in the sector at that time, due to lack of widespread background (especially within decision-makers) and technology maturity. As a consequence, direct exploitation of results was more limited than expected and a few years were needed for the research to have a continuation.

  InteGRail however paved the way to the introduction of semantic technologies in the railway sector, also defining new concepts which are now commonly accepted. This came from a logical evolution of technology, following research done in a sequence of previous projects (ROSIN, TrainCom, EuRoMain), where the background for such evolution was defined.

[56] www.integrail.info

[57] The term “sub-system” is here used with the meaning defined in the Directive 2008/57 related to the interoperability of the rail system within the Community, e.g.; rolling stock, infrastructure, energy, traffic operation and management, etc. (Directive 2008/57/EC has been repealed by Directive (EU) 2016/797 with effect by 16 June 2020 with no significant change in the definition of sub-systems).

- **ON-TIME: Optimal Networks for Train Integration Management across Europe**

  The aim of ON-TIME[58] (FP6 – 2011-2015) was a step-change in railway capacity by reducing delays and improving traffic fluidity. The project proposed a distributed architecture to integrate different algorithms to solve typical problems in railway operations, such as timetable planning, dynamic re-planning of services at the macroeconomic and microeconomic levels following disturbance, resource management and rostering. The key-purpose of the architecture was to define a distributed, configurable and flexible infrastructure to exchange data and messages between different modules. The advantage of using a distributed architecture in this context is the ability to collect and exchange data on systems that are by their own nature loosely coupled and create a coherent, dynamic communication context in which such information can be exchanged.

  One of the main results of the project was the development of the ON-TIME Real-Time Traffic Plan (RTTP), which aims to provide information about predicted train sequences on a selected network part. Train routes and position and type of stops with their arriving and departure times are provided as well. The ON-TIME RTTP data resulted from processing dynamic traffic data and static traffic data (in railML) made by a Perturbation Management Module dealing with traffic state prediction and track/connection conflict detection and resolution.

  From available documents, there is no evidence of using semantic technologies. However, a semantic approach is considered, as the architecture is based on a data model (then described using RailML, with a suitable extension) and its graphical representation looks very much like an Ontology.

  The project also provided inputs for standardisation, as they state that this is necessary “in order to easily integrate a collection of systems and to represent data in a way such that regional differences between neighboring systems will have a small impact on the communication semantic”.

  The project can be considered as a milestone in the direction of achieving a wide information interoperability in railways, as it is referenced by many other research projects.

- **CAPACITY4RAIL: Capacity for Rail**

  CAPACITY4RAIL[59] (FP7 – 2013-2017) is focused on the railway infrastructure and it is aimed at paving the way for the future railway system, delivering coherent, demonstrated, innovative and sustainable solutions for:

  - Track design: transversal approach for infrastructure solutions for conventional mixed traffic and very high speed, integrated monitoring and power supply, reduced maintenance, new concept for highly reliable switches and crossings.
  - Freight: longer trains, lower tare loads, automatic coupling, enhanced braking, modern, automated, intelligent, fully integrated system for efficient, reliable and profitable freight operations.

58 www.ontime-project.eu

59 www.capacity4rail.eu

- Operation and capacity: traffic capacity computation for freight and passenger, models and simulators for planners: capacity generation, traffic flow, resilience to perturbations, ability to recover from disturbance, computerised real time information to customers and operators at any time.
- Advanced monitoring: Integration of Advanced Monitoring Technologies in the design and building process, for an easier-to-monitor (self-monitoring) infrastructure with low cost and low impact inspection.

The full sustainability of the developed solutions and innovations have been assessed and scenarios for a smooth migration of the system from its current to its future state have been evaluated.

More in detail, some objectives relevant for interoperability have been identified:

- provide an overall increase in railway capacity by developing a holistic view on the railway as a system of interacting technical components driven by customer demand;
- develop the vision and requirements of the railway system in 2030/2050 including environmental and socio-economic aspects in terms of infrastructure and operation (including provision for an increase of freight traffic);
- identify the technologies and development/implementation steps in short-, medium- and long-term, necessary to move towards the railway system for 2030/2050;
- specify the requirements of an efficient freight rail freight system 2030/2050 to fulfil the EU targets;
- design modern fully integrated rail freight systems for seamless logistics and network-based performance and develop their catalogue, standardisation pack and technical specification.

Building on top of previous projects (NEW OPERA, INNOTRACK, ON-TIME), the project will improve the use of ICT solutions in the area of monitoring, maintenance and operation support. There is no evidence of usage of semantic technologies so far. However, final results and related documents are not yet available.

- **IT2Rail**

Shift2Rail Lighthouse Project IT2Rail[60] (H2020 – 2015-2017) is one of the projects which started the Shift2Rail research programme deployment in general, paving the way for further research envisaged within the Innovation Programme 4 (IP4). The main goal is to put the passengers at the center of the system and provide new services which enable them to organise, book and pay for their complete journey, from door to door, by means of a unique interface which allows them to express their needs from their own point of view and which provides a complete journey solution, hiding all the multiplicity, heterogeneity and complexity of the interactions needed to achieve it. Other major objective is to reduce the cost of interoperability by means of semantic technologies, as they allow to solve the problem of semantic heterogeneity by means of a general methodology enabling the interpretation of data, exchanged in any format and with any protocol.

IT2Rail highlighted the importance of Open Data and Semantic Interoperability as essential building blocks for improving customer experience of multi-modal networks. The critical areas

[60] www.it2rail.eu

are identified, as the development of an Ontology to define and model the railway data concepts and the need for specific services covering a broad range of applications. IT2Rail is finalizing a partial implementation of the Interoperability framework (IF) which enables technical interoperability of all multimodal services by relieving consumer applications from the task of locating, harmonizing and understanding multiple and independent sources of data, events, and service resources, which are consequently made available “as a service” (i.e., packaged resolvers in IT2Rail).

In compliance with the principle of openness, the IF is agnostic with respect to any application requiring its services, thus allowing multiple and concurrent implementations of the multimodal services and travel companions to access the full range of available data.

The IF realizes its definition by:

- providing travel applications with a uniform abstraction of data and services distributed over the world wide web as a “web of transportation data” in the form of *linked* data and service descriptors annotated with machine-readable logical statements which describe their semantics;
- providing applications with technical means to operate on such “web of transportation data”, e.g. publishing, querying, etc. where the semantic annotations are used to automate the process of discovering and matching data sets and service descriptors.

Further research work will be needed in order to complete the realization of the IF, as already in progress in several projects within the Shift2Rail programme (see next section).

- **Shift2Rail**

The semantic interoperability approach is now an integral part of the ambitious Multi Annual Action Plan (MAAP) of the Shift2Rail Joint Undertaking, particularly within its IP4 “innovation program” targeted at generating innovative ICT Passenger solutions for an attractive Railway. The semantic interoperability approach is an enabler for the interoperability of a number of rich passenger-centric applications operating on a distributed “web of transportation data” realized as a graph of semantically annotated and linked data and service resources from any number of providers in any format. The work programme incorporates results and experience from InteGRail, including lessons learned and the identification of the specific challenges associated with the adoption of the technology.

From the MAAP, specific Annual Work Plans are derived, in order to implement the S2R objectives by means of a number of research projects, selected through Annual Calls for Proposals. There are two types of calls: (i) Call for Members (CFM), reserved to the Members of the Shift2Rail JU, and (ii) Open Calls (OC), which are open to all organisations (except the Members of the S2R-JU).

Current active CFM projects include **ATTRACkTIVE** (Advanced Travel Companion and Tracking Services) and **Co-Active** (CO-modal journey re-ACcommodation on associated Travel serVices).

ATTRACkTIVE proposes new capabilities such as the capacity to create a “one stop shop” that helps customers to easily select and purchase an itinerary and assist her/him throughout her/his whole journey. In this respect, the solutions of IT2Rail will be expanded and further developed. It will guide, support, inform, and even entertain users throughout their entire itinerary, adapting to unforeseeable interruptions and events in order to propose alternative routes, including in the first and last miles. A real door-to-door travel solution including all modes of transport will be developed along with new forms of traveller experiences aiming to transform the travel itself into an “ATTRACkTIVE” part of the journey.

Co-Active addresses the general enrichment of the ‘one-stop-shop’ capability as initiated in the IT2Rail project and further completes the scope of functionality by addressing post-sale business transactions, and an underlying payment-settlement solution for co-modal retailed products and services. This provides the opportunity to focus specifically on those aspects whose level of customer-perceived risk discourages the advance purchase of co-modal travel entitlements.

Besides GoF4R, there are other four OC projects. The **ST4RT** (Semantic Transformations for Rail Transportation) project deals with semantic, ontology-based automation of transformations between heterogeneous data formats of carriers and operators in the rail transport. The **CONNECTIVE** project (Connecting and Analysis the Digital Transport Ecosystem) focuses on providing a technical framework and a set of tools that will foster the digital transformation of rail. The **COHESIVE** project (Coherent Setup and Demonstration of Integrated Travel Services) focuses on integrating and demonstrating the various technological innovations developed in the other Shift2Rail projects. Finally, The **My-TRAC** project deals with the design and development of an enhanced version of the IT2Rail Travel Companion supporting passengers in multi-model travels.

2.1.3 Semantic interoperability in academic research

In this section, we describe state-of-the-art solutions to semantic interoperability proposed by academic research. The solutions are focused, on one hand, in the transportation domain, since it is the domain of GoF4R project. On the other hand, in the domain of public administration at different levels (Europe, national or local). It is important to highlight that there are many different solutions, especially in the transport domain, that are currently in the state of the art like Transmodel[61], Netex[62] or Datex II[63]. These solutions are very relevant at European level and a lot of companies are following these models but they are not semantic solutions so they are not being involved in this analysis.

This analysis focuses on academic research on semantic interoperability solutions since it provides a wide vision of the current state of the transport and public administration domains. The analysis aims to understand if these solutions are being applied to real uses cases and which are the main improvements that they introduced to solve the semantic interoperability problem.

[61] http://www.transmodel-cen.eu/

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

[63] http://www.datex2.eu/

Transport and mobility

Different transport operators (of different modes) use different and often heterogeneous systems to support/implement their transport services. Specifically referring to the digital aspect of transport systems, a number of standards, specifications and conventions indeed exist, but they lack a broad and general interoperability; in some cases, interoperability is solved at syntactic level (by agreeing on the use of a common standard or by developing ad hoc adapters for bilateral solutions).

Interoperability is one of the IP4 Shift2Rail challenges. Shift2Rail would like to embrace the opportunity to achieve interoperability at semantic level by developing a shared model of the meaning of the exchanged information in the transport domain and by fostering the adoption of multilateral solutions.

The achievement of such semantic interoperability requires:

1) **Semantics**: a shared formalization of the transport domain itself, usually encoded in a formal model like an ontology.
2) **Data**: the representation of facts, events or phenomena of the travel domain, expressed in the respective specific models, standards and languages.
3) **Interpreters**: a set of transformations to convert existing data sources into a representation compliant to the shared semantics.

Semantic interoperability solutions

In what follows, the analysis of academic research on semantic interoperability mainly deals with innovative solutions to define a shared formalization of the transport domain encoded into a formal ontology.

- **Public transportation ontology for travel planning**

  The public transportation ontology [13] considers different concepts related to the best and more relevant planning for the passengers. The ontology is proposed to represent the different options that a passenger could have when (s)he wants to go from one place to another using public transport. The academic research described in [13] aims at specifying the process used for the ontology development and the analysis made about the main concepts of the public transportation domain for including them in the ontology. The ontology is divided in different parts. In Figure 3, Figure 4, and Figure 5 the multi-modal system, the geographic elements and the characterization of the stops points are shown, respectively. The proposed ontology appears to be a limited conceptualization of the public transportation domain focused only on travel planning by transport lines. For example, it covers only concepts relevant for travel planning whereas ticketing is not supported. Moreover, it covers the travel planning on public transport lines but not on on-demand public transportation solutions.

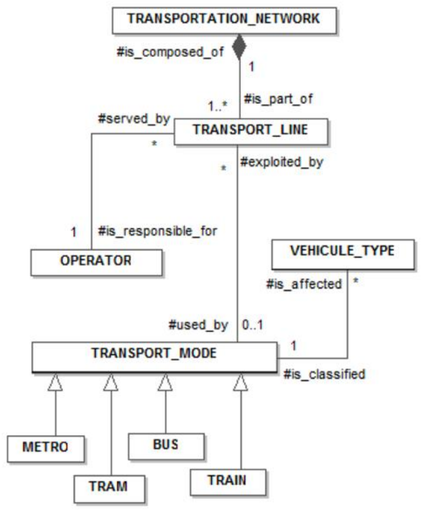

Figure 3: Public Transportation Ontology: Multimodal system [13]

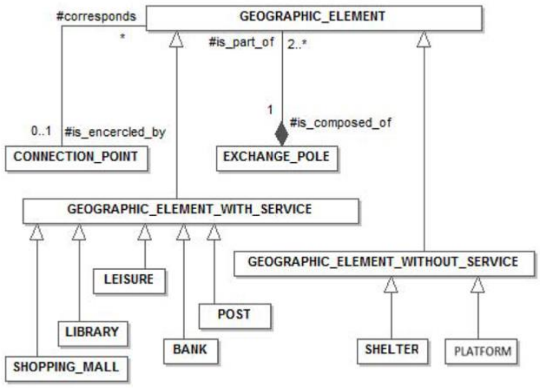

Figure 4: Public Transportation Ontology: Geographic elements [13]

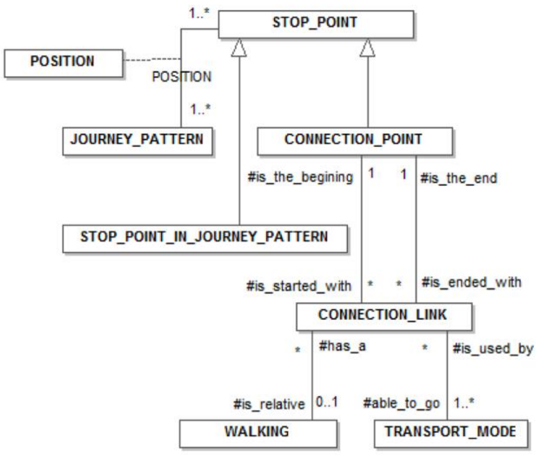

Figure 5: Public Transportation Ontology: Stop points [13]

- **The transport disruption ontology**

The Transport Disruption ontology [14] is a first attempt of a formal framework for modelling travel and transport related events that have a disruptive impact on traveller's journeys. The academic research described in [14] details the use of the framework with an interlinked repository of the travel information to support intelligent transport systems. The ontology was defined by the authors after an analysis of disruption information provided by transport authorities and operators of bus, rail, ferry, and air public transport services in the UK. The Transport Disruption ontology identifies the requirement to capture the occurrence of an event in terms of its type, location, time period, causal relationships to other events, and any impact for agent(s) that have to adapt their plan(s) because of it. The Figure 6 shows the general schema of the proposed ontology. Also in this case, the proposed ontology appears to be a limited conceptualization of the transportation domain focused on transport disruption. For example, it does not cover the passenger perspective in the description of a transport disruption (e.g., passenger rights, re-routing obligations depending on the content of the passenger's entitlement).

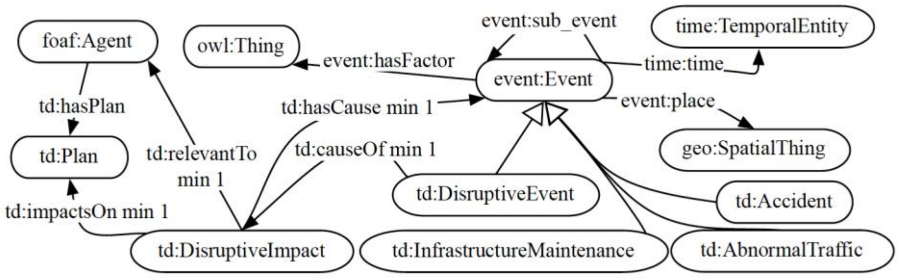

Figure 6: Transport Disruption Ontology: the general schema [14]

- **The railway core ontology**

The Rail Core Ontology (RaCoOn) [7] is an ontology model specifically tailored for use within the mainline railway industry. Although initially developed with the representation of signalling and rail infrastructure in mind, the model rapidly developed into a general model for the railways, including a core of generic railway concepts with extensions capturing particular subdomains (infrastructure, timetabling, rolling stock etc.) and an upper level model to define general concepts (e.g., the concept of "passenger") that characterize not only the railways but all the transportation domain.

A novel ontology engineering technique based on the NeON methodology [15] has been employed in designing the RaCoOn ontology, based around extracting knowledge from existing railway models and domain experts to inform and validate design decisions. This technique comprised three major steps:

1. **Specification**: High level requirements were defined. Several individual ontology modules were defined according to reusability and level of domain detail: an *upper* module for

domain-agnostic concepts, a *core* module for railway knowledge, and several subdomain-specific vocabularies including *infrastructure* and *rolling stock*.

2. **Conceptualisation, formalisation and implementation**: Each ontology module was created by repeatedly iterating over two methods: a *top down* method that draws upon expert knowledge to build a hierarchical model of a domain, and a *reuse-oriented* method where existing knowledge was extracted from models such as RailML, UIC RailTopoModel, Railway Domain Ontology (constructed in the frame of the EU FP6 InteGRail project), National Public Access Node (NaPTAN) and ArcGIS Esri. In both cases, ontology implementation was performed by defining ontology design patterns (ODPs): sets of concepts, relationships, and documentation that define how a particular concept should be encoded in the semantic data model.

3. **Evaluation and documentation**: Ontology modules were evaluated throughout the design process and then validated at the end of the design process by means of a proof-of-concept demonstrator named Train Locator.

The Train Locator demonstrator implemented a use case in which ontologies and semantic web technology were used to contextualize and enrich information from multiple systems. Simulating two different data sources, the demonstration showed how domain models allow presentation of data independently of its original syntax, such that two applications could be driven by data not originally intended for that use.

In particular, the demonstrator shows how the use of ontology can provide a bridge between legacy systems and newer replacement services without sacrificing functionality, and how interfaces between such legacy systems and more contemporary linked data-based systems can be set up.

- **The transit ontology and linked GTFS**

The Transit Ontology [16] is based on the General Transit Feed Specification and provides a vocabulary for describing transit systems, routes, stops and schedules. Figure 7 shows the core classes of the Transit vocabulary.

As stated on the website of 'Transit Vocabulary'[64], this schema has been superseded by 'Linked GTFS'. The Linked General Transit Feed Specification (Linked GTFS) is a mapping of the GTFS in CSV reference towards RDF[65]. It stays as close as possible to the CSV reference, including small additions to enhance usability.

[64] https://github.com/iand/vocab-transit

[65] https://github.com/OpenTransport/linked-gtfs

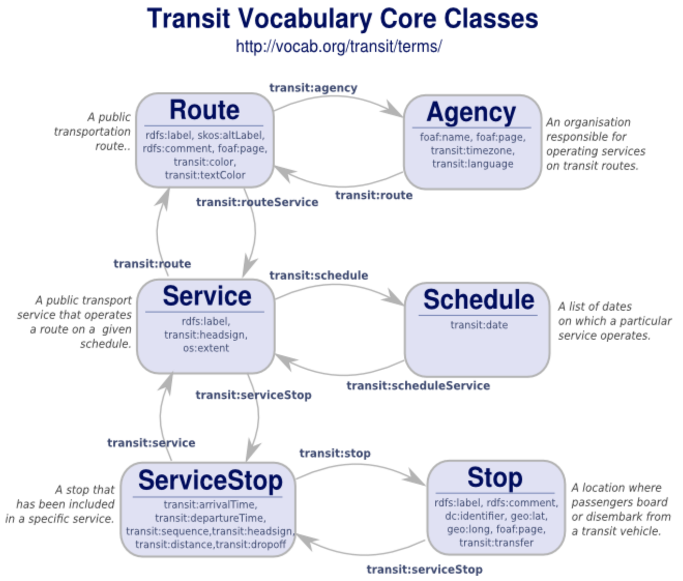

Figure 7: Transit Vocabulary Core Classes [16]

- **Web of public transport data**

  The Public Transport Domain ontology is an OWL ontology defined on existing data models and standards in the transport domain. The goal of this work [17] is to unlock the potentiality of public transport data by modelling the data in an ontology and linking it to the Web of Data. On the top of this data and by means of semantic technologies, advanced passenger information systems can be developed.

  The following steps have been followed for the creation of the ontology:

  - **Step 1: Scope of the Ontology and Competency Questions.** First, through interviews with domain experts, reviews of statistical information and focus groups, a catalogue of requirements was established. In order to outline the scope of the ontology, competency questions (e.g., "How should a user travel from A to B so he/she may get his/her shopping done on the trip?", "Where are restaurants for a romantic dinner in medium price range, open tonight, that are reachable using the bus lines 7 or 10?") has been defined.
  - **Step 2: Consider Existing Ontologies.** The authors decided to rely upon existing ontologies for some concepts not specific to public transport, such as the geographical schema of GeoSPARQL[66] and the vCard Ontology to describe postal addresses.
  - **Step 3: Terminology and Conceptualization.** In order to model the central concepts of the ontology in a class hierarchy, they first have to be selected and named. Common concepts and terminology were extracted from e.g., Transmodel, SIRI and NeTEx.
  - **Step 4: Classes and Class Hierarchies.** Based on the identified concepts, the central taxonomy of the ontology was modeled by creating the first OWL classes. Then, a top down approach was followed to define their subclasses. For each class, the corresponding concepts in the different standards were examined. In those data models, concept

[66] http://www.opengeospatial.org/standards/geosparql

hierarchies can be found as flat hierarchies in tables, expressed in UML or as entity-relationship diagrams and in XML schemas. The subclasses that are relevant for the scope of our ontology were selected and added to our class hierarchy.

- **Steps 5 and 6: Properties and Constraints.** Datatype properties were extracted from XML schemas that contained datatype properties for core elements and from UML and entity-relationship diagrams of existing data models. Based on an additional analysis of real data from public transport systems, the ontology was complemented with relevant object properties. For further refinement, domain, range and cardinalities were added to the properties.
- **Step 7: Instances.** The last step in the development process is the integration of instances into the ontology. First, there are some instances that are fixed for the domain, for example, reasons for situations. They were taken from existing enumerations. The second type of instances is application specific, for example, trips or lines. Depending on the application that uses the data, scope and detail level of the required instances may vary.

The Figure 8, Figure 9 and Figure 10 represent the main concepts of the Public Transport Domain ontology.

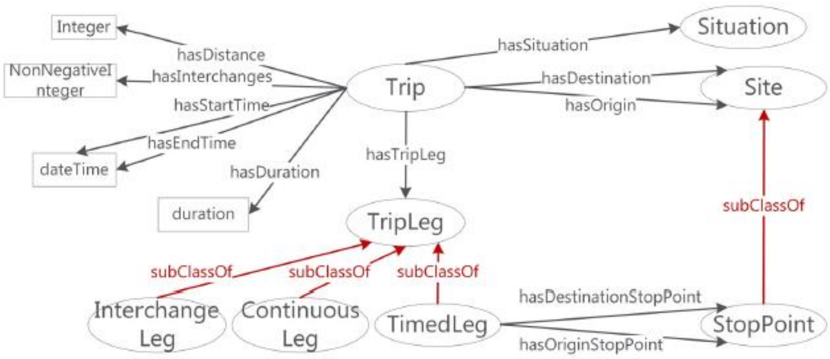

Figure 8: Public Transport Ontology: Trips [17]

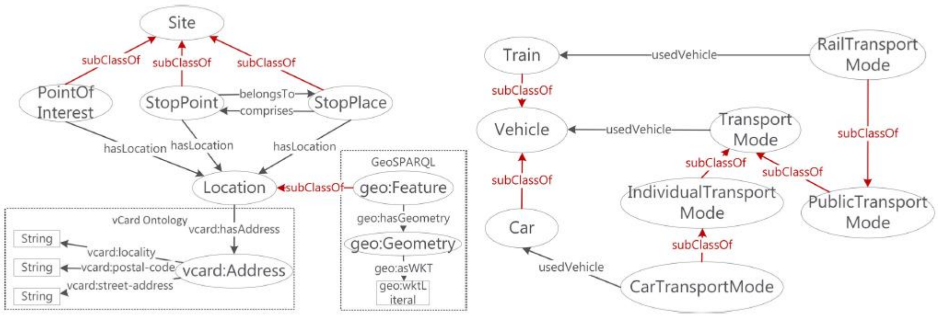

Figure 9: Public Transport Ontology: Geographical Entities and Means of Transportation [17]

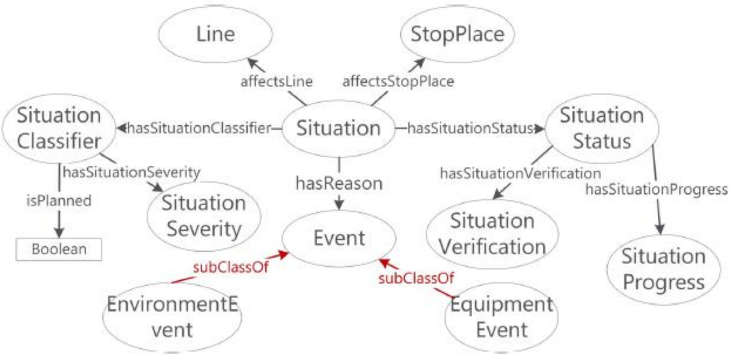

Figure 10: Public Transport Ontology: Situations and Events [17]

- **Linked connections**

  Linked Connections [18] defines a way to publish raw transit data, so that it can be used for intermodal route planning. Instead of solving each route planning query on the server-side Linked Connections enables route planning with a linear growing number of fragments with the number of connections.

  A connection, in the Linked Connections terminology, is a hop from one transit stop to another. A connection is defined by an indication of the location and time of departure, and an indication of the location and time of arrival. Connections can be extracted from a set of rules, called a transit schedule, such as a GTFS dump. When all the connections of a certain transit system are available in one sorted (e.g., descending by departure-time) array, a basic algorithm can scan the list once and come up with the earliest arrival times for all stops.

  A Linked Connection is a representation of a connection on the Web. It contains the data needed to define a connection, as well as useful links to other resources: e.g., links to GTFS terms, links to Linked Connections from different transit agencies, links to accessibility information, or links to nearby points of interests on Geonames.

  Tools for publishing Linked Connections on the Web and to transform GTFS files in Linked Connections are available for download at: http://linkedconnections.org/.

Advantages and obstacles of the analysed solutions

After the above analysis, several obstacles and advantages are identified. It should be noted that almost all the analysed solutions are focused on developing ontologies for modelling specific and limited part of the transportation domain. Due to these limitations, they cannot be adopted as reference formalization of the domain and their real applicability in industry is poor.

Moreover, the development of an ontology by itself would not be enough for a complete semantic interoperability solution; other types of features must be taken into account. For example, there are almost no solutions which propose methods of data exchange between systems adopting different data formats, as proposed in Linked GTFS.

Other solutions, such as Linked Connections, are more focused on providing a viable way of querying large amounts of data in the transport domain without focusing on current open issues such as data availability.

Therefore, there are many solutions in the state of the art that propose to model the transport domain from a semantic perspective. So the current needs of the field would be: (i) to choose a standard semantic model for domain representation where a good starting point would be the Linked GTFS since GTFS (described in ANNEX I) is a standard de-facto in the field used by several application such as Google Maps; (ii) to create easy and fast methods of data exchange between the most relevant vocabularies of the state of the art as NeTEx, Datex II or GTFS.

Public administration

Local and regional institutions, national governments and also the European Commission have put a lot of effort to create semantic interoperability solutions in the Public Administration. As we describe across this chapter there are a lot of vocabularies, initiatives or projects that have been developed in the recent years. Here, we describe and analyse some research solutions in the public administration and we identify some advantages and obstacles of them that would should take into account in next interoperability solutions developments in this domain.

Semantic interoperability solutions

In this section, we describe two recent academic research solutions based on the on W3C's Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe and one national solution from Spain focused on identify and formalize vocabularies for a smart city.

- **TransportDCAT-AP**

  TransportDCAT-AP[67] is created in the context of the CEF-OASIS project[68] with a focus on the representation of metadata about open public transport data. As transport data is considered a key asset for the development of many types of applications, it is really important to provide good metadata for open data portals focused on the provision of public transport data, or general-purpose open data portals that cover public transport among their datasets. For that reason, this profile was developed to get a common metadata model for the metadata representation of public transport data at a European level, taking into account its special characteristics.

  This profile modifies the DCAT-AP ontology to be focused on the public transport domain. TransportDCAT-AP applies a few but important changes in the standard ontology to improve the representation of the metadata in the domain:

  - The relation (dcat:Catalog, dct:spatial, dct:Location) changes from **optional** to **mandatory**.
  - The relation (dcat:Dataset, dct:spatial, dct:Location) changes from **optional** to **mandatory**.
  - The relation (dcat:Dataset, dct:keyWord, rdfs:Literal) changes from **optional** to **mandatory**.
  - The range of the relation (dcat:Dataset, dct:keyWord, rdfs:Literal) changes from a free string to a list controlled keywords properties following the format:

[67] https://oasis.team/storage/app/media/O1.2%20TransportDCAT-AP%20and%20Controlled%20Vocs.pdf

[68] https://oasis.team/

- dct:keyword A, where A is the type of transport. The available types of transport are an extension of the **route_type** property in GTFS: Tram, Metro, Rail, Bus, Ferry, Cable car, Suspended cable car, Funicular or Other. This property can be repeated if the dataset involves multiple types of transport.
- dct:keyword B, where B is the type of the dataset that A represents. The available types of datasets are followed GTFS model and are: Calendar, Fare, Frequencies, Routes, Shapes, Stops, Transfers, Trips or Other. This property can be repeated if the dataset involves multiple sets.
- dct:keyword C, where C is an INSPIRE feature for address, the AdminUnitName, where the level must be specified according to the official documentation of the framework (AdminUnitLevel3 (Province): Madrid, AdminUnitLevel4 (Municipality): Madrid).

CEF-OASIS team implements these changes since the nature of the public transport data is geospatial, so they consider that it has to be mandatory to provide this information in the generation of metadata for datasets in the catalogues of these open data portals. In addition, the modification of the keywords' property is implemented in order to facilitate querying and data visualization, for example, consulting all the datasets of a city or country associated to a specific type of public transport could be possible. Two use cases using the metadata from Madrid and Belgium transport companies were also developed[69,70].

This profile was validated by 10 companies and institutions, some of them are now describing their metadata following TransportDCAT-AP, like the transport datasets in the open data portal of Caceres city[71].

- **BotCAT-AP**

BotDCAT-AP is an RDF vocabulary[72], released with a CC-BY-4.0 license. BotDCAT-AP is also listed on Linked Open Vocabularies[73].

The vocabulary was developed starting from the DCAT and its Application Profile (DCAT-AP) with the objective to facilitate the implementation of chatbot systems by providing a formal description of all the external datasets containing useful information. In particular, the vocabulary has been used for the description of datasets published by Public Administrations. An extended description of BotDCAT-AP is provided in [19].

BotDCAT-AP enables the description of *intents* (i.e., the actions users want to accomplish by interacting with a chatbot) and *entities* (i.e., individual information units associated to an intent) supported by a dataset and the *method* to access it. A simplified UML Class diagram of BotDCAT-AP is depicted in Figure 11, where additions to the main classes and properties of DCAT-AP are highlighted. The *bot:Intent* class is designed to represent any possible intent

[69] http://crtm.linkeddata.es/metadata/CatalogCRTM.ttl

[70] https://gtfs.irail.be/metadata

[71] http://opendata.caceres.es/dataset/autobuses-caceres

[72] openly accessible at http://swa.cefriel.it/ontologies/botdcat-ap

[73] http://lov.okfn.org/dataset/lov/vocabs/bot

supported by a dataset. The relation *bot:hasEntitiesList* connects an intent to a list of supported entities enclosed in an instance of the class *bot:EntitiesCatalog*. Entities can be represented in different ways since BotDCAT-AP allows both standard and ad-hoc entities to be specified.

A dataset is represented as an instance of the class *dcat:Dataset*, and can optionally have multiple distributions denoted by *dcat:Distribution* accessible through a reference exposed by the relation *dcat:accessURL*. As of today, DCAT-AP supports only the description of data catalogs and datasets published on the web; BotDCAT-AP overcomes this limitation giving the possibility to also define different access methods to a particular dataset. This is done through the use of the relations *bot:hasMethodURL*, *bot:hasAssetURL* and *bot:hasDocumentation*, corresponding respectively to access points offered by a simple REST API, a SPARQL endpoint or any other documented method (e.g., a SOAP-based web service documented by a WSDL file). This extension supports the delivery of information that improves and speeds up the creation of the application logic needed by the chatbot system to operate at run-time.

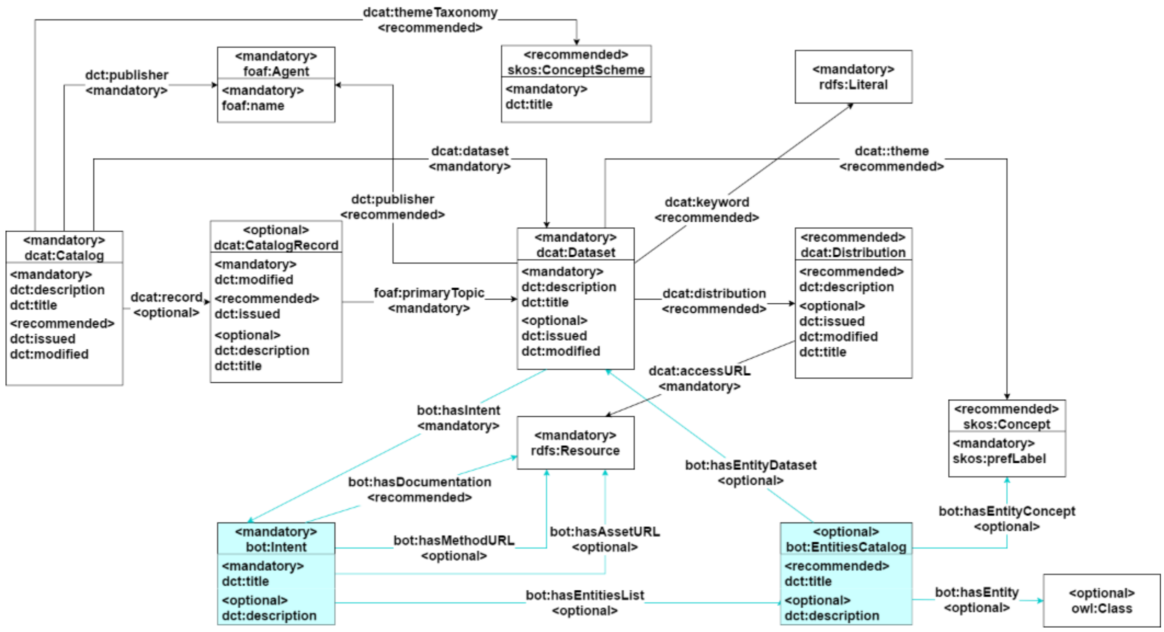

Figure 11: BotDCAT-AP simplified UML Class diagram [19]

- **OJOALDATA100**

OJOALDATA100[74] is an initiative from a set of important institutions, research groups and companies in Spain with the goal of identifying which are the 100 datasets that a city should publish in their open data portals or APIs. The main problem is that currently it does not exist any clear guideline for cities on what to publish, so there is a lot of heterogeneity in the domain and it is difficult to integrate the data from different cities. After the identification of the 100

[74] https://www.youtube.com/watch?v=v5FejbIvDrc

datasets the main goal was to define the shared data structures[75] to be used by all cities. The main idea of this initiative is to provide a simple way for the cities to publish their data following a formal and shared model (an ontology) that improves the semantic interoperability between their open data. Although the current state is not the definitive, some vocabularies are open and available[76] and the work was presented at the International Open Data Conference 2016[77].

Advantages and obstacles of the analysed solutions

After the analysis of these relevant solutions, we identified some developments following the European standard for representing metadata that improve the semantic interoperability between datasets published by Public Administrations. We also identified a national initiative focused on improving the semantic interoperability by creating formal and shared vocabularies for the most relevant 100 datasets in a city.

The main advantages of the solutions are: in one hand, that the extensions of the DCAT-AP and the vocabularies developed in the OJOALDATA100 initiative are reusing common and shared vocabularies, which improves the interoperability between the data, and in the other hand, that we have identified two real initiatives that are trying to improve the interoperability in a complex domain like public administration. They can be good examples of how other data integration process following a semantic approach can be done, as well as the formalization and homogenization of the vocabularies of a domain.

The main obstacle identified is that it does not exist any repository or way to know if somebody is doing a similar work in other place, and the solutions could be duplicated. As most solutions are open solutions, a way to know if your initiative or solution is currently developed by others or it has already been developed would be a good contribution to avoid duplicating efforts at European level. Another obstacle is how to demonstrate the relevance of metadata profiles to companies and open data portals' managers for taking them into account at the time of opening their data. The main reason for this obstacle is that, at this moment, the quality of the metadata of the open data at European level is very poor and if we are able to increase it the datasets will improve their discoverability, breaking the boundaries that still exist between the different member states.

2.2 ANALYSIS OF THE ANSWERS TO THE QUESTIONNAIRE

A survey was conducted between July and September 2017 to collect the point of view of stakeholders on the Interoperability Framework proposition and the semantic interoperability technology and skills. The questionnaire is in ANNEX IV. In total 68 companies' representatives have replied to the survey and the results are reported in the following sections.

[75] Shared data is a synonym of vocabulary and ontology

[76] https://github.com/opencitydata

[77] http://sched.co/7PVj

2.2.1 General information about the companies

In this section, relevant information about the companies that filled the questionnaire is provided. Figure 12 shows the type of the companies and their geographical distribution. Besides the information in the Figure 12, other types of organizations that filled the questionnaire are: Consulting, University, Transport Consulting Company, Service provider and Association, National Association for Public Transport Authorities, Design and consulting services for metro line, Infrastructure Manager, Start-up operating in mobility sector, NGO, Ministry of Transport and Regional Development, IT systems provider and services (i.e. call center, lost-and-found) for rail and other modes of public transport.

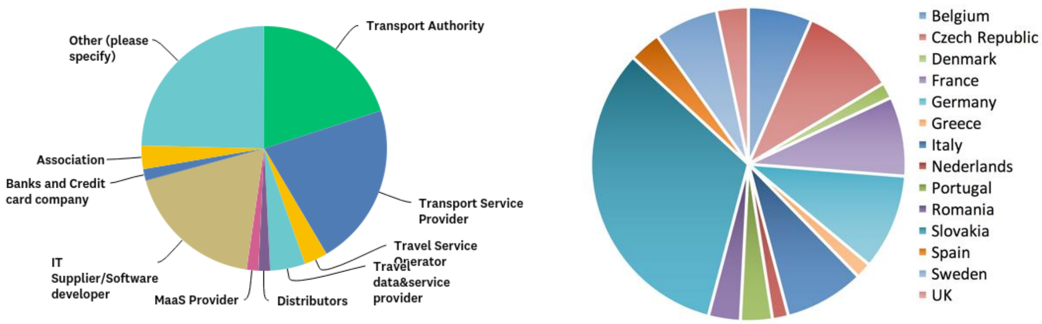

Figure 12: Company types and countries

Figure 13 shows that the participants are mainly representing the rail sector and the traditional urban transport modes (buses, metro, urban rail), although also new mobility services representatives have participated to the survey.

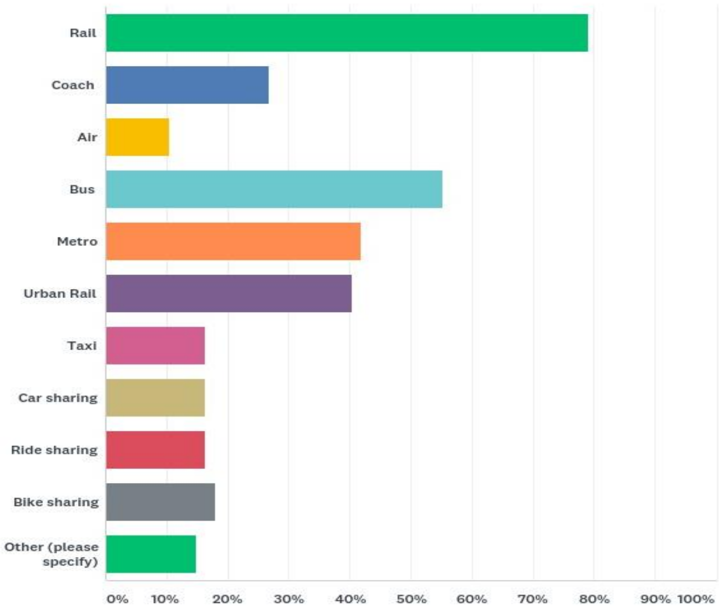

Figure 13: Companies sectors

2.2.2 Results from the survey on semantic interoperability

In this section, the results from the questions in the survey that were related with the semantic interoperability topic are shown.

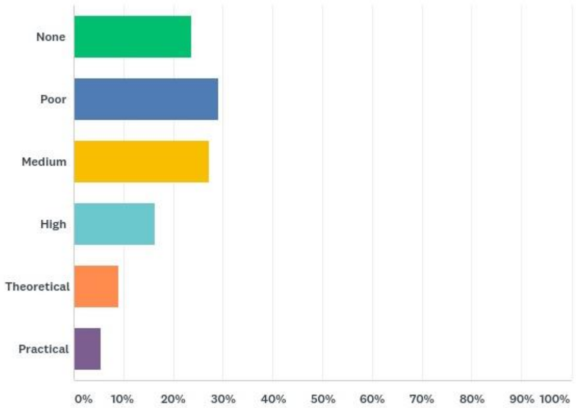

Figure 14: Q24. What is your knowledge of semantic interoperability?

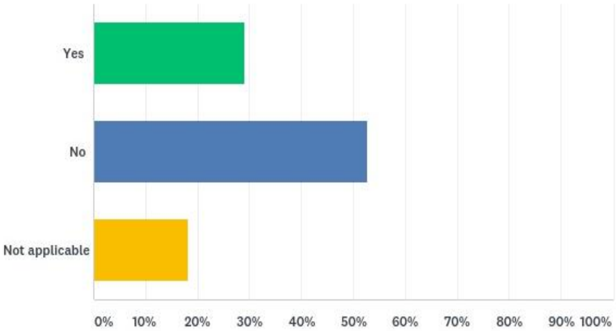

Figure 15: Q25. Does your organization adopt any semantic interoperability solution?

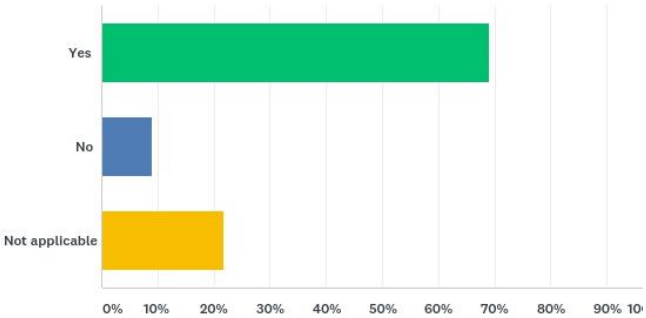

Figure 16: Q27. Do you perceive semantic interoperability solutions as beneficial for your business?

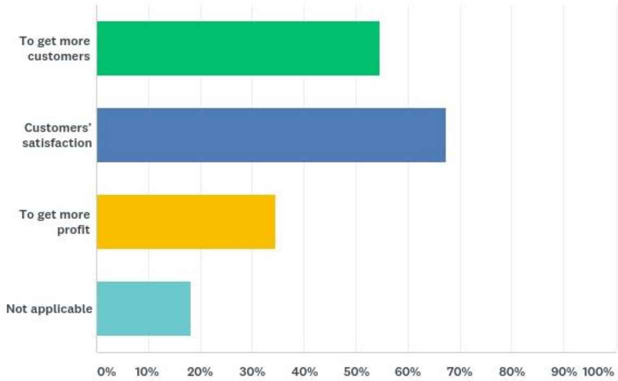

Figure 17: Q28. What is your possible business opportunities through the exploitation of such technology, which enables data and services accessibility?

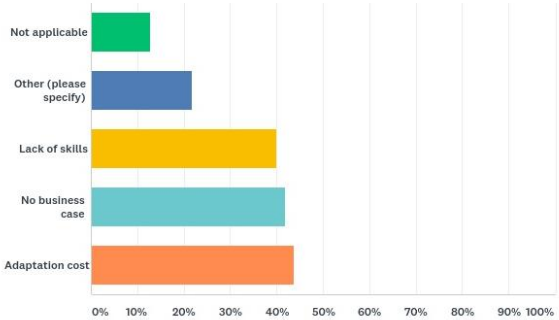

Figure 18: Q29. What are the market constraints and barriers that could hamper the adoption of such solutions?

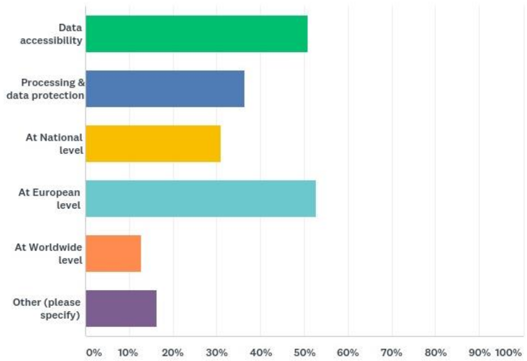

Figure 19: Q30. Do you think existing regulation framework provisions may have to be revised or upgraded?

Q31. Semantic interoperability acceptance. Do you think there are other topics which should be considered regarding the issue of “semantic interoperability” acceptance?

Replies to Q31 have been the following:

- Data sharing.
- Data reliability.
- Security of data.
- Lack of awareness of this technology.
- The interest of the customer and the public transport contracts signatories.
- Needs to be travel service provider-driven, not supply industry- or regulation-driven.
- Integration of peer-to-peer solutions is complicated: trust mechanisms should be explored since sharing of data means also sharing of customer database.
- Need for education and making the semantic modeling concepts accessible to business users.
- Solutions must adapt to a broad variety of IT maturity of the service providers. Not everyone can directly jump into such technicalities.
- Ways of easily relate the semantic interoperability mechanisms, of exposing and consuming information and transactional services, with the ways to process this services by artificial intelligence engines.
- Standardization should be considered. Enough knowledge in several projects and initiatives has been created and should start converging on a standard. Standard is a must to pave the way and trigger massive adoption.
- Hand in hand with semantic interoperability is how people expect to query over the borders of the sources. We should also carefully think about how to design Linked Data interfaces, not limiting the solutions to data dumps or SPARQL endpoints. Linked Data Fragments philosophy may lower the barriers to see the usefulness of raising the semantic interoperability.
- The experiences up to now, is that new concepts and standards has a very slow adoption, and each transport operator tends to create special solutions. Without national or European directives and/or laws this will not change.
- Stakeholder management should be improved for the acceptance of the semantic interoperability.

2.2.3 Conclusions from the survey

Although the knowledge of semantic interoperability is rather poor among the participants, this technology is perceived as beneficial for the business, mostly to reach customers’ satisfaction and gain more customers.

The majority of the stakeholders are interested in semantic interoperability beyond the current achievements and would like to understand the requirements and the skills needed to adopt it in their company. The very few companies with a knowledge of the semantic interoperability technologies have staff with education background mainly on reasoning, description logics and RDF, although they are not offering any training to further develop the knowledge in semantic technologies. This can be seen as relevant barrier for the adoption of the Interoperability Framework and it highlights the necessity to address a cultural change, engaging with the needed skills, within the companies that want to provide interoperable products and services.

However from the answers to the survey on the skills today available in companies, half of them have the necessary technical background (e.g. computer science master's degree or engineering) and are very open to innovative solutions. This means that potential skilled people are not well trained in terms of semantic web technologies, due to the lack of academic offer.
Organizations also expect a return on investment in training the staff on semantic interoperability, although it is still difficult to predict it as the IF has not reached yet the critical mass that would allow a true pan-European interoperability. For this reason, today most of the companies do not plan to contract specialized personnel to develop semantic interoperability: they are waiting for the evolution of the market.

In order to improve the acceptance of the semantic interoperability technologies, the potential adopters of the IF suggest to address the issues of data sharing, data reliability and data security. Furthermore companies are not aware of this technologies and awareness campaigns should be conducted to enlarge the community. It is also recommended to find solutions to easily relate the semantic interoperability mechanisms, of exposing and consuming information and transactional services, with the ways to process this services by artificial intelligence engines. Linked Data Fragments philosophy may lower the barriers to see the usefulness of raising the semantic interoperability.

In conclusion, the survey on the semantic interoperability technology has revealed several barriers for its adoption and for the market uptake that can be synthetized with: high adaptation cost perceived, lack of business case and lack of skills.

2.3 Feedback from the workshop in Cologne

In this section, we provide the feedback received from a set of relevant transport experts during the 1st GoF4R Workshop co-located with the 9th IT Combined Meeting at Cologne on 19 September 2017. The workshop was divided in two sessions. In the first session, GoF4R partners presented the project, a summary of the results from the questionnaire (see Section 2.2) and a brief introduction about the relevance of semantic interoperability in the transport domain[78]. In the second session, the stakeholders were assigned to three different working groups to discuss about relevant topics of interoperability: 1) Impact assessment of the Interoperability Framework, 2) Major use cases and relevant market actors, 3) Semantic interoperability.
The first two topics were selected to debate and collect advises from stakeholders on the user demand of the Interoperability Framework and the willingness of operators to be interoperable. Concern was raised that the operators may incur financial burden of implementing IF especially if this is brought in through mandatory legislation. Operators have to make revenue (selling tickets) and the investment on the IF should be worth and cost effect. For industries, sharing data and Interoperability Framework is seen as an opportunity to expand the business and increase customer satisfaction. The debate on sharing data is still open and unsolved as data ownership is seen as a competitive advantage by public and private operators. This is major barrier for the market uptake of the IF as it is based on the network effect. Complete results of the discussion are reported in the deliverable *D2.2 Analysis of the demand of market actors for the IF*.

Particularly relevant for GoF4R WP4 tasks, the third working group was organised to debate and collect feedback from the stakeholders on the work done in GoF4R on the semantic interoperability market watch. On this regard, the moderator discussed three case studies with the group in terms of (i) relevance of the case study for the transport domain, (ii) challenges and barriers, and (iii) the current available semantic interoperability solutions available for it.

The results of the discussion for the three case studies are hereafter reported.

Case Study 1: Transport data integration

Description of the case study 1: Multimodal transport is characterized by the use of more than one transportation mode within the scope of a single end-to-end journey. The IT2Rail Travel Companion acts, among others, as a multimodal journey planner promoting a more effective use of existing network capacity both within rail and in the wider multimodal transportation system. The IT2Rail Travel Companion must be supplied with data from transport operators concerning routes, stations, timetables, and fares. Such data are in various data formats (e.g., NeTEx, GTFS, railML) and, currently, the interoperability between them is only partially covered. Moreover, it can happen that the comparison of data from different sources may also lead to the detection of inconsistencies and, more in general, different levels of data quality.

[78] https://www.slideshare.net/DavidChavesFraga/semantic-interoperability-in-transport-domain

The challenge provided by case study 1 consists in identifying a semantic interoperability solution for multimodal journey planners enabling (i) batch integration of data concerning routes, stations, timetables, and fares defined in various data formats, and (ii) data quality profiling and improvement.

Results of the discussion on case study 1: Integration of different transport data – multi operators and multi services – are very important for the participants (e.g. operators like MOVIA in Denmark and DB in Germany). Today the tendency is for every country to have a central open common point to access the static data (as it will be also regulated by the National Access Point directive[79]), each of them adopting their own formats and models without any common ontology. From participants' perspective, the integration of transport data is an issue partly solved in terms of quality and availability, especially for the static data.

Indeed the main interest for the participants was real-time case: how to make available and integrate real-time transport data in a cost-effective and quality way? Semantic technology is not very well understood and in general the participants were not informed about its goal and its benefits.

Case Study 2: Semantic transformation of reservation Request/Response

Description of the case study 2: the system S1 adopting standard A (e.g., UIC 918 XML) receives a reservation request. In order to perform the reservation, S1 must use the system S2 adopting standard B (e.g., FSM). The system S2 performs the reservation and sends a response to S1. The two standards are characterized by different levels of granularities in the request/response representation of railway reservations. The request/response messages contain sensible data, such as passenger gender and PRM conditions. The described case study is under investigation in the S2R-H2020 ST4RT project.

The challenge provided by case study 2 consists in identifying a semantic interoperability solution for making S1 and S2 semantically interoperable by (i) performing the mapping between standard A and standard B for both the request and response messages, and (ii) preserving the privacy in the management of sensible data.

Results of the discussion on case study 2: This case study was remarked as very important by the participants of the working group. In this use case, semantic technologies should support the user preferences; for instance solutions of the travel shopping component should give options relevant for the user habits, e.g. solution with a longer but cheaper route. Moreover, for the participants, semantic technologies should support the implementation of a one-stop-shop for booking and purchase the tickets from the different organizations/companies at European level. The issue to solve today is that users have to access the web page of each travel service provider to buy the single tickets for multimodal journey planning and booking.

[79] Commission Delegated Regulation C (2017) 3574 final supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services.

Case Study 3: Real-time semantic data integration to better support passengers

Description of the case study 3: In addition to static data on routes, stations and timetables of the multimodal transportation network, real-time data (e.g., live delays and cancellation) are of critical importance to the effective use at network capacity. Real-time data, defined in various data formats (e.g., GTFS-Realtime, SIRI, ON-TIME RTTP) are necessary for the IT2Rail Travel Companion to re-schedule multimodal planned journey in case of disruption. For managing such real-time data on the transportation network conditions, the technical state of the infrastructure assets and vehicles must be monitored to detect faults or failures. Moreover, the traffic state (vehicle positions and speeds) as captured by trackside and onboard sensors must be detected and continuously compared with the timetable in order to identify deviations.

The challenge provided by case study 3 consists in identifying a semantic interoperability solution for multimodal journey planners enabling (i) integration of static data on routes, stations and timetables with real-time data on traffic state, and (ii) early warning for disruptions on planned multimodal journeys.

Results of the discussion on case study 3: This case study was considered as the most important for the working group. Participants agreed that today real time data is rarely available and with low quality; furthermore when available real-time data is difficult to access and integrate. Operators and authorities involved in the discussion of this working group stressed the fact that real time data is even difficult to integrate at local level; within the working group only one of the participants provides real-time data for open access, the other ones have very poor real-time data quality to provide it to third parties.

The group perceived difficult to apply semantic technologies to this use case: how could a global and common model be created? How could the performance of the solutions be improved to apply it to real cases? These are open questions that semantic interoperability should address according to the group of participants.

2.3.1 Conclusions from the workshop

In conclusion, semantic interoperability as presented by GoF4R is perceived as a very useful technology for the stakeholders' business, but the participants of the working group are not really confident in its application and use in real cases. Semantic interoperability can surely change the vision of the transport service providers. Today they first look at the local needs and only in few cases integrate international services. With the semantic interoperability, providers can design and implement their services at international level, integrating local services as part of the whole offer.

The working group suggests to review the case studies and merge the Case Study 2 with the others, as it is complementary to them.

An additional use case was suggested: how to apply semantic interoperability solutions to the passenger rights at the moment of buying a ticket (e.g. in Belgium the public transport ticket lasts 1h and in Germany 2h). This use case is particularly relevant for tourists and travelers visiting a new city and not confident with the local public transport.

3. INITIAL ANALYSIS OF THE AVAILABILITY OF SEMANTIC INTEROPERABILITY SKILLS AND PROFESSIONAL SERVICES SUPPLY MARKET

The analysis of the availability of semantic interoperability skills and of the professional services supply market conducted in Task 4.3 is aimed at understanding the potential difficulties in the adoption and evolution of semantic solutions like the Interoperability Framework in the transport domain.

The analysis is divided into three parts:

1. Interviews and surveys with transport sector companies, to understand their existing semantic technology skills and their difficulties in adopting semantic technologies.
2. Survey of academic courses/curricula and professional courses provided by companies related to semantic technologies, to understand the availability of potential skills to support the adoption of semantic technologies.
3. Survey of job offers related to semantic technologies, to understand the current amenability of companies to the adoption of semantic technologies.

The results of these analyses should constitute a valuable input to the definition of the governance policies for the Interoperability Framework.

The rest of this section is structured as follows. Section 3.1 gives an account about the interviews and surveys conducted with the transport sector companies, while Sections 3.2 and 3.3 analyse the surveys of academic and professional courses/curricula and job offers.

3.1 Interviews and Surveys with Transport Sector Companies

In order to better understand the knowledge and the usage of semantic interoperability technologies among the transport stakeholders, transport sector companies have been asked about their existing semantic technology skills and their difficulties in adopting it as part of the survey presented in Section 2.2. The following specific questions have been prepared to address these aspects:

- **Q41: If you employ (semantic) interoperability technologies, what is the education background of the people who use these technologies in your company? Are they trained in any of the following topics?**

The aggregated results of the survey are reported in Figure 20, which shows the poor knowledge of this technology in the transport sector.

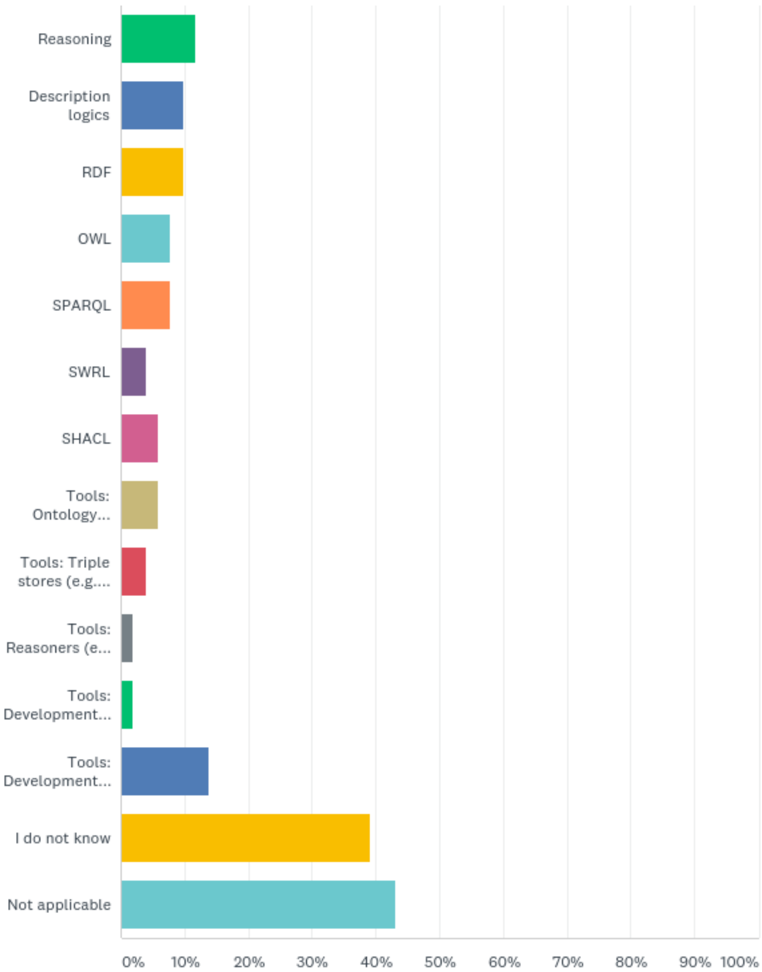

Figure 20: Percentage of employees with specific semantic technology skills

- **Q42: If you employ (semantic) interoperability technologies, what training do you offer among the following topics?**

The aggregated results of the survey are reported in Figure 21 the following diagram, which shows that most of the companies do not offer any training in the semantic interoperability related topics.

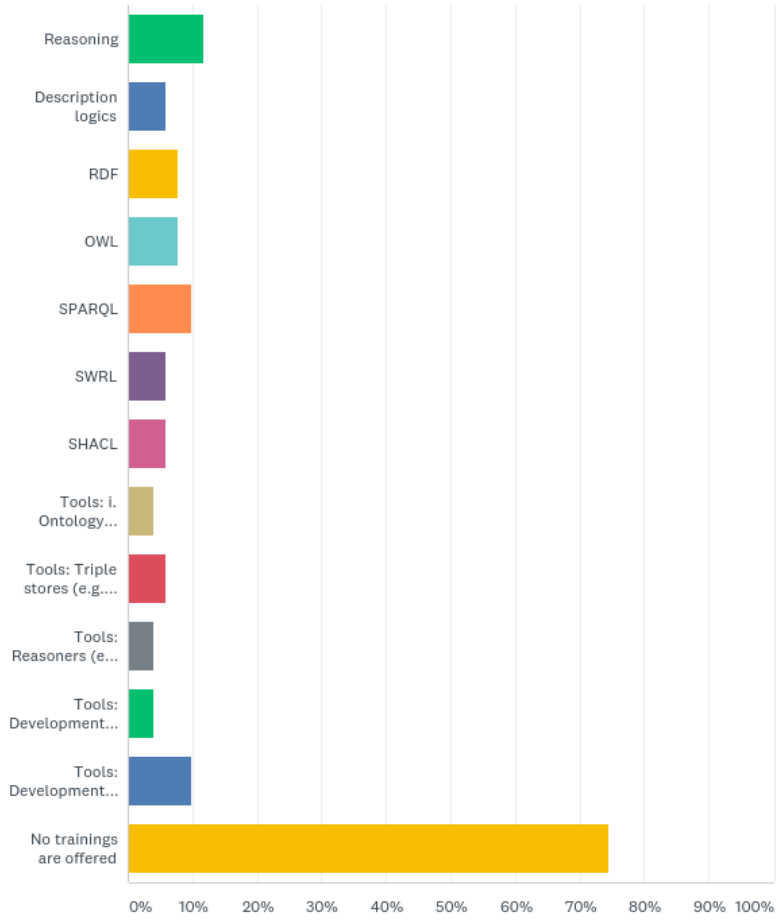

Figure 21: Percentage of companies offering specific training on semantic technology

3.2 Survey of Academic and Professional Courses Related to Semantic Technologies

This section presents a summary analysis of the surveys of academic and professional courses related to semantic technologies. The details of the surveys are given in Annexes IV (academic courses) and V (professional courses).

3.2.1 Methodology

The initial search of academic and professional courses in the first year of the project has been carried out in a rather empirical way. In more detail, relevant courses have been identified using the following three methods:

- Manual search of the Internet using keywords typical of the domain of semantic technologies, like “semantic technologies”, “semantic web”, and “ontology”.
- Just for the academic courses, explicit check of the curricula offered by well-known universities.

- Just for the academic courses, explicit check of the courses held by well-known professors of the field.

The search has been limited to the courses held in the English language.

Regarding the courses delivered by companies, only courses providing live sessions have been considered, both online or face-to-face. Therefore, courses constituted only by pre-recorded sessions have not been included in this survey; we have made this choice because pre-recorded courses do not seem to be very interesting, since they are somehow similar to written books.

Table 3 reports the information that has been retrieved for each course/curriculum.

Table 3: Information retrieved for each course/curriculum

Content item	Description
Course provider	Name of the university or the company delivering the course
Course name	Title of the course
Course URL	URL where the course is published
N. of ECTS credits (just for academic courses)	Number of credits in the European Credit Transfer and Accumulation System (if applicable)
Duration (just for professional courses)	Duration of the course
Short description of the course	A brief textual description of the contents of the course
Topics covered	Topics covered in the following list: Semantic Web, Linked Data, Open Data, ontology engineering, reasoning, description logics, RDF/RDFS, SKOS, OWL, SPARQL, SWRL, SHACL
Software tools	List of software tools taught in the course, classified in the following categories: ontology editors, triple stores, reasoners, development libraries, development frameworks

3.2.2 Results

Our search has resulted in 22 courses (20 master and 2 bachelor) and one complete master curriculum about semantic technologies, all held in Europe or in North America. Moreover, 8 professional courses delivered by companies have been found from Europe, North America and South Africa.

Table 4 shows the number of courses/curricula covering the considered topics, while Table 5 summarizes the teaching of tools in the analysed courses/curricula.

Table 4: Number of courses/curricula covering the considered topics

Topic	Number of academic courses/curricula	Number of professional courses
Semantic Web	22	8
Linked Data	19	6
Open Data	19	6
Ontology engineering	23	8
Reasoning	5	1
Description logics	12	1
RDF/RDFS	23	6
SKOS	2	3
OWL	23	5
SPARQL	19	5
SWRL	3	1
SHACL	0	0

The first aspect that can be noticed is that 21 academic courses/curricula belong to master programs, and just two academic courses belong to bachelor programs. This suggests that within universities semantic skills are considered advanced and specialized, and not much as core competences.

The skills that are most conveyed are similar in academic and professional courses/curricula. In more detail, from the analysis above it is possible to observe that the most considered topics in the analysed university courses/curricula are Semantic Web, Linked Data, Open Data, ontologies, RDF/RDFS, OWL and SPARQL; some courses deal with description logics and reasoning, while almost no consideration is devoted to SKOS, SWRL and SHACL. Likewise, the most popular topics in the professional courses provided by companies are Semantic Web, Linked Data, Open Data, ontologies, RDF/RDFS, OWL and SPARQL; some courses deal also with SKOS, while almost no consideration is devoted to description logics, reasoning, SWRL and SHACL.

Not all the courses deal with tools. However, when tools are considered within universities the most taught ones are surely Protégé and Jena, while no dominant tools can be identified within professional courses.

In addition, it must be noticed that when we have searched for university courses dealing with semantic interoperability we have observed that many universities do not offer any such courses. Similarly, the number of professional courses provided by companies that we have found in our search is quite small.

Therefore, it is possible to conclude that skills related to semantic interoperability are available, but the knowledge of students and graduates is more focused on practical aspects like OWL or SPARQL than on theoretical and formal ones like description logics and reasoning. Moreover, since not all the universities offer courses related to semantic interoperability, a homogeneous expertise on these subjects among graduates in IT disciplines cannot be expected.

The professional courses are focused on the more practical aspects as well, completely neglecting the theoretical and formal ones. Moreover, the small number of courses that we have found suggests that the need to train personnel might be difficult to satisfy, especially regarding the more theoretical notions.

Table 5: Number of courses/curricula teaching the considered categories of tools

Tool		Number of academic courses/curricula	Number of professional courses
Number of courses/curricula teaching at least one tool		12	5
Number of courses/curricula teaching at least one ontology editor		10	2
	Protégé[80]	10	1
	Oops![81]	1	0
	TopBraid	1	1
Number of courses/curricula teaching at least one triple store		2	3
	Virtuoso	1	1
	Fuseki[82]	1	0
	GraphDB	0	2
Number of courses/curricula teaching at least one reasoner		1	0
	Pellet[83]	1	0
Number of courses/curricula teaching at least one development library		6	0
	Jena[84]	5	0
	OWL API[85]	1	0
Number of courses/curricula teaching at least one development framework		0	0
Number of courses/curricula teaching other kinds of tools		0	1
	Anzo Suite[86]	0	1

[80] https://protege.stanford.edu

[81] http://oops.linkeddata.es/

[82] https://jena.apache.org/documentation/fuseki2/

[83] https://github.com/stardog-union/pellet

[84] http://jena.apache.org/

[85] http://owlapi.sourceforge.net/

[86] www.cambridgesemantics.com/

3.3 SURVEY OF JOB OFFERS RELATED TO SEMANTIC TECHNOLOGIES

This section presents a summary analysis of the survey of job offers related to semantic technologies. The details of the survey are given in Annex VI.

3.3.1 Methodology

The survey has been conducted by using the job-oriented social network LinkedIn[87]. In more detail, a keyword search using the words “ontology” and “OWL” has been executed. The word “ontology” has been chosen because it is the one that best characterizes semantic technologies. The word “OWL” has been added in order to exclude irrelevant results associated with other meanings of the word “ontology”. Then, the retrieved job offers have been considered individually, and the relevant ones have been included in the tables contained in Annex VI. Only job offers published using the English language have been taken into account.

Table 6 reports the information that has been retrieved for each job offer.

Table 6: Information retrieved for each job offer

Content item	Description
Name of the company	Name of the company offering the job
Location	City and country where the job is offered
Job title	Name of the position as published on LinkedIn
LinkedIn link	URL to reach the job offer on LinkedIn
Industry sector	The field in which the company operates (e.g., software, financial services, …)
Job functions	Roles that will be covered by the candidate
Short description	Brief description of the offered position
Experience	Level of experience required to perform the offered job
Semantic skills explicitly required by the job offer	Semantic skills explicitly mentioned by the job offer, in the following list: Semantic Web, Linked Data, Open Data, Ontology engineering, Reasoning, Description logics, RDF/RDFS, SKOS, OWL, SPARQL, SWRL, SHACL
Skills on specific software tools required by the job offer	List of software tools whose knowledge is explicitly required by the job offer, classified in the following categories: ontology editors, triple stores, reasoners, development libraries, development frameworks

3.3.2 Results

Our search has resulted in 22 job offers, located in United States, United Kingdom, India and Singapore.

Table 7 and Table 8 show the job functions of the retrieved job offers and the industry sectors of the companies. Only the job functions and the industry sectors associated with at least two job offers are included in the tables. Table 9 shows the number of job offerings requiring the considered skills, while Table 10 summarises the required knowledge of tools in the analysed job offers.

[87] http://www.linkedin.com

Table 7: Job functions of the retrieved job offers

Job function	Number of job offers
Computer science	15
Engineering	11
Marketing	2
Research	2
Sales	2

Table 8: Industry sectors of the companies offering the jobs

Industry sector	Number of job offers
Computer science and services	17
Software	12
Financial services	6
Internet	5
Hospital and healthcare facilities	2
Marketing and advertising	2
Research	2
Space and defence	2

Table 9: Number of job offers requiring the considered skills

Topic	Number of job offers
Semantic Web	7
Linked Data	5
Open Data	0
Ontology engineering	22
Reasoning	0
Description logics	0
RDF/RDFS	13
SKOS	3
OWL	22
SPARQL	13
SWRL	0
SHACL	0

Table 10: Number of job offers requiring the considered categories of tools

Tool		Number of job offers
Number of job offers requiring at least one tool		5
Number of job offers requiring at least one ontology editor		4
	Protégé	3
	Pool Party	1
	OBO-Edit	1
Number of job offers requiring least one triple store		1
	GraphDB	1
Number of job offers requiring at least one reasoner		0
Number of job offers requiring at least one development library		0
Number of job offers requiring at least one development framework		0
Number of job offers requiring other kinds of tools		0

Table 7 and Table 8 highlight that the job market deems the semantic technology skills as typical of IT specialists: according to Table 7, in fact, the candidates are almost always allocated to the computer science and engineering functions. Moreover, most of the companies seeking candidates with semantic technology skills are from the technical sector (computer science, software or internet), but these abilities are requested also in other sectors (especially financial services).

Regarding the specific competences that are mentioned in the job offers, it is possible to observe that the most requested skills are those with a more immediate practical impact: ontologies, OWL, RDF/RDFS and SPARQL. Some job offers require also Semantic Web, Linked Data and SKOS. We have not found any job offers mentioning Open Data, description logics, reasoning, SWRL and SHACL. It is possible to notice also that only few job offers require the knowledge of specific tools, like Protégé.

As a final consideration, we can observe that the fact that just few positions require prior knowledge of specific tools is a somewhat expected result. Indeed, a person with a deep knowledge of semantic technologies can easily learn the needed tools.

4. CONCLUSIONS

The back-office analysis on the interoperability of data formats and structures in the railways and transport sectors have revealed that current interoperability standards deal with different types of data: *infrastructure data* (i.e. topological information on transportation networks), *static transportation data* (i.e., information on timetables, stations, fares and electronic tickets), and *real-time transportation data* (i.e., real-time data on the status of infrastructures and vehicles).

These types of data are characterized by different interoperability challenges that can be clustered in two common data management activities:

- *Data Mapping*: mapping between data models characterized by different levels of granularities;
- *Data Integration*: integration of data specified according to different models and with different levels of Quality of Data (e.g., completeness, accuracy, timeliness and consistency).

The market watch of semantic interoperability tools has revealed that no tool covers the entire set of requirements for a semantic data integration solution in the rail and transport sector. A combination of tools appears to be required. However, since the current analysis has been performed using only online documentation of the selected tools, an extended analysis with the involvement of tools providers is planned to perform the enhanced analysis of the semantic interoperability technology market that will be described in D4.2.

The initial analysis on the adoption of semantic interoperability solutions in industry and government has revealed that the current use of this type of solutions is very low. This is the main feedback received from the survey and the workshop organized by GoF4R. The main problem of the adoption of semantic interoperability solutions is that they require (i) a shared conceptualization of the domain provided by the data owners, and (ii) a transformation/mapping process of local data schemas to the new shared conceptualization. The two requirements are not easily satisfied by e.g., transport companies because the required knowledge and skills on semantic technologies are normally not owned by these companies and institutions. Therefore, the process of adopting semantic interoperability solutions is a difficult and expensive work and it is hard for a company or institution to see the benefits of this change. Currently, some prototypes like Tripscore[88] [21] or iRails[89] have been developed but they are research and innovation prototypes that could only provide an idea of what could be the advantages of implementing this solutions in the transportation domain.

Throughout the survey and the workshop organized by GoF4R, the participants have been asked to assess the impact that the IT2Rail Interoperability Framework based on sematic technologies may have on their business. Both workshop and survey highlight the importance for the business actors and the transport authorities of the multimodal interoperability. The Interoperability Framework based on sematic technologies appears as an interesting solution and its governance should define rules, roles and tasks, and should be market driven, open and inclusive. The main barriers for market adoption of the IT2Rail Interoperability Framework identified by the participants to the survey and

[88] http://tripscore.eu/

[89] https://irail.be/route

the workshop can be synthetized with: high adaptation cost, lack of business case, unclear understanding of the benefits and lack of semantic technology skills.

The surveys about the courses/curricula and job offers in the scope of semantic technologies have identified a correspondence between the skills required in the job market and those provided by the existing courses. Indeed, the analysed job offers are mostly requiring skills with immediate practical impact like ontologies, OWL, RDF/RDFS and SPARQL, and these topics are the same being mainly conveyed by the analysed courses/curricula. Nevertheless, it has emerged also that recruiting personnel proficient in semantic technologies might be complicated. The surveys in fact have underlined how the competences supplied by universities are actually not homogeneous, and as a consequence not all the IT specialists have the necessary expertise in the semantic technologies domain. Also, it is not so simple to find professional courses to adequately train the employees.

Transport operators willing to adopt the Interoperability Framework would need specialized people. Given the possible difficulties in finding personnel with the necessary skills, companies should take a special care about the selection process. The potential difficulties of transport operators in finding appropriate personnel should be taken into account by the definition of the governance schemes for the Interoperability Framework.

4.1 Input for the Definition of the Interoperability Framework Governance Structure

Input from T4.1 “Maturity of the Semantic Interoperability Solutions Supply Market”: the back-end analysis of main standards in the railways and transport sector reveals that a large range of solutions related to the interoperability of data formats and structures already exist. Solutions which include semantic technologies exist but seem to be not completely developed or focusing on one transport sector. In some cases, the interoperability solutions are supported by stakeholders organisations, which agree to cooperate in the creation and maintenance of common platforms. The definition of the governance structure of the Interoperability Framework that is the main objective of GoF4R-WP5, should consider these currently existing examples of governance structures in the field.

The GoF4R project identifies a list of asset types that are used by the IT2Rail Interoperability Framework to enable technical interoperability of all multimodal services in the transportation domain. The identified asset types, also shown in Figure 22, are:

- ***Ontologies***: formal specifications of shared conceptualizations of specific domains (or a part of them).
- ***Web Service Descriptors***: metadata descriptions of the functionalities offered by specific Web services and their binding information.
- ***Business Rules and Processes***: rules and processes that characterize a specific standard and/or a system adopting it.

- ***Schemas and Mappings***: schemas or mappings between schemas/ontologies aiming at supporting the technical interoperability between services/systems adopting different standards in the transportation domain.

- ***Data Sets***: data sets providing information (e.g., timetables) on multimodal services in the transportation domain. This asset type presents two levels to be managed and governed: the metadata description of the dataset and the data included into the dataset. An example of this asset type is a NeTEx dataset containing the timetable of a specific transportation mode that it is also described by means of a DCAT-AP metadata description.

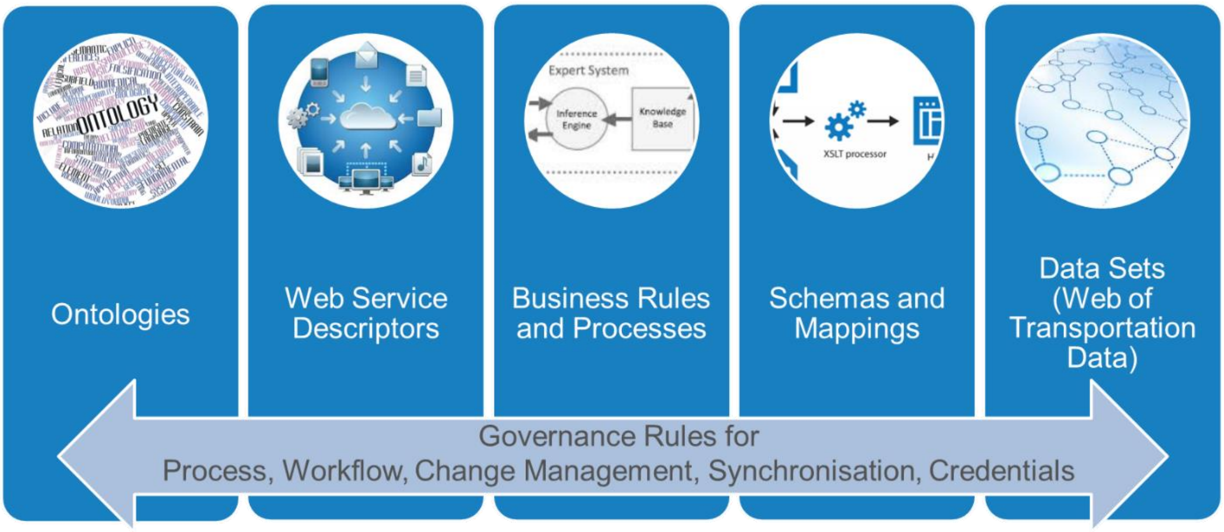

Figure 22: Interoperability Framework Assets

The IT2Rail Semantic Assets Manager, fully described in IT2Rail D1.2 "Semantic Web Service Registry" [20], is the component devoted to govern the process of maintaining a shared repository of the mentioned asset types. It is an organized collection and storage of these assets, enhanced by tools to support a workflow process for review, versioning, approval and publishing of the assets. One of the objective of the IF governance process that GoF4R will define concerns the design of the lifecycle of each asset type.

Here, we aim at providing input to GoF4R-WP5 "Governance and management structure for interoperability framework" concerning which assets have to be governed to support the development of semantic data integration solutions described in Section 1.2. Each asset is described by: (i) asset type (see Figure 22); (ii) the actor(s) who acts on it to realize and use the mentioned solutions; (iii) how the actor(s) acts on it; (iv) potential risks for the asset governance. The result is in Table 11.

Table 11: Relevant assets for semantic data integration solutions

	Asset Type	Involved Actors	Actors' needs	Risks
Reference Ontology	Ontologies	Solution Designer and Developer	Download the asset; Use the asset.	Versioning of the asset.
Data Schemas *(i.e., schemas of the data to be integrated)*	Schemas and Mappings	Solution Designer and Developer	Download the asset; Use the asset.	Versioning of the asset.
Data Sets *(i.e., the data sets to be integrated)*	Data Sets	Solution Designer and Developer. Solution users.	Download the asset; Use the asset; Integrate the asset.	IPR; heterogeneous data formats; Versioning of the asset.
Data Sets Metadata Descriptions *(i.e., the metadata of the data sets to be integrated)*	Data Sets	Solution Designer and Developer.	Download the asset; Use the asset.	Heterogeneous metadata descriptions; Versioning of the asset.
Local Ontologies	Ontologies	Solution Designer and Developer	Publish the new assets.	Quality of the assets; Language used.
Mappings between Ontologies *(i.e., mappings between local and reference ontologies)*	Schemas and Mappings	Solution Designer and Developer	Publish the new assets.	Quality of the assets.
Data Warehouse Ontology / Global Ontology	Ontologies	Solution Designer and Developer	Publish the new assets.	Quality of the assets; Language used.

<u>Input from T4.2 "Adoption of Semantic Interoperability Solutions in Industry and Government"</u>: The EU Commission Delegated Regulation C (2017) 3574 final, supplementing Directive 2010/40/EU with regard to the provision of EU-wide multimodal travel information services, provides a guidelines for publishing open transport data that every state member will have to take into account in the next four years when going through this process.

These recommendations suggest that each state member should publish all the transport (static and real-time) data in a unique and common **National Access Point (NAP)** in a machine-readable format. The main motivation is to provide a centralized and common way to query all the transport data at EU level to improve the developments of application that will be able to get data from distributed sources as if they were centralized. To do this, one option could be to do a data dump into a centralized data warehouse, but it has a lot of problems, for example related on how to provide access to the real time data or how to maintain the database updated if it depends on a large number of sources. Other option could be to apply semantic interoperability solutions that have a big potential and will be able to resolve the issues that the domain is demanding. However, as remarked in Section 2, the maturity and adoption of these type of solutions in the transportation domain is low. The input for the GoF4R-WP5 from T4.2 consists in considering the realization of a NAP to static and real-time data as a use case to be studied in WP5. The feasibility to realize a NAP upon the IT2Rail Interoperability Framework should be investigated and, in case, a proper governance should be proposed.

Other input for the GoF4R-WP5 from T4.2 are the feedbacks gathered from the survey and the workshop organized by GoF4R.The Interoperability Framework based on sematic technologies appears as an interesting solution in the transportation domain but its governance should clearly define rules, roles and tasks, and should be market driven, open and inclusive. Moreover, in order to define a proper governance process, GoF4R-WP5 should considered the following main barriers for market adoption of the Interoperability Framework identified by the participants the survey and the workshop: high adaptation cost, lack of business case, unclear understanding of the benefits and lack of semantic technology skills.

Input from T4.3 “Availability of Semantic Interoperability Skills and Professional Services Supply Market”: The analysis of the skills conveyed by academic and professional courses related to semantic technologies, and the job offers requiring personnel trained in this domain has highlighted a few facts that are relevant for the definition of the structure of the governance of the Interoperability Framework. In particular, among the conclusions drawn above, two are very relevant in this regard. First, it has been found that people skilled about semantic technologies are available, but they might be hard to find, and a homogeneous expertise on these topics cannot be expected among IT graduates. In addition, knowledge in the field of semantic technologies appears to be mostly a prerogative of IT specialists.

The possible lack of adequately trained personnel leads to the risk of having actors not skilled enough in semantic technologies to interact in a suitable way with the assets of the Interoperability Framework. This means that when defining the governance workflows and rules, it will be necessary to devote a special care to all activities allowing the modification of the assets. It will be essential to envisage very robust review and versioning processes supporting the changes.

The fact that the experts in semantic technologies are usually IT specialists, on the contrary, might create problems to the definition of the committees managing the assets, another important activity that should be foreseen in the governance structure. IT specialists, indeed, may be appropriate to manage the assets under a technical point of view, but they usually lack the domain knowledge that is required to understand whether the assets are being used in the proper way, and are undergoing reasonable changes. When defining the committees, therefore, it will be crucial to place side by side experts in semantic technologies and also domain experts, the latter providing the essential domain knowledge.

4.2 Next Steps

The GoF4R Advisory Group, composed by relevant stakeholders of the rail and transport domain, will be involved in the validation of the initial analysis of the semantic interoperability market described in this deliverable. Moreover, the GoF4R Advisory Group will be involved in the identification of additional drivers to be used to extend and enhance the analysis.

Concerning T4.1 “Maturity of the Semantic Interoperability Solutions Supply Market”, the following next steps are planned:

- Back-office analysis of interoperability standards in the aviation sector and in other potential sectors suggested by the GoF4R Advisory Group and/or emerged from Shift2Rail collaboration meetings. The objective of the analysis is to understand to what extent semantic interoperability has been taken into account.
- Revision of the list of requirements for semantic interoperability in the transportation domain (described in Section 1.2) guided by additional case studies. Candidate case studies are:

- Semantic transformations of reservation request/response messages between systems adopting different standards (from the S2R IP4 ST4RT Project);
- Real-time semantic data integration to better support passengers in case of disruptions.

- Extended analysis of semantic interoperability tools (described in Section 1.3.2) by involving the providers or developers of the tool, which will be asked to answer a questionnaire and possibly participate in an interview.

Concerning T4.2 "Adoption of Semantic Interoperability Solutions in Industry and Government", the following next steps are planned:

- Selection of additional application domains (e.g., media industry, medical domain, public sector information) and scouting of adopted semantic interoperability solutions. Back-office analysis of the identified solutions, aimed to understand the success conditions, the lessons learned and the best practices.
- Identification of (i) relevant dimensions to be used for the evaluation of semantic interoperability solutions; (ii) cross-domain lessons learned and the best practices, and (iii) the advantages as well as the possible obstacles to adoption.
- An analysis shall be made of the proposals made in the EC Communication COM (2017) 134 final adopted on 23 March 2017 on the European Interoperability Framework regarding the actions to be developed for adoption at a large scale of semantic interoperability solutions. This analysis shall be presented to the UITP experts for debate in a workshop organised jointly with WP5.

Concerning T4.3 "Availability of Semantic Interoperability Skills and Professional Services Supply Market", the following next step is planned:

- Extension of the analysis performed in the first year by considering further semantic interoperability skills, identified by the GoF4R Advisory Group and/or emerged from the case studies analysed in T4.1. Candidate new skills to be included in the analysis are: transformations between ontologies and other data models, semantic mappings, ontology matching and alignment, ontology merging.
- Extension of the analysis performed in the first year by searching more academic courses/curricula, professional courses and job offers. The driver to the extension of the list of courses and job offers should be the attempt to cover in a systematic way all the EU countries, in order to identify potential differences between them.

REFERENCES

[1]. EU Commission. EU-report on 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]. General Transit Feed Specification (GTFS). Information can be found from https://developers.google.com/transit/gtfs/

[3]. CEN/TS 16614. Network and Timetable Exchange (NeTEx). Information can be found from http://netex-cen.eu/wp-content/uploads/2015/12/06.NeTEx-Network-WhitePaper_1.08.pdf

[4]. Bellatreche L., Khouri S., Berkani N. (2013) Semantic Data Warehouse Design: From ETL to Deployment à la Carte. In: Meng W., Feng L., Bressan S., Winiwarter W., Song W. (eds) Database Systems for Advanced Applications. DASFAA 2013. Lecture Notes in Computer Science, vol 7826. Springer, Berlin, Heidelberg.

[5]. De Cesare S., Foy G., Lycett M. (2016). A Semantic Data Integration Framework. In Proceedings of *ICEIS 2016*, pp. 127-134.

[6]. Bellatreche, L., et al.: Contribution of ontology-based data modeling to automatic integration of electronic catalogues within engineering databases. Computers in Industry Journal Elsevier 57(8-9), 711–724 (2006).

[7]. Tutcher, J., Easton, J. M., & Roberts, C. (2015). Enabling Data Integration in the Rail Industry Using RDF and OWL - the RaCoOn Ontology. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*.

[8]. Keller, C., Brunk, S., & Schlegel, T. (2014). Introducing the Public Transport Domain to the Web of Data. WISE 2014, Part II, LNCS 8787, S. 521–530

[9]. Shvaiko, P., & Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering*, 25(1), 158-176.

[10]. Cruz, I. F., & Xiao, H. (2005). The role of ontologies in data integration. Engineering intelligent systems for electrical engineering and communications, 13(4), 245.

[11]. Vdovjak, R., & Houben, G. J. (2001, April). RDF-Based Architecture for Semantic Integration of Heterogeneous Information Sources. In Workshop on information integration on the Web (pp. 51-57).

[12]. Cruz, I.F., Xiao, H., and Hsu, F.. Peer-to-Peer Semantic Integration of XML and RDF Data Sources. In *The 3rd International Workshop on Agents and Peer-to-Peer Computing (AP2PC 2004)*, July 2004.

[13]. Houda, M., Khemaja, M., Oliveira, K., & Abed, M. (2010, May). A public transportation ontology to support user travel planning. In Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on (pp. 127-136). IEEE.

[14]. Corsar, D., Markovic, M., Edwards, P., & Nelson, J. D. (2015, October). The transport disruption ontology. In International Semantic Web Conference (pp. 329-336). Springer International Publishing.

[15]. Suarez-Figueroa, M. C., Gomez-Perez, A., and Fernandez-Lopez, M. (2012). The NeON Methodology for Ontology Engineering." Ontol. Eng. a Networked World, M. C. Suarez-Figueroa, A. Gomez-Perez, E. Motta, and A. Gangemi, eds., Springer Berlin Heidelberg.

[16]. TRANSIT: A vocabulary for describing transit systems and routes. Available at: http://vocab.org/transit/terms

[17]. Keller, C., Brunk, S., & Schlegel, T. (2014). Introducing the Public Transport Domain to the Web of Data . WISE 2014, Part II , LNCS 8787, S. 521–530 .

[18]. Colpaert, P., Llaves, A., Verborgh, R., Corcho, O., Mannens, E. and Van de Walle, R. (2015), “Intermodal public transit routing using Linked Connections”, in *Proceedings of the 14th International Semantic Web Conference: Posters and Demos*, Vol. 1486, CEUR Workshop Proceedings.

[19]. Cappello, P., Comerio, M., & Celino, I. BotDCAT-AP: An Extension of the DCAT Application Profile for Describing Datasets for Chatbot Systems. In proceedings of the 4th International Workshop on Dataset PROFIling and fEderated Search for Web Data (PROFILES 2017) co-located with The 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22, 2017.

[20]. IT2Rail D1.2 “Semantic Web Service Registry”. Available at: http://it2rail.eu/download.aspx?id=eb11d71c-ecf7-4ce9-82ae-c7285531645c

[21]. Chaves-Fraga, D., Rojas, J., Vandenberghe, P.-J., Colpaert, P. & Corcho, O. (2017). The tripscore Linked Data client: Calculating specific summaries over large time series. In proceeding of the Workshop on Decentralizing the Semantic Web (DeSemWeb) co-located with The 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22, 2017.

ANNEX I

Description: Annex I is the document titled “Scouting of Interoperability Standards in the Rail and Transport Sector” provided together with D4.1. The document takes into consideration all regulations, standards and specifications related to the transportation sector, mainly focusing on the railway and public transport modes, with a first analysis related also to automotive and aviation modes. Some document which apply to all transportation modes have been included as well.

In accord with such structure, the contents are organised according to the following table.

2. ANALYSIS OF INTEROPERABILITY IN THE RAIL SECTOR			
	2.2 Regulation documents		
		2.2.1	Interoperability Directive
		2.2.2	TAP TSI
		2.2.3	PRM TSI
		2.2.4	Regulation on rail passengers' rights
	2.3 Standardisation documents		
		2.3.1	EN IEC 61375 series
		2.3.2	EN IEC 62580 series
		2.3.3	EN 50463 series
	2.4 Specification documents		
		2.4.1	IRS 50405
		2.4.2	UIC 918 XML
		2.4.3	railML
		2.4.4	UIC RAILTOPOMODEL / RAILML 3
		2.4.5	THE FULL SERVICE MODEL (FSM)
		2.4.6	ON-TIME RTTP – Real Time Traffic Plan
3. ANALYSIS OF INTEROPERABILITY IN THE PUBLIC TRANSPORT SECTOR			
	3.2 Regulation documents		

	6.2 Regulation documents	
		6.2.1 Multi-Modal Information
		6.2.2 Privacy and Personal Data Protection

Further evolution of the document, especially related to the automotive and aviation modes, is planned during the second year of project work.

ANNEX II

Description: Annex II is composed by two documents provided together with D4.1.

Annex IIA is the document titled “Standards Metadata Descriptions” that aims at describing the interoperability standards identified in ANNEX I, using the sets of metadata defined by the GoF4R project. The aim of the metadata description is (i) to support the identification of the requirements for semantic interoperability tools in the railway and transport domain and (ii) to understand to what extent semantic interoperability has been currently taken into account.

Annex IIB is an Excel Document with the same contents and scope of Annex IIA. Filters and selection functions supported by Microsoft Excel can be used to search and compare the analysed interoperable standards.

ANNEX III

Description: Annex III reports the survey conducted by the GoF4R project to evaluate the knowledge of semantic web technology in the transportation domain.

The survey has been conducted by means of a questionnaire and interviews to collect information from European main transport stakeholders about the current market of interoperability platforms and frameworks in Europe. The questionnaire aims at exploring the needs for interoperability for a multimodal and seamless end-to-end mobility offer to customers, as well as the benefits and the barriers. Furthermore, the information collected would help explore market actors' interests in the Interoperability Framework technology and achieve an overall understanding about if and how these market actors approach the issue of semantic interoperability and semantic web technologies.

Finally, the survey aims also at addressing the analysis of interoperability and semantic technology skills just available in transport companies.

ANNEX IV

Description: Annex IV provides a list of academic courses or curricula related to semantic interoperability, which has been compiled by selecting some of the most relevant Universities. Each course/curriculum is classified on the basis of the topics and technologies that are dealt with. Annex IV shows the results of the analysis that aims to identify the semantic interoperability skills that may be acquired through academic courses.

ANNEX V

Description: Annex V shows a list of training courses provided by companies related to semantic interoperability. Each course is classified on the basis of the topics and technologies that are dealt with. Annex V shows the results of the analysis that aims to identify the semantic interoperability skills that may be acquired through training courses provided by companies.

ANNEX VI

Description: Annex VI provides a list of jobs offered by companies related to semantic interoperability. Each job is classified on the basis of the skills and knowledge it requires. Annex VI shows the results of the analysis that aims to identify the semantic interoperability skills that are required in the job market.