ST4RT NEWSLETTER

November 2017

INTRODUCTION AND SCOPE

Following a quick introduction of what the ST4RT (Semantic Transformations for Rail Transportation) project is about, this newsletter will give you some useful information about the latest developments, since the project has reached its halfway point. The ST4RT project started on the 1st November 2016 and is set to last for 24 months. By the time you read this newsletter, you should have a good idea of the overall concepts, the developments so far, along with the challenges and the ambition for the final results. After giving a brief overview of the project, the newsletter will explain in more detail the ST4RT approach that has been followed. Then more details on the use case will be given, which will be the basis for the ST4RT demonstrator, followed by some explanations related to the annotation and conversion processes that are being used in the project.

The primary objective of the ST4RT project is to develop a demonstrator tool that will provide ontology-based transformations between different standards and protocols, resulting in enhanced semantic interoperability between disparate, heterogeneous legacy systems.

Such transformation technology is essential to achieving the goals for the Interoperability Framework that is at the core of the fourth Innovation Programme (IP4) of Shift2Rail. The Interoperability Framework will provide the right tools in order to introduce seamless mobility services, foster the development of multi-modal travel services and help to overcome the obstacles currently impeding the development of market innovation and limiting a large acceptance of the semantic web for transportation. The main technical ambition of the Interoperability Framework is "the provision of the tools and technologies that will allow data exchange among different actors of the transport ecosystem, providing mechanisms to abstract data consumers from the complexity of varied data formats and non-integrated services, facilitating interoperability among systems and the creation of added value services for achieving a seamless multimodal door-to-door experience". In line with the objectives stated in IP4, such transformation technology is necessary to assure that technical interoperability can be deployed effectively and cost-efficiently by market actors in order to create service offerings that substantially improve mobility. Such transformation technology is a powerful tool that will allow us to meet the challenge to overcome the complex misalignment of eco-system services due to differences in business models and legacy systems.

The Semantic Transformation, as an element of the underlying Interoperability Framework technologies and standards as developed in IP4, will address fundamental obstacles to interoperability by overcoming incompatibility. This technology, if widely deployed, will be able to:

- overcome the fragmentation of multiple data formats and communication protocols;
- connect multi-modal providers and the services sectors;
- lower the cost of accessing data that is openly discoverable but of low quality or availability;
- maximise growth potential for the development of new products and services by removing ICT systems incompatibility, e.g. enabling market players and new entrants' capabilities, thereby reducing the cost and time-to-market for the ICT integration.

ST4RT APPROACH

Interoperability refers to the ability of devices or systems to participate in the coordinated execution of tasks and functions in some business process, in which data/information exchange happens in a transparent way, according to common protocols and formats. While interoperability can be achieved at the "syntactic" level, a full interoperability should be sought at "semantic" level – i.e., when interoperating systems are able to interpret the exchanged information automatically and meaningfully, so that exchanged messages are unambiguously defined and understood by the different parties. The Interoperability Framework at the core of IP4 of Shift2Rail enables the latter approach through the creation of an explicit, formal, shareable, machine-readable and computable description of the semantic reference model associated with data descriptions and exchanges; the ultimate goal is to allow for a higher degree of automation of distributed processes across multiple data formats and spanning unspecified actors.

The ST4RT project builds on this idea of a common shared ontology to enable the communication between services that adopt different standards for the representation of data. The general approach adopted by the ST4RT project is shown in Figure 1: two different standards A and B are semantically annotated in order to create mappings from their data models to the global reference ontology – currently represented by the IT2Rail ontology, but possibly extended to a network of interrelated domain ontologies. The resulting mappings are the basis for the semantic transformations of ST4RT, so that data expressed in the two standards can be converted into their respective ontological version.

Whenever two systems that adopt the different standards A and B need to exchange information, the semantic transformation takes place. As in the depicted example, a message originally expressed with regards to standard A – because it is generated by a system adopting standard A – is "lifted" to its ontological version, by means of the mapping between standard A and the reference ontology; once in its ontological counterpart, the message can then be "lowered" to a message expressed with regards to standard B, by means of its respective mapping to the reference ontology, and can be consumed by the target system. Of course, the same procedure can be followed when the message exchange is initiated by the system adopting standard B: a message expressed in standard B is first lifted to its ontological version and then lowered to a message expressed in standard A to be consumed by the target system.

This approach avoids the costs associated with the creation and maintenance of several point-to-point conversions that, with the increase of the number of participant systems and their respective data models, quickly become unmanageable. On the contrary, with the adoption of semantic interoperability technologies, only the mappings to the ontological level must be maintained and aligned over time.

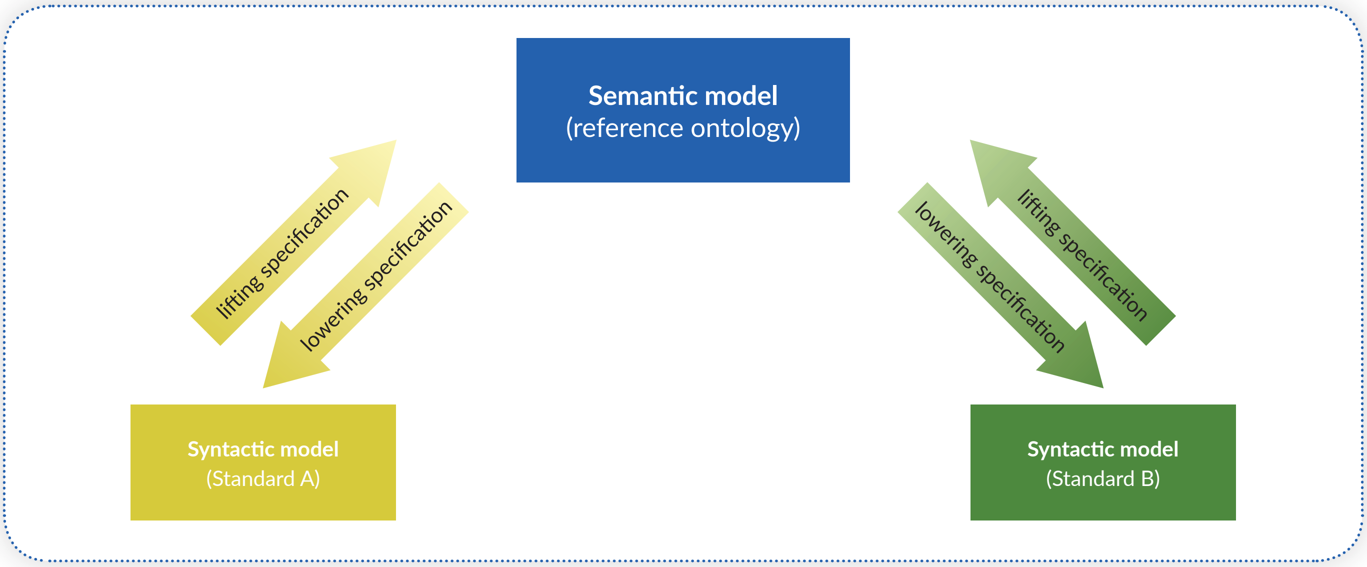

Fig. 1 ST4RT approach to semantic interoperability

CASE STUDY

The Case Study taken into account for the ST4RT demonstrator aims at showing how the semantic approach can solve the problem arising from the existence of many different communication standards, hindering the interoperability of the rail distribution systems, not to mention the sale of intermodal products.

Historically, the sale of tickets by a rail company for the trains of another one was done offline, based on catalogues exchanged once per year. The on-line reservation was introduced firstly in the 70's to allow a customer to make sure to travel seated or on a sleeping accommodation. More recently the system was extended to also sell the so-called IRT (Integrated Reservation Tickets), products including both the transport title and the reservation.

For those sales European railways have always used a standard created by the UIC (International Union of Railways) and called 918 as it is described in the UIC leaflet 918-1. The 918 standard defines a set of messages to be exchanged between a requesting and an answering reservation system to sell a ticket or a simple reservation, or to cancel it.

There exist two versions of the 918 messages, an older one in binary format (still used by almost all European railways) and an XML one. The first set of messages is the one adopted by the TAP TSI Regulation, in its Technical Document B.5. The XML set of messages could be included in B.5 in a next revision of TAP TSI, and is therefore called in the ST4RT project "TAP TSI 918 XML".

In the last years the sector, including Rail Undertakings (RU) and ticket vendors, have created a new standard called FSM (Full Service Model), with a more complete and flexible approach than 918. The latter works basically with "one request – one answer". There is one RU requesting the reservation and one RU allocating the place where the passenger will travel. It only covers the reservation process, leaving aside the steps before (journey planning) and those afterwards (payment, revenue apportionment).

FSM, on the contrary, supports a variety of business models, including intermodality, and can be used by a wide number of applications, overcoming and improving the current limitations in railway distribution. The problem is that FSM has a completely different semantics from 918, but their interconnection is essential to allow some railways/ticket vendors to start using FSM without losing the ability of exchanging reservations with those still using 918. This is where ST4RT comes to the rescue.

As the goal of the demonstrator is to show the feasibility of a conversion through the ontology-based approach, and not of course to develop a translator able to convert all possible exchanges of messages, we will concentrate on a specific use case. We namely imagine that a customer goes to a French rail station and requests a reservation to travel on a sleeping car in an Italian train. This use case was agreed as it involves business logics that are similar in the two standards, and allows the conversion of one FSM request message into one 918 request, and the conversion of the corresponding 918 answer into one FSM answer.

Once the ST4RT demo will have proven the feasibility of the ontological conversion, it will be possible to study in a next step more complex use cases, where one message in one standard translates into different messages in another one.

ANNOTATION PROCESS

The overarching idea of the solution proposed within the ST4RT project to the interoperability problem outlined in Section 2 is to annotate legacy data representations to be able to match the legacy concepts to those of the reference ontology, and vice-versa. In the ST4RT approach, the assumption is that legacy data representations and the corresponding data instances are given in terms of standard technologies, such as XSD (XML Schema Definition) and XML (eXtensible Markup Language).

The starting point of the approach is that data schema definitions given in terms of XSD files are first translated into declarations in the Java programming language. Then, the declarations of Java classes, methods and attributes are annotated to match them with concepts in the reference ontology. The annotations defined in the ST4RT approach are compatible with (and extend) those defined in the Pinto framework (https://github.com/stardog-union/pinto). For example, Java classes can be mapped to entities in the reference ontology using annotation @RdfsClass, as in the following example, which maps the ContactInformation concept from the FSM standard to the Passenger class of the ontology.

```
@RdfsClass(“customer:Passenger”)
public class ContactInformation extends FSMID
```

It can happen that the values to be created during the lifting process must be built by piecing together information from the incoming message, but also from the ontology. This is achieved through queries in the SPARQL language, to be performed on the ontology. The converter can be instructed to execute SPARQL queries through the @Sparql annotation, which is exemplified in Figure 2. More precisely, the @Sparql annotation shown in Figure 2 is applied to attribute country; its result is mapped to the object of an RDF triple whose subject and property are, respectively, an instance of customer:Passenger (since ContactInformation is the enclosing class) and st4rt:isInCountry_Alpha2.

```
@RdfsClass(
 “customer:Passenger”
)
public class ContactInformation
  extends FSMID {
[…]

@Sparql(
  name = “getAlpha2ByName”,
  inputs = {“country”}
)
@RdfProperty(
  “st4rt:isInCountry_Alpha2”
)
@XmlElement(name = “Country”)
protected String country;

[…]
}
```

```
PREFIX countries: <http://www.sample-country-ontology.org/
Countries#>

SELECT ?alpha2
  WHERE {
    ?subject countries:countryCodeISO3166Alpha2
    ?alpha2;
    countries:countryNameISO3166Short ??input1.
}
```

Fig. 2 Lifting @Sparql example and corresponding SPARQL query

The @Sparql annotation indicates that the value must be fetched through the getAlpha2ByName SPARQL query, whose definition is shown in the right-hand side of Figure 2.

CONVERSION PROCESS

In the following we briefly describe the process followed by a ST4RT converter to realise the mechanisms described in section 2. More precisely Figure 3 shows a UML Sequence Diagram (SD for short) depicting the general steps performed by a ST4RT converter during the transformation process.

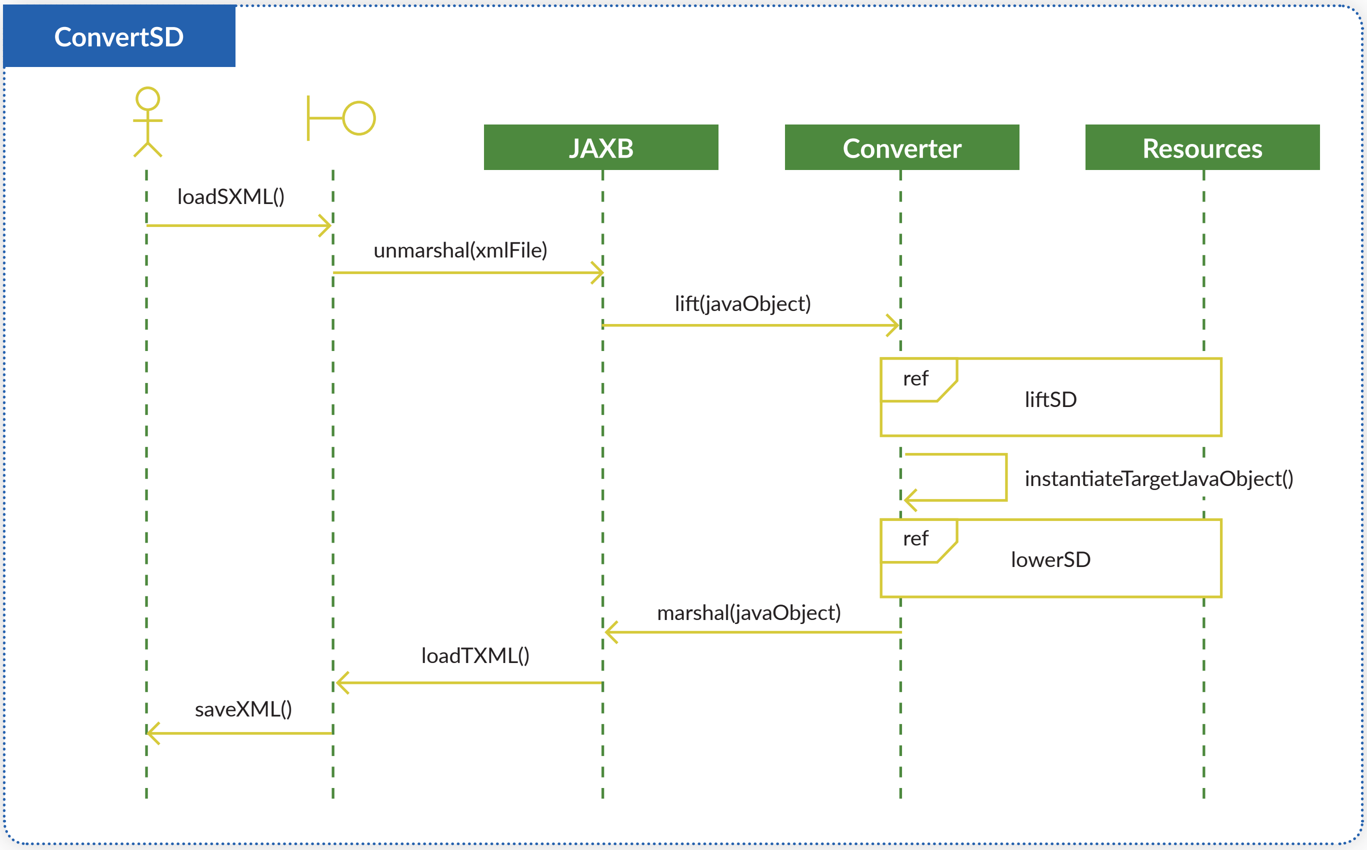

Fig. 3 General conversion mechanism

As the figure above shows, first the XML file conforming to the source data representation (e.g., FSM) is loaded, and it is used by the JAXB framework to instantiate a corresponding Java object. Then, the ST4RT converter uses the annotations defined for the source standard, and additional resources such as repositories of RDF triples on which to run SPARQL queries, to lift the Java object to the integrated ontology. After the lifting process has been completed, a new Java object, capturing the concepts of the target standard (e.g., TAP-TSI 918 XML) is instantiated, and it is filled out with the relevant information during the lowering phase (again, by exploiting resources such as SPARQL endpoints). Finally, the Java object is used by the JAXB framework to create an XML file that conforms to the target standard.

Figure 4 depicts the details of Sequence Diagram liftSD (Sequence Diagram lowerSD is similar, and is not shown here for the sake of brevity).

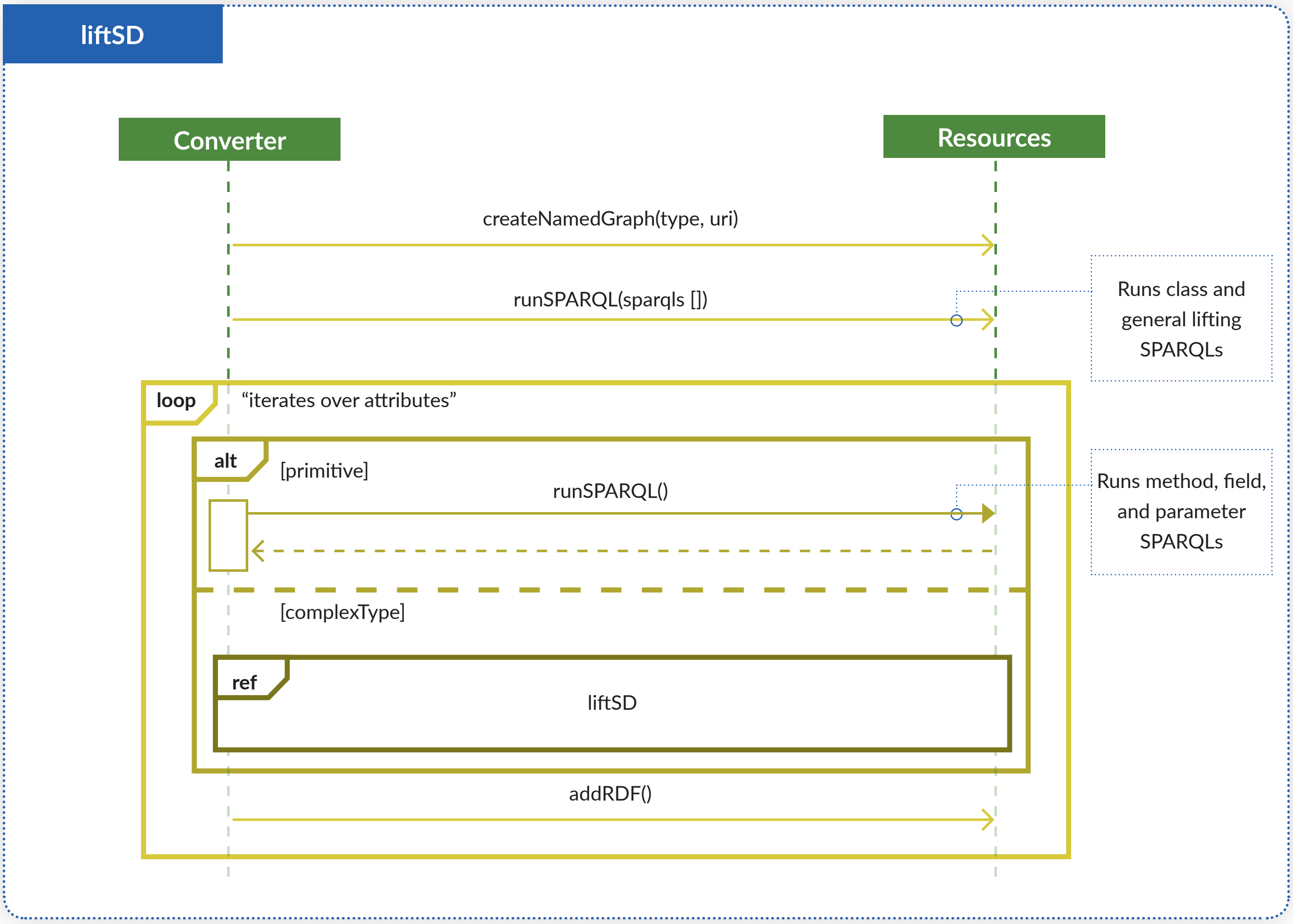

Fig. 4 Detail of the liftSD Sequence Diagram

As Figure 4 shows, during the lifting process the converter analyses the @Sparql annotations of the source standard, and runs them on a suitable resource (i.e., an RDF triple store). If the element to be lifted is a complex one, this requires itself a lifting through the same mechanisms. After all necessary elements have been lifted, the generated RDF triples are added to the knowledge base.

DEMO

The final part of the project will be a practical demo, where a Hardware/Software platform will be made operational for a real-life test of the concepts developed in the previous tasks. To this scope the project group will have to:

- define in detail the use case(s) mentioned in above chapter 3,
- define the test data, in the form of sample bookings,
- identify in detail the HW/SW environment where to run the demo and how to adapt the available one if needed,
- identify the participants to the demo and their respective roles,
- define the KPIs and metrics needed to measure the results of the demo,
- provide and setup the run time implementation environment and reporting tools,
- run the demo,
- generate and evaluate the results of the demo,
- produce the final report.

The demo will adopt as basis an existing solution, designed by Hit Rail and developed by OLTIS, which allows the conversion of binary 918 messages into XML ones and back. This converter is currently used for the interconnection of innovative railways already using 918 XML with those still using the older format. The ST4RT demo will take advantage of the middleware of the existing converter (management of input/output queues, protocols, logging, …), while of course the application software will be reviewed and the entire "ontological section" will be added (lifting/lowering of messages, triple store, SPARQL queries, …).

The demo will also reuse a separate tool, called Boomerang, created by Hit Rail and OLTIS to support the railways in developing/maintaining their reservation systems. Boomerang simulates a reservation system and allows a user to send a reservation message to it and receive back a plausible answer as if it was sent by a real railway. Boomerang will need to be enhanced so that it can be used to close the loop in ST4RT between the lowered request and the answer to be lifted.

Obviously, the demo will have to be run repeatedly, because it is likely, as in any project, to detect bugs in the work done in the preliminary tasks. It is also possible that, despite the absence of real bugs, the process defined on paper in the previous phases (with lifting/lowering, annotations, triples and queries), when put in operation produces response times that are considered unacceptable. In this case the group would try to adopt corrective measures to improve the performances, e.g. by an optimal mix of processing data in central memory vs accessing external repositories.

At the end of the project, the group intends to demonstrate a concrete application of the concept of "translation through ontology", showing the realistic integration of the semantic approach defined in the previous parts of project with an interoperability converter originated by existing systems on the European market.

LINKS WITH S2R

ST4RT is one of the 1st Open Call projects of Shift2Rail and as shown in the figure above, it is also going to build on the work that has been carried out in the IT2Rail 'lighthouse' project and will also feed future Shift2Rail projects. Many partners of the ST4RT project are also involved in IT2Rail, which has eased the knowledge transfer needed for such collaborative projects.

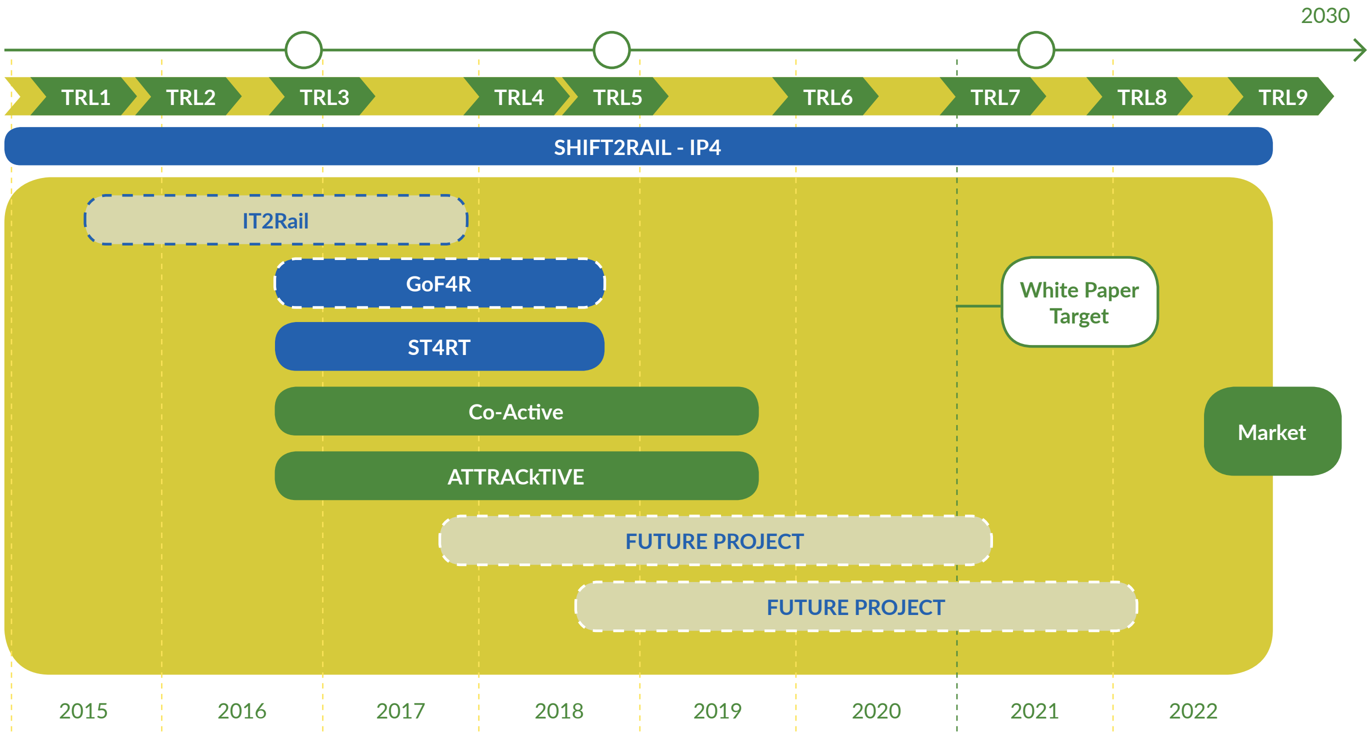

Fig. 5 Links with Shift2Rail's 4th Innovation Programme

PROJECT STRUCTURE, FACTS & FIGURES, PROJECT PARTNERS

The diagram below shows how the project is structured into the five different Work Packages.

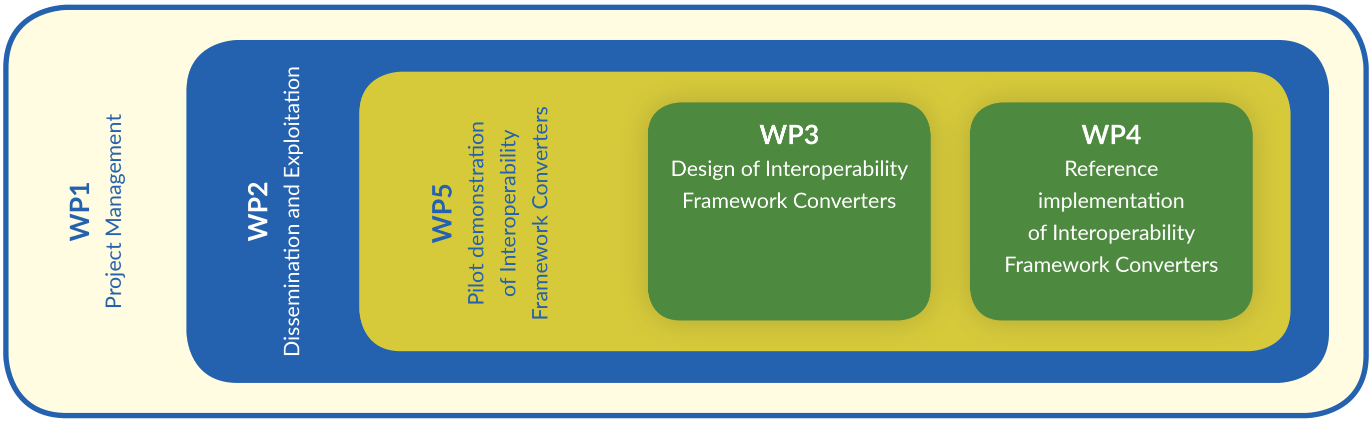

Fig. 6 Project structure of ST4RT

WP3 will produce the conceptual design of the converter-generation procedure, whilst WP4 will study the best approach to implementation of the design, including a selection of tools and frameworks, and provide a specific implementation of the design with the support of established software developer partners active in the provision of solutions and services to the Rail Industry. Towards the second half of the project, WP5 will analyse the selected target use case, develop the test cases and performance measures, and implement the actual demonstration scenario providing quantitative analysis of the results.

FACTS AND FIGURES

Total Budget
€1 million
(100% EU funded)

Duration
24 months

Number of Partners
8

Grant Agreement No
730842

Project Start Date
1st November 2016

Project End Date
31st October 2018

PROJECT COORDINATOR

unife
THE EUROPEAN RAIL INDUSTRY

TECHNICAL LEADER

TRENITALIA
GRUPPO FERROVIE DELLO STATO

PROJECT PARTNERS

Cefriel

RINA CONSULTING

hitrail
european railway IT & data communications

oltis group

POLITECNICO
MILANO 1863

uic
INTERNATIONAL UNION OF RAILWAYS