ATMO: Autonomous Train Map Ontology*

Nadia Chouchani[a] and Sana Debbech[b]
IRT Railenium, Technopole Transalley 180, rue Joseph-Louis Lagrange, 59308 Valenciennes Cedex, France

Keywords: Ontology Engineering, Modularization, Model-Based Engineering, Conceptual Modelling, Autonomous Train

Abstract: Infrastructure is the backbone of the railway industry which is an extensive and particularly complex system. However, the usage of different national and international standards for its design has led to a large number of incompatible systems. In this paper, we present *ATMO*, ***A**utonomous **T**rain **M**ap **O**ntology*, a modular ontology for modelling an on-board digital map for autonomous trains representing the railway infrastructure, line-side signalling and associated buildings. Semantic web technologies, knowledge and ontology engineering are adopted to integrate information from heterogenous sources and diverse standards such as *RailSystemModel*, *EULYNX* and *IFC Rail* ; to ensure semantic consistency of digital map objects. The development process is based on *METHONTOLOGY* and produced : (*i*) *UML* model as lightweight ontology; and (*ii*) *OWL* formal machine-readable specification as heavyweight ontology. The latter was evaluated using a railway use case.

1 INTRODUCTION

Physical infrastructures such as railway networks are key elements for passenger and freight transport. Modelling rail infrastructure is a critical process because of the heterogeneity of the information to be modelled such as tracks, signalling and control systems ; as well as safety regulations, engineering conventions and design rules to be respected. As part of the autonomous train project, we are working on the modelling of the railway digital map ; based on different standards. It includes all information on physical tracks, signalling assets and even buildings such as tunnels and bridges, eventually in *3D* representation. Data integration and interoperability are complex challenges for this task consisting on modelling a complete and extensive digital map ; due to the heterogeneous nature of data and underlying standards. To overcome this problem, we propose to apply semantic data modelling techniques to allow integration of heterogeneous information and make consistency of mapping elements. In this paper, we adopt recent researches in semantic web, ontology engineering and information architecture to develop ***A**utonomous **T**rain **M**ap **O**ntology*, *ATMO*, a modular ontology for autonomous train on-board map. As developing ontologies is not an easy task, it is compulsory to restrict the studied domain knowledge (Gruber, 1993). Our proposed solution represents a railway map ontology, which will describe the concepts and relations related to infrastructure, signalling and *BIM* (*Building Information Modelling*) (BSI, 2022). Its usefulness consists in ensuring interoperability, defining common domain knowledge and its sharing. *ATMO* federates knowledge from different source models or railway standards such as RailSystemModel, *RSM* (UIC, 2022) and IFCRail (BSI, 2022) that are initially developed separately and use different terminology of railway system components. In this sense, *ATMO* will be exploited by multidisciplinary stakeholders and different subsystems of autonomous trains such as environment monitoring during the description of the occurrence of train stations, tunnels, etc. ; and the *ATO-OB*, *A*utomatic *T*rain *O*peration *O*n-*B*oard, to identify the topological description of the railway network.

Competency questions of *ATMO* are raised to define which precise kind of information this semantic model will provide. Here the adopted methodology is *METHONTOLOGY* (Fernandez-Lopez et al., 1997) whose choice will be justified in the following section. It has resulted in, first, lightweight ontology represented using a *UML*[1] conceptual model, and second, heavyweight ontology as machine-readable specification obtained by applying a *Model Based En-*

[a] https://orcid.org/0000-0002-2660-1352
[b] https://orcid.org/0000-0002-4003-6505

*This research work is funded by the French program "Investissements d'Avenir" and is part of the French collaborative project TASV (Train Autonome Service Voyageurs), with Railenium, SNCF, Alstom Crespin, Thales, Bosch, and Spirops.

[1] https://www.omg.org/spec/UML/2.0

Chouchani, N. and Debbech, S.
ATMO: Autonomous Train Map Ontology.
DOI: 10.5220/0011893200003402
In *Proceedings of the 11th International Conference on Model-Based Software and Systems Engineering (MODELSWARD 2023)*, pages 283-290
ISBN: 978-989-758-633-0; ISSN: 2184-4348

gineering transformation from *UML* to *OWL*[2]. ATMO is developed to fulfill the following Research Goal (RG):
RG. *How to provide a structured and shared view of autonomous train map concepts and their relations in order to deal with the semantic heterogeneity of railway standards?*

The remainder of this paper is organised as follows. In section *2*, we present a background on the existing data models, standards and related work. In section *3*, we detail the methodological framework of the ontology development. In section *4*, we discuss the obtained results. Finally, we conclude and describe future works.

2 RELATED WORK AND MOTIVATIONS

An ontology is defined as an explicit formal specification of a conceptualisation model which is an abstraction of a domain of interest (Gruber, 1993). It was exploited in many domains specially for Safety Critical Systems (SCS) (Uschold and King, 1995). Ontologies have powerful capabilities to represent knowledge from different domains and ensure their federation and semantic clarification. Reasoning using the machine-readable version allows an efficient communication between stakeholders involved in the development process of SCS. Furthermore, the implemented ontology facilitates the decision-making process by data retrieval and ensures traceability of safety decisions and design choices. However, the ontology development process is costly and time consuming. In order to reduce its complexity, the concept of modular ontologies has been introduced (D'Aquin et al., 2009). Being an incremental process, modularization offers a scalable and efficient development. This process is the composition of an ontology into smaller modules. The latter can then be partially reused, which improves the management of the complexity of the ontology and the associated reasoning (Pathak et al., 2009). An ontological module is therefore defined as a reusable component of a larger and more complex ontology. To be reused, this component must be autonomous, i.e logically coherent and independent of the other modules (Pathak et al., 2009). Nevertheless, it must be connected to it without being isolated or dis-joined to form a single coherent ontology. Moreover, its deletion does not affect the ontology as a whole. As a description of the railway infrastructure can contain several aspects and views of the system such as signalling, track construction or the catenary network, modularity seems to be a suitable solution for the development of ATMO. Indeed, each module would be devoted to a view of the infrastructure digital map. This ensures semantic cohenrencies when ansewring the multidisciplinary users and needs of the digital map.

Several methodologies for ontologies development, which vary according to the context of use of the ontology and the target users, have been proposed. A methodology is defined in *IEEE std.730.1* as "a comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought-intensive work ought to be performed" (IEEE, 1996).

These methodologies have been classified into four groups according to their characteristics and purpose (Debbech, 2019). Early methods such as *ENTERPRISE* (Uschold and King, 1995) assumes a single, non-iterative design process. The second category includes iterative methods that emphasise initial formal specification and advocate the reuse of existing models and ontologies, such as *METHONTOLOGY* (Fernandez-Lopez et al., 1997). A third class includes post-semantic web methods such as *SABiO* (de Almeida Falbo et al., 1998) a systematic approach that emphasises collaboration and flexibility. And finally ontological learning methods which are based on the use of automated or semi-automated tools to reconfigure knowledge.

In our case, we are based on existing standards for modelling the railway infrastructure. This motivates our choice of *METHONTOLGY* which allows this reuse and adaptation. Moreover, it is the only one which specifies in its steps how the ontology will be populated.

Recently, semantic data models have been proposed allowing data context storing in a machine-readable format (Berners-Lee et al., 2001). In this context, some ontologies for the railway domain are proposed. *InteGRail*, a project led by the *European Union's (EU) 7th Framework Program (FP7)* designed a standard approach for architecture and communication (InteGRail, 2022). One developed ontology (*RDO*) was intended to create an opportunity for improved performance. Another research work proposed a domain ontology for railways which aims to integrate information from heterogeneous sources (Tutcher et al., 2017). Based on semantic web, it presents a proof-of-concept real time passenger information system.

Regarding railway infrastructure, there are some modelling standards. RailSystemModel (*RSM*) is a conceptual model which is partly devoted to mod-

[2]https://www.w3.org/TR/owl2-syntax/

elling this part of the rail system (UIC, 2022). Based on *ISO 19148* for *Linear Referencing*[3], *RSM* provides a description of the railway infrastructure based on the topology of the track. The latter is represented by objects in the "*Topology*" package which are the carriers of other information. This description is based, on the one hand, on an operational breakdown of the network infrastructure in the "*Network*" package and on the positioning of these objects on the earth's geode through association with concepts from the "*PositioningSystem*" package. The information attached to the topology can be geometric in the context of the "*Location*" package and/or functional in the context of the "*NetEntity*" package. IFC Rail norm attempts to represent the geometry of construction tracks based on the Building Information Modelling (BIM) (BSI, 2022). Finally, the *EULYNX* model includes a large set of objects relating to signalling objects (signals, locks, etc.) and related concepts (routes, needle protection, etc.) (EULYNX, 2022). Its main purpose is to allow interoperability through the exchange of signalling information between infrastructure managers and signalling system providers.

The work presented in this paper is situated at the intersection of several domains. Our motivation is to model a digital map for autonomous train by developing a modular ontology. The modularization allows semantic knowledge to be presented in a structured, formal and expressive model. We noticed that *METHONTOLOGY*, whose steps are shown in the *Figure* 1, is the most suitable methodology for our use case.

In the literature, one common problem in designing semantic models for railway is the lack of use of international standards. Therefore, our choice of knowledge sources is based on *RSM*, *EULYNX* and *IFC Rail*. To the best of our knowledge, our work is the first to propose such a mapping between standards for semantic modelling of on-board railway digital map.

3 METHODOLOGICAL FRAMEWORK

In this work, we have adopted the ontological modularity for modelling the on-board railway map. The idea is to define individual modules, as shown in *figure* 2 which are then assembled in the same modular ontology *ATMO*. The latter allows reasoning on knowledge of the field of railway digital map. Its development requires to acquire knowledge from experts in various fields and reuse existing resources and finally to compose the first developed modules. Each module has been developed respecting the methodology *METHONTOLOGY*.

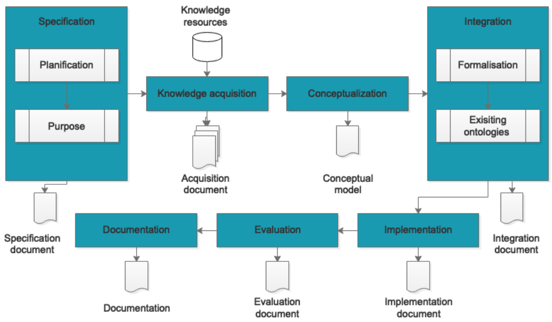

Figure 1: The methodology *METHONTOLOGY*: activities and steps.

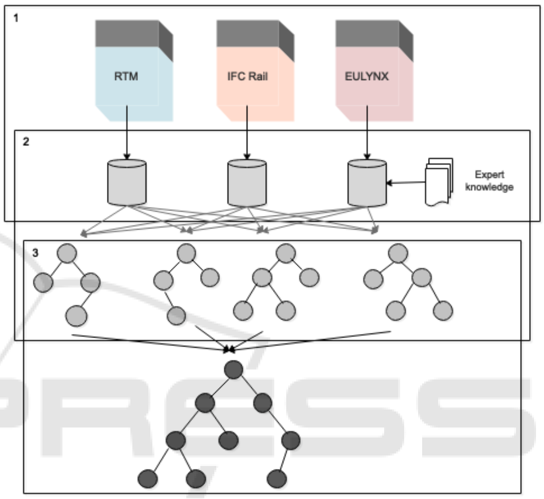

Figure 2: Overall approach of *ATMO* development: (1) knowledge extraction; (2) ontology modules construction; and (3) final ontology composition.

In the following, we present the general framework of the developed modules, from the specification to the evaluation.

3.1 Specification

In this step, we defined high level requirements. The scope being the railway map, this ontology will answer the questions of interoperability of national and international systems and standards as well as the reuse of domain knowledge. This semantic model can be used for the implementation of on-board autonomous train map, and also for the project carried out by the *International Union of Railways* (*UIC*) aiming to creating a global dictionary for the railway domain. Different modules have been defined

[3]https://www.roadotl.eu/static/eurotl-ontologies/iso19148_doc/index-en.html

according to the reuse, the domain and its level of detail; which are: “Railway”, “Track”, “TrackSide” and “Operational”.

The specification of the data model is defined by a set of functional and non-functional requirements derived from the established needs of the implementation of the autonomous train.

3.2 Knowledge Acquisition

Several areas of knowledge are at the heart of this work. This step was carried out by defining Ontology Design Patterns (ODPs). It involves defining all the concepts to be used in the ontology, the relationships between them and also a documentation corresponding to the different concepts. In order to extract knowledge from the domain of the ontology, we used three sources for explicit and implicit acquisitions. First bibliographic research of articles and books was necessary to form a background on the whole field and questions on more specific use cases. Then we collaborated with experts, especially in the signalling field. We had discussions around *EULYNX UML* model to which we had a read access. Finally, the reuse and reengineering of non-ontological resources were applied to the construction of the different models. The analysis of the various cited resources allowed to define a knowledge model that meets the needs covered by the *ATMO* ontology.

The two approaches “top-down” and “re-use” were used to extract knowledge for the ontology. Each module of the latter was created by repeatedly iterating over these two approaches. The first relies on the knowledge of experts to build a model. It comprises the following stages:

1. Determine the concepts within the scope of the ontology after discussing with experts;
2. Decompose the concepts into subcategories around which to create competency questions;
3. Examine the scope of new concepts in order to decide whether they will be implemented or reused according to the re-use approach;
4. Re-engineer the concepts where appropriate.

During this process, the method “top-down” aimed to develop a high quality model for the knowledge of on-board map in order to model it exhaustively and fill the lack of links between the models when reusing existing ones. As for the second approach of “re-use’, knowledge is extracted from existing models. It revolves around the following stages:

1. Identify the concepts to be reused after iterations of the two approaches ;
2. Examine the documentation in order to analyse the semantics of the concepts;
3. Re-engineer the concepts in the model by reusing or extending it from existing models, respecting the following methodology:
   - Refer to *RSM* where useful ;
   - Acknowledge the dual nature of concepts by instantiating several classes;
   - Avoid mutual and strong dependencies;
4. Consider new questions of competency on the new concepts within the scope of the ontology.

3.3 Conceptualisation

During the previous step, the basic concepts and classes specific to the domain are identified. In our work, three main categories of knowledge must be represented in the model, which are the following :

- **Infrastructure:** includes the topological description of the railway network as well as all the geographic data making it possible to geo-locate the train in a 3D representation;
- **Signalling:** involves the recognition of the various signals encountered by the train, more specifically their structures;
- **Building Environment:** describes the occurrence of buildings on the network such as stations, bridges and tunnels, this then enables the environmental monitoring subsystem to be supplied.

Based on these types of knowledge and the specification, a set of informal competency questions expresses the problem solving goals. In the initial *ATMO* prototype, some of the competency questions are listed below :

- How can an autonomous train be geo-located in a *3D* representation ?
- Which are the signals encountered by a train along the route ?
- What is the global environment in the network including tunnels, bridges and stations ?

Answering these questions was carried out by building and extracting knowledge from domain models and experts. The vocabulary and the *ATMO* model modules are mainly based on the elements of *RSM*, *IFC Rail* and *EULYNX*, relying on both their *UML* models and natural language documentation. In fact, the used modelling language of the lightweight ontology is *UML*. This language provides a standard and tool-supported notation. In addition, the sources models are designed in the same language.

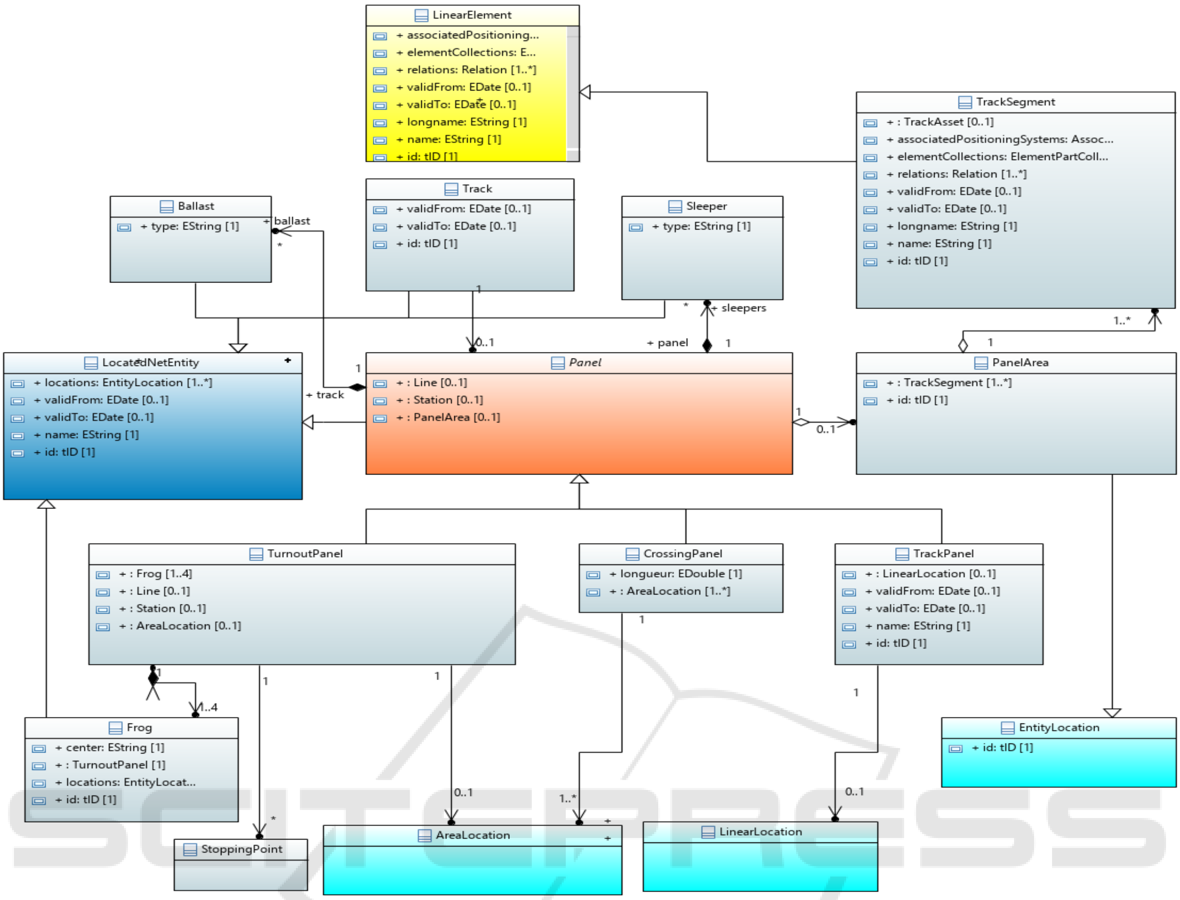

Figure 3: An excerpt of the "Track" *UML* package.

The designed model contains four packages, each one references one module of the *ATMO* ontology. Excerpts from the *UML* packages of "Track" and "TrackSide" are shown, respectively, in *figure* 3 and *figure* 4. We have chosen a modular design from the beginning of the development cycle. The methodology of this design follows a composition approach. The different modules, each corresponding to a dimension of the railway map, are constructed and subsequently composed to constitute the global ontology.

3.4 Integration

The reuse of ontological resources could not be implemented because if there are ontologies in the railway domain, standards like *RSM* and *IFC* are not used. Nevertheless, the research of such resources was carried out by exploiting the ontology search engines as well as bibliographic searches. However, we have reused existing vocabularies and concepts from the standards mentioned above. *Table* 1 provides a documentation of some of the reused concepts.

3.5 Implementation: From Lightweight to Heavyweight Ontology

For mapping the *UML* model to the *ATMO* modular ontology, we have used the *Model-Based Engineering* (*MDE*) approach. Based on the *OMG* classical *Model-Based Architecture*, *MDE* ensures automatic generation of the ontological modules from the *UML* packages. Using these packages, the modules have been generated in *OWL/XML* automatically. *OWL* is the ontology encoding language used to implement *ATMO*. Based on description logic, it is a standard language to represent knowledge in semantic web assuring interoperability. The transformation from *UML* to *OWL/XML* is shown in the *figure* 5.

The transformation process is divided into two steps : first the *ATMO* model is generated and conforms to the *OWL* Metamodel. Second, the ontology is created in OWL/XML format. The ATL rule executing this transformation is detailed in the *figure* 6.

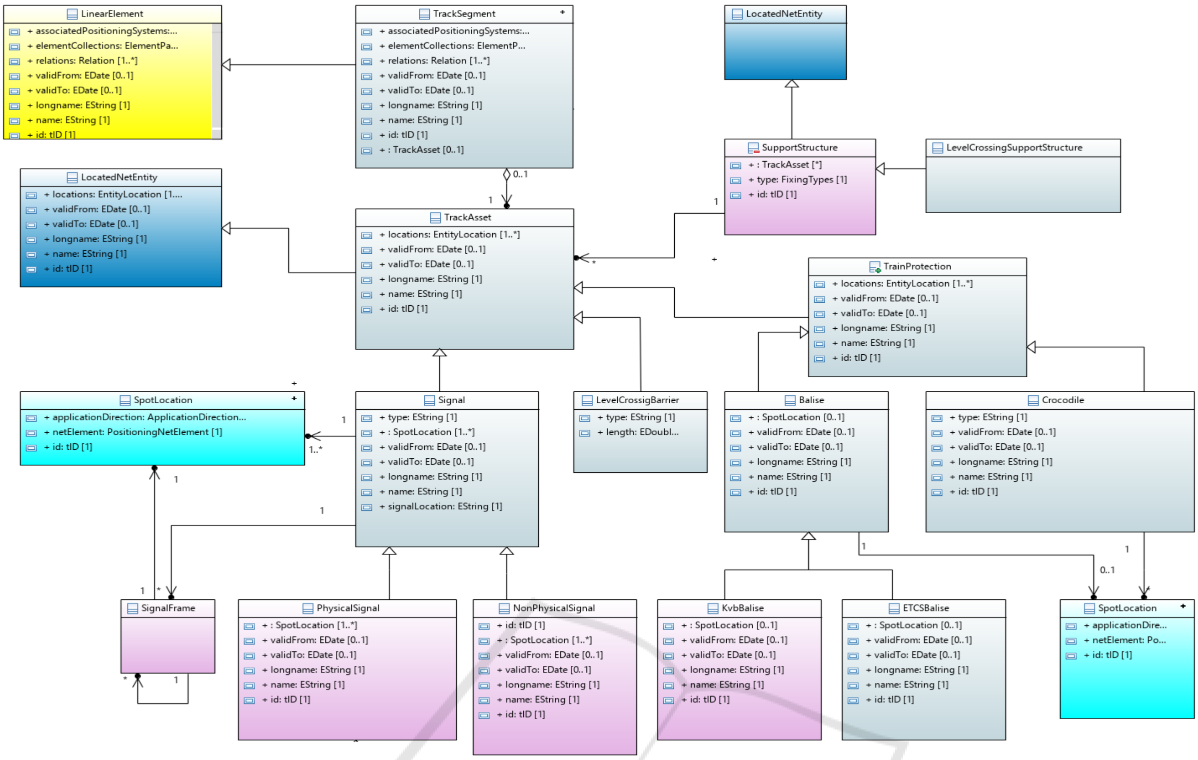

Figure 4: An excerpt of the “TrackSide” *UML* package.

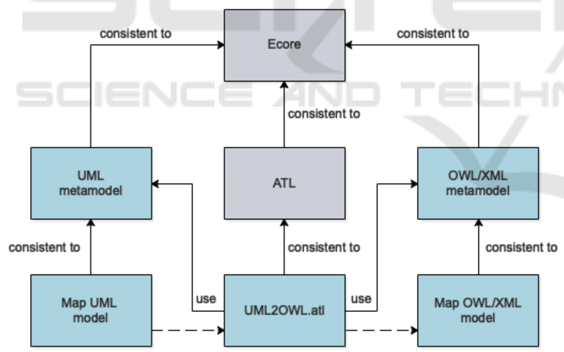

Figure 5: *ATL* transformation from *UML* to *OWL/XML*.

```
rule Package2Ontology {
    from
        p : UML! Package (
            p.oclIsTypeOf(UML! Package)
        )
    to
        o : OWL!OWLOntology (
            uriRef <- u,
            OWLImports <- p.packageImport->collect( e | e.importedPackage ),
            owlGraph <- UML! Model.allInstances()->asSequence()
                    ->any( m | m.oclIsTypeOf(UML! Model)),
            owlUniverse <- p.getPackagedElements()
                    -> select(e | e.oclIsTypeOf(UML!Class) or e.oclIsTypeOf(UML!Property))
        ),
        u : OWL!URIReference ( fragmentIdentifier <- l, uri <- uri ),
        l : OWL!LocalName ( name <- p.name ),
        uri : OWL!UniformResourceIdentifier (
            name <- thisModule.addNamespace(p.name)
        )
}
```

Figure 6: *ATL* rule transformation of *UML* package to ontological module.

3.6 Evaluation

A first evaluation of the developed lightweight ontology was carried out by one of the project partners in relation to the requirements specified in the specifications. *ATMO* heavyweight ontology evaluation was defined in the context of two concepts : verification and validation. From this perspective, we created several instances. During ontology instantiating, we verified that all the needed information required to support the on-board map were represented.

A typical scenario of use for *ATMO* is supporting the net entities positioning. In this scenario, a positioning system supported by the *ATMO* on-board map can geo-locate a signal in the railway network presented in *figure* 7.

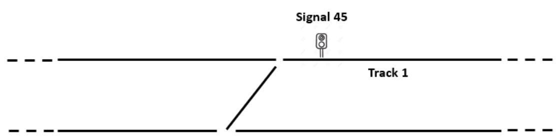

Figure 7: Description of a railway network.

Geographical elements are collected and used to create instances in the ontology. Let us suppose that the system needs to know the position and location of the signal “Signal_45”: individuals instanciating from *ATMO* is shown in Figure 8.

Table 1: Documentation of some of the reused concepts.

Package	Class	Provenance	Description
Track	EntityLocation	*RSM*	The located net entities, such as a signal, have one or more location relations. A location relation has one “EntityLocation”. “SpotLocation”, “LinearLocation” and “AreaLocation” are kind of entity location. They refer to network elements on the topology.
	LocatedNetEntity	*RSM*	It is a kind of “NetEntity” which represents a functional object and is associated to a location.
	TrackPanel	*IFC Rail*	The track is the logical element of a train line. In terms of *RSM*,it is represented as a located net entity. It refers to a linear location of the topology.
TrackSide	Signal	*EULYNX*	The signal is an element of the track that sends a message to the train. It can be physical i.e. a fixed display element or non physical i.e. virtual or fictive one.
	KvbBalise	*EULYNX*	It is an object for train protection. Its position refers to a “SpotLocation” such as a buffer-stop or point.

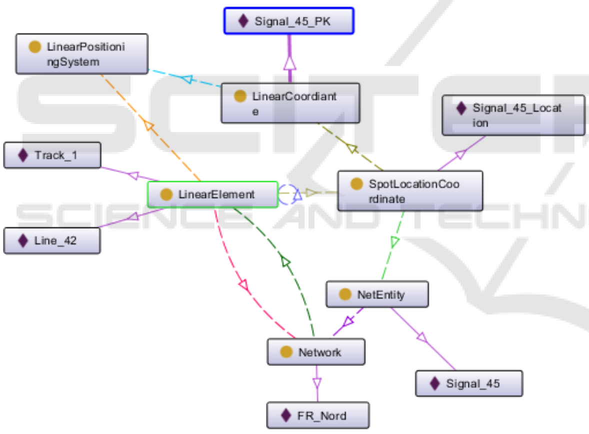

Figure 8: An example of a signal positioning in the railway network using *ATMO*.

- A NetEntity “Signal_45” object which represents the functional object, it is associated to a the railway network “FR_Nord” (of type “Network”). This signal includes:
- A Location Object : SpotLocationCoordinate named “Signal_45_Location” to represent its geometric footprint on the railway infrastructure, which references two objects:
- A PositioningSystem : LinearCoordinate Object named “Signal_45_PK” to represent the precise position in a linear frame of reference attached to the LinearPositioningSystem topology
- A Topology : LinearElement object named “Track_1” to represent the portion of track to which the signal is attached and which is part of a topological element of the same nature but at a higher level “Line_42”.

4 DISCUSSION

In this work, we have chosen a modular conceptualisation of the railway on-board map. The different modules developed corresponding to the domains of the ontologies are then composed to constitute the global ontology. This composition is achieved through the existing links between the different modules which are : “Railway”, “Track”, “TrackSide” and “Operational” (see *figure* 9).

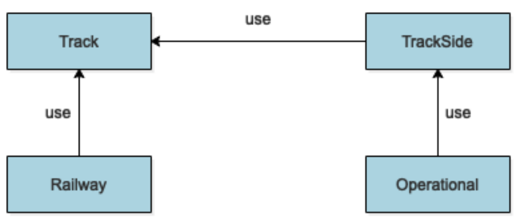

Figure 9: Modules structuring.

The reasoning, inference and updating needs of this model will be more simple thanks to the modularity. A reflection is currently underway to define axioms and assess the global ontology. Indeed, quality and correctness are the two important aspects concerning heavyweight ontology evaluation. With these perspectives in mind, we identified different metrics to measure the ontology quality (computational efficiency, adaptability and clarity) and the ontology correctness (accuracy, completeness, clarity, and consistency) (Hlomani and Stacey, 2014).

Modularity also allows the re-usability of the ontology or of its modules. Indeed, we proposed a framework to guarantee the safety of the autonomous system from the upstream design phases thanks to the use of *ATMO* and its alignment with the safety rules (Chouchani et al., 2022).

5 CONCLUSIONS AND FUTURE WORKS

In this work, we presented the methodological framework adopted to develop *ATMO* the modular ontology of on-board map of autonomous train. The main contributions of this work are : (i) the use of standards to provide a semantic map model ; (ii) the detailed description of the ontology development methodology *METHONTOLOGY* ; and (iii) the modularization paradigm used to manage the complexity of the ontology.

knowledge acquisitions were explicit and implicit by referring to bibliographic research, expert opinions as well as national and international standards and models. After a validation of the resulted lightweight ontology, presented in the form of the *UML* model, the ontology is sufficient enough to experiment with reasoning and evaluate the heavyweight ontology.

The problem of building a modular ontology approached in this work, can serve as a basis for a reflection on the approach of developing an ontology of the railway domain in general. Indeed, this proposal will be discussed in future work within the framework of the *OntoRail* project aiming to create a global dictionary and to unify the vocabulary used by the various international actors and standards like *RSM*, *IFC* and *EULYNX* (OntoRail, 2022).

REFERENCES

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*, 284(5):34.

BSI (2022). Building smart ifc, homepage. https://technical.buildingsmart.org/.

Chouchani, N., Debbech, S., and Perin, M. (2022). Model-based safety engineering for autonomous train map. *Journal of Systems and Software*, 183:111082.

D'Aquin, M., Schlicht, A., Stuckenschmidt, H., and Sabou, M. (2009). Criteria and Evaluation for Ontology Modularization Techniques. In Stuckenschmidt, H., Parent, C., and Spaccapietra, S., editors, *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, volume 5445 of *Lecture Notes in Computer Science*, pages 67–89. Springer.

de Almeida Falbo, R., de Menezes, C. S., and Rocha, A. R. (1998). A systematic approach for building ontologies. In Coelho, H., editor, *IBERAMIA*, volume 1484 of *Lecture Notes in Computer Science*, pages 349–360. Springer.

Debbech, S. (2019). Ontologies pour la gestion de sécurité ferroviaire : intégration de l'analyse dysfonctionnelle dans la conception.

EULYNX (2022). Eulynx homepage. https://www.eulynx.eu/.

Fernandez-Lopez, M., Gomez-Perez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*, pages 33–40, Stanford, USA.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220.

Hlomani, H. and Stacey, D. A. (2014). Approaches , methods , metrics , measures , and subjectivity in ontology evaluation : A survey.

IEEE (1996). Ieee guide for software quality assurance planning. IEEE Std 730.1-1995.

InteGRail (2022). Integrail project, homepage. https://www.integrail.info.

OntoRail (2022). Ontorail project, homepage. https://ontorail.org/ontorail/index.php?title=Main_Page.

Pathak, J., Johnson, T. M., and Chute, C. G. (2009). Survey of modular ontology techniques and their applications in the biomedical domain. *Integr. Comput.-Aided Eng.*, 16(3):225–242.

Tutcher, J., Easton, J. M., and Roberts, C. (2017). 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*, 3(2):F4015001.

UIC (2022). Railtopomodel homepage. https://www.railtopomodel.org/en/.

Uschold, M. and King, M. (1995). Towards a methodology for building ontologies. In *In Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95*.