NEWSLETTER

DECEMBER 2020

CONTENT

Foreword 02

Project scope & structure 03

Requirements and alignment with Shift2Rail IP4 04

Performance and scalability of the IF 04

Semantic automation for service integration 06

Proof-of-concepts implementation and validation 07

Facts & Figures 08

FOREWORD / INTRODUCTION

In this 2nd and final newsletter of the **SPRINT** (**S**emantics for **P**erfo**R**mant and scalable **IN**teroperability of multimodal **T**ransport) project, you will learn about what has been achieved until the conclusion of the technical developments.

To find out more about SPRINT, to get a closer look at the project partners and to access the public deliverables that are already available, please visit our website at:

www.sprint-transport.eu

Enjoy the read!

PROJECT SCOPE & STRUCTURE

The Shift2Rail Interoperability Framework (IF) aims to facilitate multimodal travel in a highly diverse environment and with many transport modes. It is intended as both as a data integration/conversion layer and as a data and service sharing ecosystem. As a data integration/conversion layer, it provides 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. As a data and service sharing ecosystem, it allows Travel Service Providers to advertise their services and to describe the meaning of the carried information using Semantic Web technologies.

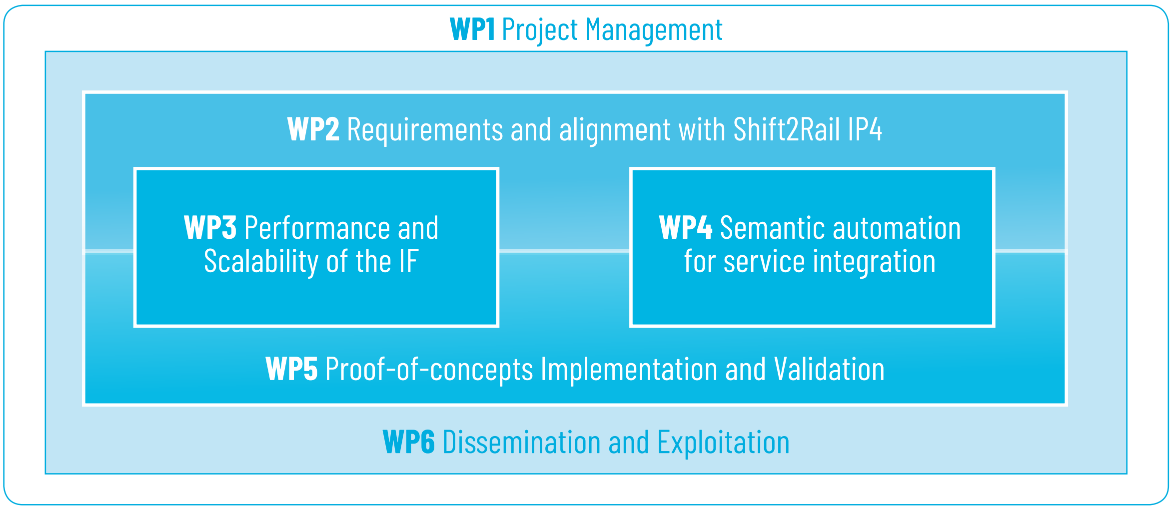
Building upon the work already initiated by previous Shift2Rail projects, the SPRINT project pursued the following objectives:

- Define a reference architecture for the Interoperability Framework, taking into account recent advances in the design and development of distributed systems, and in particular of cloud-based ones;
- Define techniques facilitating, in particular by increasing their level of automation, activities that are central to the concept of IF, such as the collaborative creation and management of ontologies and of semantic-based mappings between heterogeneous data representations;
- Define how Semantic Web technologies can be used to streamline data conversion processes, thus improving time to market and easing adoption;
- Provide an Asset Management tool to let Travel Service Providers contribute to the creation of an interoperability-oriented sharing ecosystem according to the principles of openness and sovereignty;
- Demonstrate the proposed improvements to the IF through a proof-of-concept implementation reaching at least TRL 4.

Moreover, SPRINT contributes to the realisation of the IF by analysing which technologies and architectures allow improving its performances and scalability, aiming for an EU-wide deployment of the solution.

The way in which the project is structured is shown in the diagram below:

FIGURE 1
SPRINT project structure



REQUIREMENTS AND ALIGNMENT WITH SHIFT2RAIL IP4

One of SPRINT's goals is the definition of requirements for the IF architecture design and alignement of the IF architecture with S2R IP4. Besides this, SPRINT accepted the challenge to understand the needs and requirements of National Access Points (NAPs) and the IF architecture's compatability and complementarity with NAP architecture and data-exchange models.

The ambition of SPRINT is to contribute towards the realisation of a Single European Railway Area (SERA) and towards the development of S2R IP4. To do this, SPRINT partners analysed related S2R IP4 projects results and ongoing projects, as well as EU initiatives and projects related to the IF concept, particularly those linked with the public transport sector and semantic technologies.

The first report that the project partners published was the analysis of relevant initiatives and projects, high-level requirements and specific requirements to the IF architecture. Based on the outcomes of this high-level requirements analysis, the report Requirements for an IF architectural design aligned with S2R IP4 and other initiatives was delivered. The procedure for the identification of the main requirements for the IF included an analysis of the main concerns, challenges and contributions of these projects. The produced requirements are represented by three viewpoints: data management, service management, and system management. The final report on Requirements for an IF architectural design (F-Rel) provides the final list of the requirements collected from the active Shift2Rail IP4 projects and from the analysis of the NAPs.

Besides the recommendations regaridng the IF architecture, SPRINT aimed to produce recommendations to the IP4 ecosystem. The intermediate version of the recommendations was delivered in the summer of 2020. At the final stage of the project's developments, SPRINT partners will update the recommendations list for S2R IP4 future research opportunities based on the outcomes of ongoing work. The report will be delivered in the end of the project.

PERFORMANCE AND SCALABILITY OF THE IF

The main goal of the SPRINT Work Package 3 is to revise and consolidate the architecture of the IF, to make it suitable deployment in real-life scenarios. This overarching goal is broken down into three sub-goals, which can be summarised as follows:

- First, the technical requirements of the IF need to be identified, in particular those concerning the performance and scalability features of the framework, which need to be fulfilled to ensure that the IF can handle the complexity of the European transportation domain.
- Second, a reference architecture, which identifies the main elements of the IF and their relationships, must be defined.
- Third, the means to verify that an implementation of the defined architecture actually meets the stated requirements must be defined; in particular, the infrastructure to test the IF against its performance and scalability requirements must be defined, together with the relevant test cases.

The work of WP3 has been completed during the second year of the project, with the release of the final versions (F-REL) of the artifacts mentioned above. In particular, the main achievements have been the following:

- It has consolidated the set of performance and scalability requirements for the IF; in particular, it has revised the requirements introduced in the first year of the project, and it has added new requirements.
- It has produced the final proposed IF architecture, which expanded on the C-REL version by detailing the relevant features of the components of the frameworks.
- It has defined the final version of the testing infrastructure for the architecture, by defining the general testing method to be applied, and the test cases for each relevant component of the IF.

The figure below shows the consolidated version of the IF architecture described through a UML Component Diagram. The core of the IF is the Asset Manager, which includes components dealing with issues such as Asset registration, discovery, and lifecycle management. The IF is designed to interact with external services to handle user authentication and authorisation. The IF relies on asset meta-data (created at the moment of asset registration) for various tasks such as asset discovery; asset meta-data is stored in a triple store, which is another core element of the IF.

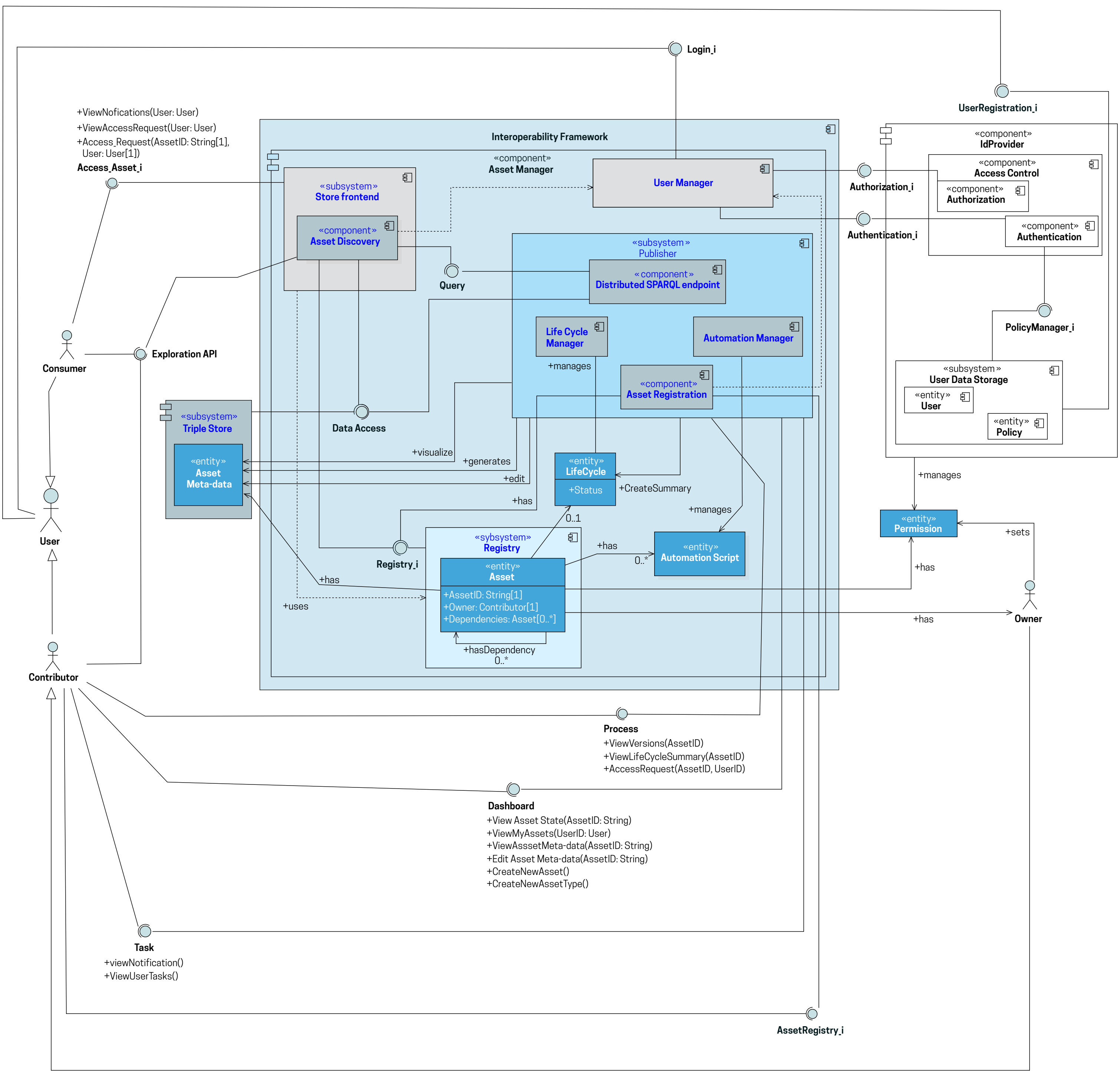


FIGURE 2
SPRINT IF architecture

SEMANTIC AUTOMATION FOR SERVICE INTEGRATION

The SPRINT IF supports by design a wide array of automation solutions, and can used as a starting point to implement an IP4 ecosystem. To that extent, the Shift2Rail IP4 ecosystem (as shown in Figure 3) is just one of the possible interoperability solutions which can be implemented using the IF.

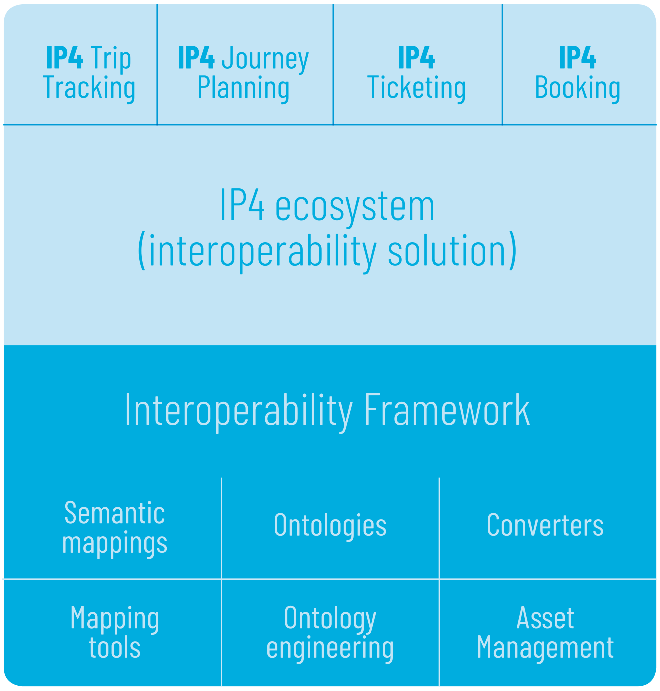


FIGURE 3
Interoperability Framework and IP4 ecosystem

The establishment of an IP4 ecosystem encompasses several different activities which must be taken care of. Such activities stem from the creation of a common model to the development of the integration solutions to the usage of sharing platforms to raise awareness and ease the adoption. A high-level list of activities is reported here:

1. Develop the ontology;
2. Analyse existing ontologies;
3. Develop mappings;
4. Develop converters;
5. Develop resolvers;
6. Publish artifacts;
7. Gain access to existing artifacts;
8. Perform tasks after the successful publication of artifacts.

Automating parts of this long list of activities can play a key role in establishing an efficient ecosystem, lowering the maintenance costs and helping govern a complex distributed system ran by different transport operators. Some of those activities can be fully automated, while others require human intervention.

In the ontology development and maintainance topic, we applied automation to two different issues: starting the development of a new ontology according to pre-existing knowledge in the form of XML Schemas, and easing the collaborative ontology engineering effort. Standards for interoperability are decades old, and during the creation of each standard a specific aspect of the transportation domain has been modeled. Performing a conversion of the XML Schemas modeling messages of a standard to their semantic representation in some cases can therefore lead to a faster ontology development. The second application of automation was related to ontology engineering, which is a very complex and time consuming activity. Several people can modify different parts of the same ontology, and inconsistencies must be avoided. By combining version control techniques and automatic generation of documentation, ontology quality reports and ontology diagrams, we ensure that all the developers involved in the task are able to keep track of the modifications to the ontology, and are able to receive proper information about the impacts of the proposed changes.

Interoperability based on semantic techniques means analysing both ends of a communication channel, extracting their conceptual models and then finding correspondances and similarities. Such process requires a very strong understanding of both the originating system and the destination one. We therefore decided to offer a level of automation to ease such task, and we created a mapping suggester which can, as its name suggests, provide hints to mappings developers about possible similarities between concepts and relations belonging to different conceptual models.

An EU-wide ecosystem fostering interoperability in the transportation domain requires enforcing a strong governance policy. Such governance can involve organisational aspects as well as technical aspects. Publishing information about a service, asking the permission to use such a service, or convert datasets from the original format to another one mandated by the EU regulations, are all tasks which can be automated. By analysing the initial requirements for the establishment of an IF governance provided by the GoF4R project, we realised that governance involves the coordination of people belonging to several organisations, and services which can be provided by the IF itself or by external and trusted parties. We decided to leverage on an existing and well-tested solutions for modeling such governance aspects, and we combined BPMN processes (which can orchestrate both humans and services) and Continuous Integration/Deployment tools to deal with low-level technical aspects. The resulting tool (Asset Manager, which will be described in the next section) can therefore provide a basis for the development of a sound ecosystem, allowing the publication of several types of digital assets according to the devised governance model.

PROOF-OF-CONCEPTS IMPLEMENTATION AND VALIDATION

The main objectives of this part of the SPRINT project were the implementation of the proof-of-concepts designed in WP3 (Performance and Scalability of the IF) and WP4 (Semantic Automation for Service Integration), and the setup of the technical environment defined in WP3, to test its performance and scalability. Several use case scenarios were also defined to validate the technological solutions identified in the project.

As SPRINT focused on supporting the establishment of an interoperability-focused ecosystem, we implemented tools and framework at various technical levels.

Ontology development

Ontology development can be roughly divided it into two main branches: creating a new ontology from scratch and converting an already existing data model using the W3C Semantic Web stack. The former activity requires a fully human activity since it involves understanding a domain and its rules and representing them as a set of logical axioms. Automation can support the human activity, easing collaborative editing and providing up-to-date graphical representations of the ontology. Those two aids are especially useful when the domain which is being modelled is vast, and when the team is actively working at different aspects at the same time. SPRINT is helping to address this issue by providing **Ontoology** (https://ontoology.linkeddata.es/), a tool providing collaborative ontology editing, automatic ontology documentation and automatic ontology validation.

Full automation is not possible even in the latter branch of the ontology development, namely the conversion of an existing model into an ontology. In this case, the conceptualisation of the domain has already been performed by someone else, but it has been serialised into a non-ontological format. XML or RDB schemas, UML diagrams, are all examples of such non-ontological formats. The aim of the ontology development activity, in this case, is to obtain a clean model, removing attributes and relations which are usually introduced by the specific format, while at the same time staying very close to the original model to keep compatibility. The role of automation, in this case, is to provide a first rough draft of the ontology, which can be used by ontology designers to speed up the development process. To this extent, SPRINT provides an XSD2OWL tool which is able to create raw ontologies directly from complex XML Schemas to be used as a basis for further ontology development.

Streamlining the conversion process

Integrating different systems requires overcoming multiple technical difficulties, and the integration process requires analysing several aspects, like:

- whether the two systems use the same communication approach (pull vs. push);
- whether the two systems are stateful or stateless;
- whether the two systems use compatible processes;
- how much information from a source system can be sent to the destination system.

The SPRINT project demonstrated that some aspects of the integration process can be streamlined using a combination of semantic techniques and already existing open source solutions. We implemented **Chimera** (https://github.com/cefriel/chimera), an Open Source solution for semantic-based data conversion and service mediation which leverages on the Apache Camel integration framework. With such solution we were able to break down complex integration processes into conversion pipelines composed by simpler operations. At its core, we first transform the input message into RDF according to a reference ontology using RML declarative mappings or ST4RT Java annotations, and we then extract knowledge from the resulting RDF graph to populate the output message using Apache Velocity templates or again ST4RT Java annotations. A lot of effort has been put to improve the performances and the scalability of our solution to make it usable in real-world scenarios and improve industrial adoption.

While the overall conversion process can be streamlined and automated, finding correspondences between model is a time-consuming and knowledge-demanding effort. To ease such tasks SPRINT provides a **Mapping suggester**, which can propose possible mapping candidates and automatically generate ST4RT Java annotations or RML mappings.

Governing an ecosystem

As introduced before, the IP4 ecosystem is an interoperability solution which is being built for the transportation domain using the tools of the IF. The SPRINT project provides a tool, named **"Asset Manager"**, to efficiently address ecosystem maintenance. With the term "ecosystem maintenance" we mean all the activities which contribute to keeping a distributed system running with minimal inconsistencies and downtimes. Such activities encompass the high-level governance of the ecosystem, dealing with rules and contracts between companies, and the constant checking that the information contained in the asset descriptions is up to date and does not lead to an inconsistent distributed system. The Asset Manager can help automating all the technical parts of the governance process since it represents a single source of knowledge for all the components of the ecosystem. It can, therefore, ensure that a major release of service is notified in time to all its users, it can be used to automatically execute ontology-based tests to ensure that changes do not break any existing Converters, and it can be used to automatically convert datasets into other formats (like the Transmodel-based standards required by NAPs).

FACTS AND FIGURES

Total Budget 1 999 500 100% EU funded

8 Partners

Project Start Date 1st December 2018

Project End Date 28th February 2021

Duration 27 Months

Grant Agreement No: 826172

PARTNERS

unife THE EUROPEAN RAIL INDUSTRY

Cefriel

oltis group

POLITÉCNICA

POLITECNICO MILANO 1863

RINA

TRENITALIA GRUPPO FERROVIE DELLO STATO

UITP

Shift2Rail

This project has received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 826172.