

# On Linking Heterogeneous Railway Knowledge Graphs

Challenges in Integrating ERA and OpenStreetMap Rail Infrastructure Representations

Stefan Bischof, Gottfried Schenner\*

Siemens AG Österreich, Vienna, Austria

## Abstract

In the last years railway infrastructure data has become available via public SPARQL endpoints. The ontologies/-vocabularies used to represent the topology part of the railway data are mainly based either on the Open Street Map (OSM) data model or on the UML-based RailTopoModel. In this paper we discuss some of the challenges of integrating railway infrastructure data especially topological data. As an example, we show how to link the data available for the Austrian railway network using Open Street Map data via the QLever SPARQL endpoint and data from the ERA Knowledge Graph.

## Keywords

Railway Infrastructure, Railway Topology, Data Linking, Knowledge Graph

## 1. Introduction

The digital transformation of railway systems has led to an increasing amount of structured data being available about rail infrastructure and topology. Knowledge graphs (KG) have emerged as a powerful paradigm for representing complex domains like railway systems, offering semantic richness and flexible query capabilities. However, railway data often exists in isolated knowledge repositories that follow different modelling patterns, limiting the potential for comprehensive analysis and decision support.

In this paper, we present our ongoing work on linking rail topology and infrastructure data between two knowledge graphs with fundamentally different modelling approaches: the European Railway Agency (ERA) knowledge graph and a knowledge graph derived from OpenStreetMap (OSM) geographical information system data. We selected these knowledge graphs because they are openly available, making them ideal candidates to exemplify integration challenges that commercial railway operators and infrastructure managers commonly face when attempting to integrate data with external sources.

The ERA knowledge graph represents a centralized, authoritative data source containing high-quality information provided by the main infrastructure managers across EU member states. Based on the Reference Implementation for National Inventories (RINF), it offers standardized representations of operational points (such as stations) and sections-of-line (connections between stations). Its topology model derives partially from the Rail Topological Model (RTM). While operational points in this knowledge graph have geolocations, they lack fine-granular spatial information about individual infrastructure elements, such as individual tracks, switches or signals.


In contrast, the OpenStreetMap-based knowledge graph follows a community-driven approach with broader but potentially less consistent coverage. It encompasses not only main railway infrastructure but also private facilities and offers more detailed spatial data about individual tracks, signals, and other railway assets. Additionally, this knowledge graph provides contextual information about surrounding entities such as platforms, station buildings, and various points of interest.

Linking these complementary knowledge sources is valuable as it allows users to leverage the strengths of both systems: the authoritative nature and standardization of the ERA knowledge graph alongside the broader coverage and fine-granular details of the OSM knowledge graph. However, establishing these links presents significant challenges that go beyond simple entity alignment. The different

---

*Sem4Tra, co-located with SEMANTiCS'25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria*

\*Corresponding author.

 0000-0001-9521-8907 (S. Bischof); 0000-0003-0096-6780 (G. Schenner)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

conceptual models, granularity levels, and coverage scopes necessitate more complex correspondence patterns that can bridge semantic and structural gaps between the representations.

Our work illustrates these integration challenges through concrete examples and proposes approaches to establish meaningful links between the knowledge graphs. The findings are relevant not only for the specific case of combining ERA and OSM data but also for the broader context of integrating railway data with other sources in industrial settings.

The remainder of this paper is organized as follows: Section 2 reviews related work in semantic railway models and semantic integration of railway infrastructure data. Section 3 describes the structure and characteristics of both knowledge graph systems in detail. Section 4 discusses the challenges encountered and approaches developed. Section 5 describes the case study of linking passenger stations in Austria. Finally, Section 6 concludes with implications for practice and directions for future research.

## 2. Related Work

Camarazo et al. review several existing ontologies of railway infrastructure, analyse the challenges of alignments, and propose an ontology alignment method [1]. In our work we focus on linking instance data in knowledge graphs instead.

Verstichel et al. describe and compare different approaches on how to integrate rail infrastructure data at the data level [2]. In particular UML and ontology-based approaches are compared, showing the advantages of ontology-based data integration. They acknowledge the main problem of scenarios based on multiple ontologies: creating mappings between ontologies is a tedious, complex, and error-prone task. One scenario which would partially address this problem are local ontologies which are built on top of shared ontologies. These can be either upper ontologies such as BFO [3] or vertically defined ontologies such as the Rail Topology Ontology [4].

Agreeing on such a common, shared ontology as a basis for different ontologies for railway infrastructure is a complex task in practice. So far, the different ontologies in this area are formally independent although conceptually, however some converge to RTM concepts such as the ERA KG ontology or the railML ontology, which is currently under development.

Rahmig and Richter show how to extract rail topology data from OSM and combine it with data from rail measurement campaigns [5]. The integrated data is then exported as railML XML file. The data has to be further enriched by data from additional sources to obtain 3D data to enable railway simulations. While this tool chain combines railway geodata from different sources, the data pipeline which implements the data integration is custom software for a single use case and cannot be applied to other use cases.

## 3. Preliminaries

Traditional graph representations, with nodes representing switches and edges representing tracks, prove inadequate for expressing the complex realities of railway infrastructure topologies. Such simplistic models fail to capture essential characteristics required for meaningful operational analysis, such as track directionality, physical clearance constraints, valid path combinations through junctions, and track-specific attributes like electrification systems or speed restrictions. Most critically, simple graphs cannot properly represent navigability—the precise constraints determining which physical paths rolling stock can traverse based on factors like train length, weight, or gauge compatibility.

Several more sophisticated models have emerged to address these limitations, including RTM, which forms the foundation for the European IRS 30100 standard [6]; railML (Railway Markup Language), which implements RTM and provides XML schemas for railway data exchange [7]; RINF (Register of Infrastructure) and the ERA Knowledge Graph [8], which standardizes infrastructure parameters across European networks<sup>1</sup>. These models employ concepts such as `netElements` and `netRelations`

---

<sup>1</sup>[https://www.era.europa.eu/domains/registers/rinf\\_en](https://www.era.europa.eu/domains/registers/rinf_en)

to model topological connections, distinguish between operational and physical views of infrastructure, and incorporate detailed spatial and functional attributes necessary for comprehensive railway topology representation.

The first KG we use in our case study is the ERA Knowledge Graph [8]. ERA adopted integrates its base registries in a comprehensive KG system. This approach enabled new capabilities, such as route compatibility checks, which were previously unsupported because the data was distributed over different registries (RINF and ERATV). The key contributions of the initial ERA KG included the development of an official railway ontology, a reusable Knowledge Graph of European railway infrastructure, a flexible and cost-efficient system architecture, and an open-source RDF-native web application. The ERA Ontology partially uses RTM for representing the railway topology.

Geographic Information Systems (GIS) can also model railway topology and infrastructure data. Even though GIS are more focused towards spatial modelling, GIS also model railway infrastructure to provide for example routing services. In particular we use OpenRailwayMap<sup>2</sup>, which is part of Open Street Map (OSM), for our case study. OSM is a public, collaborative effort to map geographic data. The OpenStreetMap (OSM) data model employs a topological structure composed of three primitive elements: *nodes* (point features with geographical coordinates), *ways* (ordered sequences of nodes representing linear features or area boundaries), and *relations* (groups of nodes, ways, or other relations describing complex geographical relationships). Each element can be annotated with an unlimited number of key-value pairs called *tags*, which store semantic information about geographic features. This flexible schema-free approach allows contributors to map diverse features without rigid predefined categories, while community-developed tagging conventions ensure data consistency.

OpenRailwayMap builds upon this foundation as a specialized project focusing on railway infrastructure, utilizing the same core data model but implementing domain-specific tagging schemes to represent detailed railway-specific attributes such as electrification systems, signalling equipment, maximum speeds, and track gauges. Even though OSM lacks a semantic representation, several works have provided such semantic representations. In our case study we use OSM via QLever [9]. While there exist other SPARQL endpoints, such as <https://sophox.org/> or <http://linkedgeodata.org/>, the QLever OSM KG provides custom functionality to find all nodes in a geographical area.

## 4. Challenges for linking the data

Even though both, the ERA KG and OSM via QLever, represent the European railway infrastructure in RDF, combining the two datasets and linking their entities is not trivial.

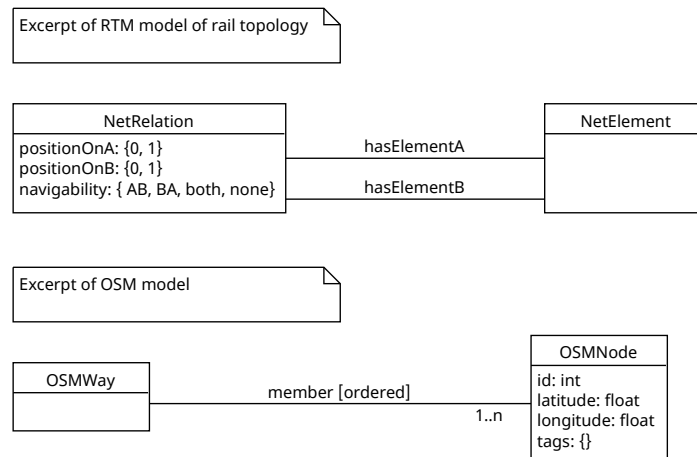
### 4.1. Topology: RTM vs OSM

Figure 1 shows the fundamentally different ways how the RTM and OSM model the railway infrastructure. In the case of RTM, NetElements represent a segment of a rail (linear NetElement) or a point in the rail network (non-linear NetElement). Linear NetElements have a start (0) and end (1) position. NetRelations connect the ends of exactly two NetElements (hasElementA, hasElementB). The positions (positionOnA/positionOnB) determine, which end (0/1) of one NetElement is connected to which end the other NetElement by the NetRelation. Finally, the navigability expresses the possible ways a train can pass the NetRelation, e.g., from A to B, from B to A, in both or in no direction. The navigability of none is used to express that a train can not directly got from the left and right branch of a railway switch.

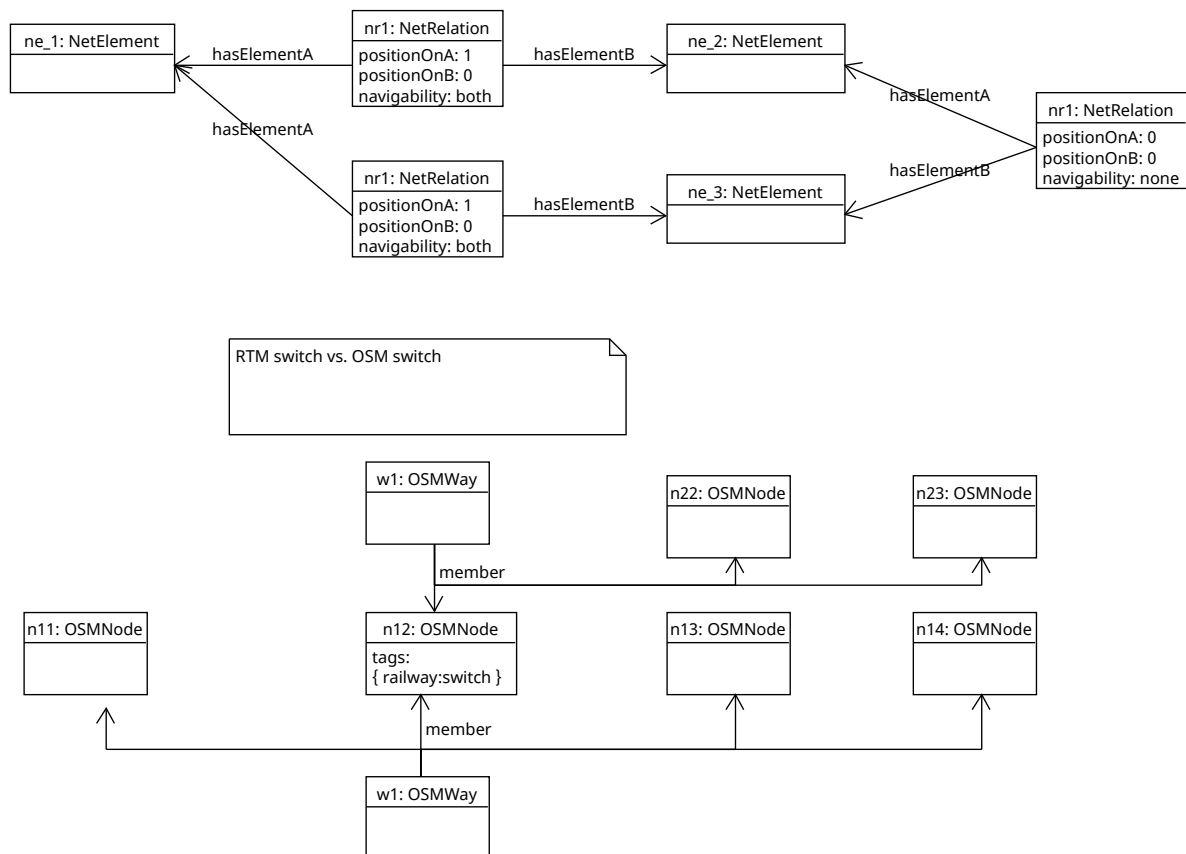
In the case of OSM data on the other hand the railway network in its most basic form is modelled by OSMWays and OSMNodes. OSMWays contain a list of ordered OSMNodes. The order determines the order in which a train would pass the OSMNodes of the OSMWay in one direction. The OSMNodes have Geo data (latitude/longitude). Both OSMWays and OSMNodes can have optional tags indicating the

---

<sup>2</sup><https://www.openrailwaymap.org/>



**Figure 1:** Modelling railway topology in RTM and OSM.

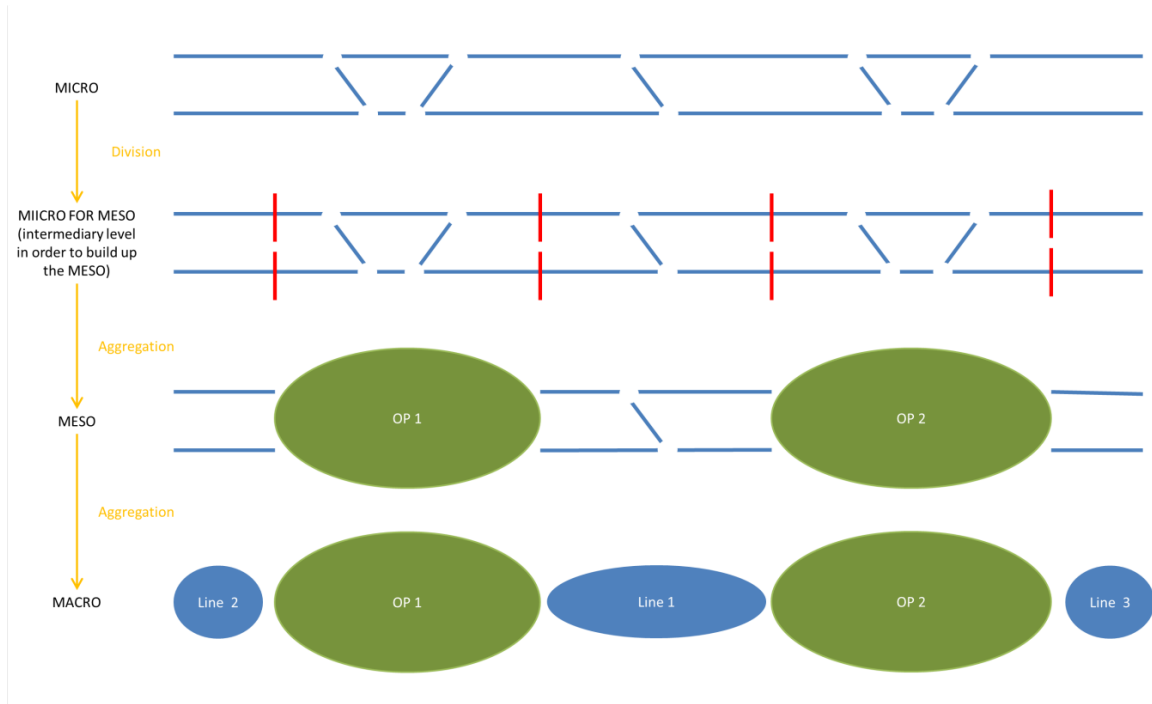


**Figure 2:** Modelling railway topology in RTM and OSM.

function of the elements, e.g., a OSMWay representing a rail will have the tag `railway:rail`, whereas a OSMNode representing a signal has the tag `railway:signal`.

## 4.2. Structural challenges

Figure 2 shows how the same switch is modelled in the different formalisms. This illustrates the challenges the arise, because RTM and OSM are using a fundamentally different way to represent the railway topology.



**Figure 3:** RTM Aggregation Level, illustration from <https://en.wikipedia.org/wiki/RailTopoModel>

In the RTM case a switch can be identified by the navigability of the `NetRelations`. The OSM-switch is identified by the tag `railway:switch`. If the tag is absent, switches can be identified by finding OSMNodes with three neighbouring nodes. Still the problem remains, which neighbouring node is the tip and which are the branches of the switch. This can only be done by reasoning about the geo data and the angle of the outgoing edges. Clearly one cannot expect the OSM data to have the same level of detail as the RTM model which was specifically created to represent railway infrastructure data.

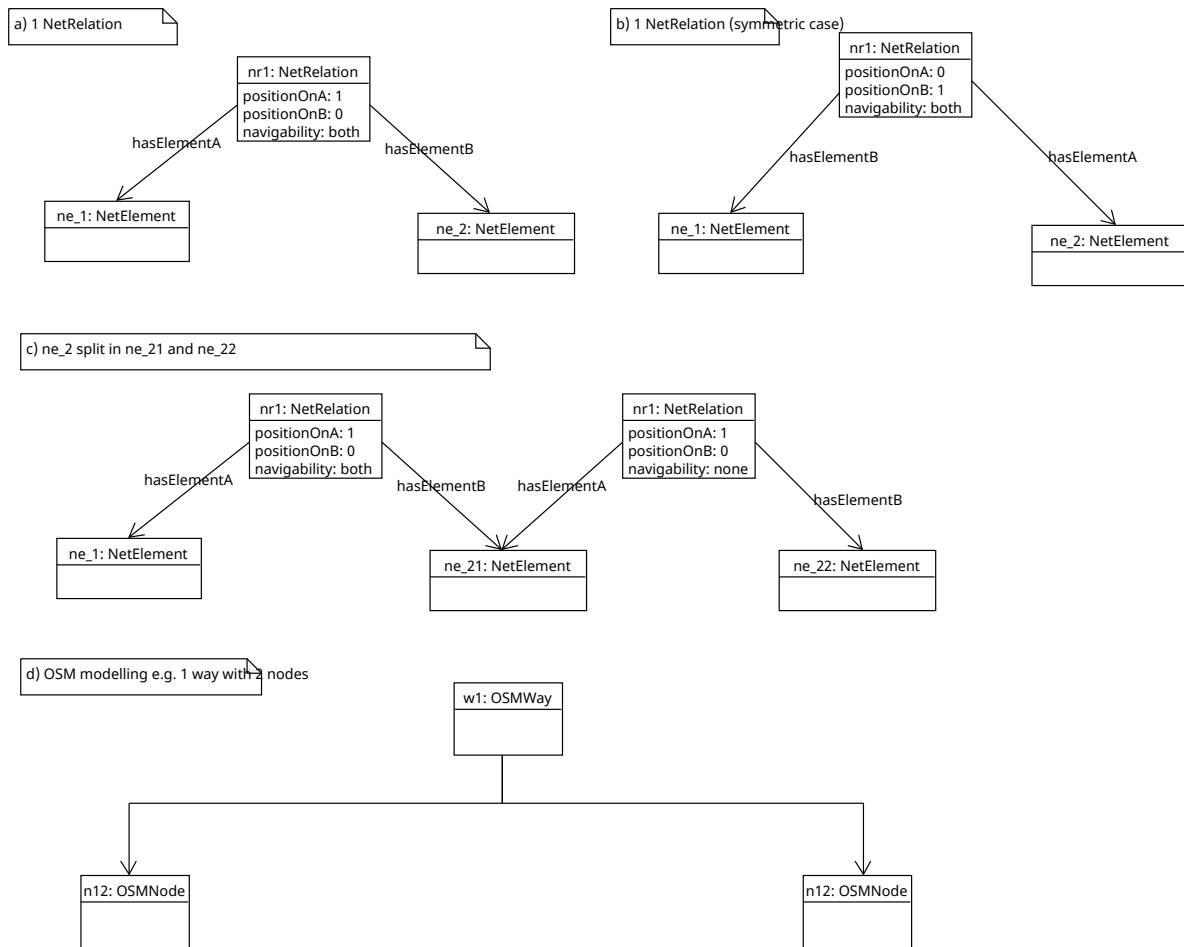
As can be seen in the figure a simple entity matching approach does not work, as there are no corresponding entities. The only candidate for a 1-1 matching would be the end (position 1) of the `NetElement ne_1` as it corresponds to the OSMNode `n12`. The same applies to the start (position 0) of `NetElement ne_2` and `ne_3` respective.

#### 4.2.1. Aggregation level

A special feature of the RTM is that it allows modelling the topology in different aggregation levels (Figure 3) within the same formalism. Only on the *micro* level the topology inside the operational points is known. The *meso* level focuses on the operational points and their connecting tracks. This corresponds to the level of the ERA knowledge graph. On the *macro* level the connecting tracks are aggregated in the information about the railway lines. In contrast in OSM everything is modelled on the same level. Sometimes special OSMRelations are used to aggregate all the topology elements of an operational point. But this is by no means required.

### 4.3. Symmetries and Granularity

Another challenge arises because of the possibility to model the same instance data in various ways. Figure 4 shows how the same segment of a railway network could be represented in various ways even in the same formalism. The cases a) and b) are symmetric cases. Case c) splits the second `NetElement` into two `NetElements`. In case d) there is only one implied `NetElement`, i.e., the track segment between the two connected OSMNodes.



**Figure 4: Symmetries and granularity**

#### 4.4. Linking Shared Entities

The easiest way to link common entities is via a common identifier, e.g., if a property or a combination of properties of one entity can be directly mapped to a property of the other entity in a specific context e.g. the ref attribute of OSMNodes with tag railway:switch correspond to rdfs:label of railML:SwitchIS entities in a given operational point. If identifiers are not available it might still be possible to use geospatial coordinates to align the entities. Depending on the type of entities the geospatial might not be exactly the same in the different data sources. Another option for infrastructure elements within the railway topology is to use topological information e.g. a switch could be identified by being the ‘second left-traversed switch from the beginning of the operational point in direction X’. A more sophisticated approach is to represent the topologies as a graph and to use graph algorithms (isomorphism, graph edit distance or similar) find a mapping between the topology graphs.

## 5. Case study: Passenger stations of Austria

As a case study we attempt to integrate the ERA and the OSM data for all passenger stops and their connections in Austria.

We use these prefixes in the following SPARQL queries:

```

PREFIX country: <http://publications.europa.eu/resource/authority/country/>
PREFIX era: <http://data.europa.eu/949/>
PREFIX ex: <https://www.example.org/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
  
```

```

PREFIX loc: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmnode: <https://www.openstreetmap.org/node/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

```

## 5.1. Getting the passenger stations

The SPARQL query in Listing 1 gets all operational points of Austria and their geo data (in the current dataset 1799 entries). It contains the (small) stations and passenger stops (1496 entries) but also other operational points like (domestic) border points, private sidings, etc. We need these (intermediate) points because we want to also reason about the topological connections of the operational points.

**Listing 1:** ERA operational points

```

SELECT * WHERE {
  ?op a era:OperationalPoint .
  ?op era:opName ?opName .
  ?op era:inCountry country:AUT .
  ?op era:opType ?opType .
  ?opType skos:prefLabel ?opTypeLabel .
  FILTER (LANG(?opTypeLabel) = "en")
  ?op era:uopid ?uopid .
  ?op loc:location ?loc .
  ?loc loc:lat ?o1 .
  ?loc loc:long ?o2 .
}

```

With QLever we can query all the OSMNodes within a certain region. The following queries OSMNodes representing railway station, stops and halts in Austria (osmrel:16239). This query finds 3848 OSMNodes, which are significant more than in the ERA case, due to the fact that every halt position within a station is represented by a different OSMNode. If we group the halt positions belonging to the same station we still get 1698 OSMNodes in contrast to the 1496 ERA entries, because the result also contains privately operated railway lines, cog railway lines etc. not represented in the ERA KG.

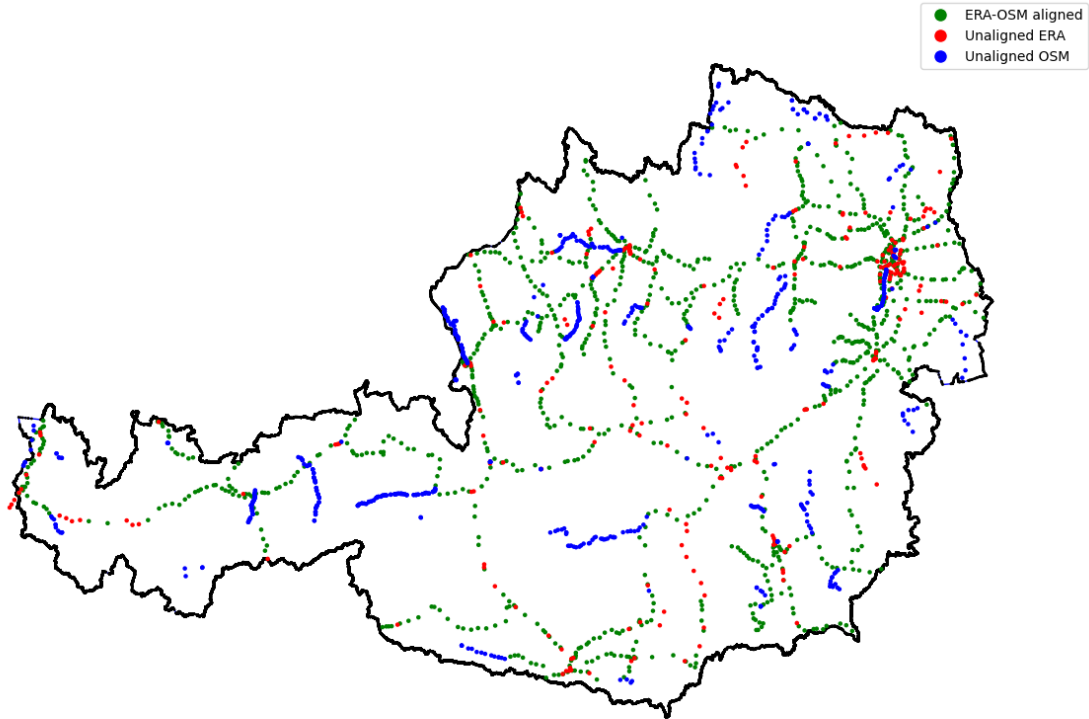
**Listing 2:** OSM Station

```

SELECT ?osm_id ?name ?type ?ref ?geometry WHERE {
  osmrel:16239 ogc:sfContains ?osm_id .
  ?osm_id osmkey:railway ?type .
  OPTIONAL {
    ?osm_id osmkey:railway:ref ?ref .
  }
  ?osm_id osmkey:name ?name .
  ?osm_id rdf:type osm:node .
  # filtering out subway nodes
  FILTER NOT EXISTS {
    ?osm_id osmkey:subway "yes" } .
  FILTER (?type = "station" || ?type = "stop" || ?type = "halt")
  ?osm_id geo:hasGeometry/geo:asWKT ?geometry .
}

```





**Figure 5:** Automatic operational point alignment Austria

How can we align the ERA operational points with the corresponding OSMNodes? As a first attempt one can look for common identifiers. Some OSMNodes with tag `railway:station` also contain an entry `railway:ref`. The value of this `ref` is the same as the `uopid` without the AT prefix, e.g., in Feldkirch the `uopid` is `ATFk` and the `railway:ref` is `Fk`. Unfortunately this is not always the case (about 40% of the OSMNodes in the result have a `railway:ref`) and even worse, sometimes the OSMNode for the railway station is not contained in any other relation i.e. it is isolated from the railway network. Train stops and halts on the OSM-level on the other hand do not have a `railway:ref`. After some trial and error as a pragmatic solution we decided to use the distance of the Geo coordinates and assign the OSMNodes to the nearest operational point within a certain range (200 m). The found alignments (green in Figure 5) are a good approximation of the desired result. Of course, one could go on and further refine the queries, e.g., by incorporating route information, but as stated earlier, this is not the goal of this paper. The main goal is to show that finding alignments is an iterative process of finding the right queries, choosing the alignment strategy, and testing, especially if one is not an expert in the involved ontologies. Some of the ERA OPs cannot be aligned with OSMNodes and vice versa for various reasons. One example is defunct railway stations that are no longer stops for passenger trains and therefore no longer tagged as `halt/stop` in OSM. Finding alignments is a dynamic process that needs to be repeated periodically because of changes in the data sources.

## 5.2. Topology

As a next step we want to align the topological connections between the operational points. In the case of the operational points the topological connections between direct neighbours are available as `SectionOfLine` entities. In the case of OSM we can derive a similar information by the `OSMRelations` corresponding to train routes (Listing 3). In the QLever ontology all contained `osm_nodes` in a relation are ordered with a `member_pos` property. The next passenger stop for a given node is then the next passenger stop with a higher `member_pos` (not necessary sequential as there are also other type of



nodes and ways contained in the same relation).

**Listing 3:** next passenger stops in OSM

```
CONSTRUCT {
  ?osm_id1 ex:nextPassengerStop ?osm_id2 .
  ?osm_id1 osmkey:name ?name1 .
  ?osm_id2 osmkey:name ?name2 .
}
WHERE {
  {
    SELECT ?rel ?m1 ?osm_id1 ?osm_pos1 (MIN(?osm_pos2) AS ?min_pos2) WHERE {
      BIND (<https://www.openstreetmap.org/relation/1658658> AS ?rel)
      ?rel osmrel:member ?m1 .
      ?m1 osmrel:member_id ?osm_id1 ;
        osmrel:member_pos ?osm_pos1 .
      ?osm_id1 a osm:node .
      ?osm_id1 osmkey:railway ?function1 .
      FILTER (?function1 != "platform")

      ?rel osmrel:member ?m2 .
      ?m2 osmrel:member_id ?osm_id2 ;
        osmrel:member_pos ?osm_pos2 .
      ?osm_id2 a osm:node .
      ?osm_id2 osmkey:railway ?function2 .
      FILTER (?function2 != "platform")

      FILTER (?osm_pos2 > ?osm_pos1)
    }
    GROUP BY ?rel ?m1 ?osm_id1 ?osm_pos1
  }
  ?rel osmrel:member ?m2 .
  ?m2 osmrel:member_id ?osm_id2 ;
    osmrel:member_pos ?min_pos2 .
  ?osm_id1 osmkey:name ?name1 .
  ?osm_id2 osmkey:name ?name2 .
}
```

### 5.3. Aligning infrastructure elements

To illustrate how to check the consistency of infrastructure elements with the integrated data, we compare the queries for getting the sections of lines with tunnels and railway switches using OSM and the ERA KG.

#### 5.3.1. Tunnels

For ERA the query looks like this: (Listing 4)

**Listing 4:** Section of lines with tunnels

```
SELECT * WHERE {
  ?sol era:inCountry country:AUT .
  ?line a era:NationalRailwayLine .
  ?line rdfs:label "20601" .
  ?sol era:lineNationalId ?line .
  ?sol rdfs:label ?sol_label .

  ?tunnel a era:Tunnel .
  ?tunnel era:netElement ?ne .
```

```

?tunnel rdfs:label ?tunnel_name .
?ne era:hasImplementation ?track_uri .
?track era:canonicalURI ?track_uri .
?sol era:track ?track .
}

```

In the OSM variant we use the previously derived information `nextStopPosition`. A tunnel is found for a set of neighbouring nodes, if there is a OSMWay with the tag `tunnel="yes"` between the nodes. In contrast to the ERA KG variant this query finds also tunnels within operational points. Due to the meso scope of the ERA KG these tunnels are not represented in the ERA KG.

**Listing 5:** Section of lines with tunnels

```

SELECT ?osm_id1 ?osm_id2 ?tunnel WHERE {
  BIND (<https://www.openstreetmap.org/relation/1658658> AS ?rel)
  ?osm_id1 ex:nextStopPosition ?osm_id2 .

  ?rel osmrel:member ?mrelway1 .
  ?mrelway1 osmrel:member_id ?osm_way1 .
  ?mrelway1 osmrel:member_pos ?way1_pos .
  ?osm_way1 osmway:member ?mway1 .
  ?mway1 osmway:member_id ?osm_id1 .

  ?rel osmrel:member ?mrelway2 .
  ?mrelway2 osmrel:member_id ?osm_way2 .
  ?mrelway2 osmrel:member_pos ?way2_pos .
  ?osm_way2 osmway:member ?mway2 .
  ?mway2 osmway:member_id ?osm_id2 .
  {
    SELECT ?tunnel ?tunnel_name ?tunnel_pos {
      BIND (<https://www.openstreetmap.org/relation/1658658> AS ?rel)
      ?rel osmrel:member ?mtunnel .
      ?mtunnel osmrel:member_id ?tunnel .
      ?mtunnel osmrel:member_pos ?tunnel_pos .
      ?tunnel a osm:way .
      ?tunnel osmkey:tunnel "yes" .
    }
  }
  FILTER (?tunnel_pos > ?way1_pos)
  FILTER (?tunnel_pos < ?way2_pos)
}

```

The label of the ERA tunnels and the `tunnel:name` tag of OSM are similar but the seldom match exactly. So to align the tunnels it is safer to use geo coordinates or topological information e.g. the *i*th tunnel after operational point X.

### 5.3.2. Railway switches

Railway switches are one of the most important elements of the railway network as they define the topology. They are not contained in the ERA KG so for this example we assume to have the information as micro-level instance data of a RailML ontology or a similar ontology that uses the RTM for the topological information. In RailML railway switches are explicitly modelled as the infrastructure element `SwitchIS`<sup>3</sup>. RailML switches have additional properties that link the branches (Left, Right) of the switch to the corresponding `NetRelations` on the RTM level.

In OSM, detecting switches is more challenging. The SPARQL query shown in Listing 6 demonstrates how to retrieve all OSMNodes within a railway area that may correspond to infrastructure elements.

<sup>3</sup><http://ontology.railml.org/railml3#SwitchIS>

Switches in OSM are optionally tagged with `railway=switch` and may include a corresponding `ref` tag. Notably, OSMNodes representing switches are typically part of either two or three OSM ways tagged with `railway=rail`, depending on whether they are located at the start/end of a rail segment or between segments. This structural information can be leveraged to identify switches in OSM, even when explicit switch tags are absent.

**Listing 6:** Switches with osm ref

```
SELECT DISTINCT ?osm_node ?way_count ?function ?ref WHERE {
  {
    SELECT ?osm_node (COUNT(?way) AS ?way_count) WHERE {
      ?feldkirch osmkey:name "Feldkirch" .
      ?feldkirch osmkey:railway "station" .
      ?feldkirch osmkey:railway:ref "Fk" .
      ?railway_area ogc:sfContains ?feldkirch .
      ?railway_area osmkey:landuse "railway" .
      ?railway_area ogc:sfContains ?osm_node .
      ?way osmway:member ?mway .
      ?way osmkey:railway "rail" .
      ?mway osmway:member_id ?osm_node .
    }
    GROUP BY ?osm_node
  }
  OPTIONAL {
    ?osm_node osmkey:railway ?function .
  }
  OPTIONAL {
    ?osm_node osmkey:ref ?ref .
  }
}
```

Aligning the identified OSM switches with the RTM-based switches presents a greater challenge. To establish a one-to-one alignment, it is essential to ensure that the same context is used in both RTM and OSM. In RTM, infrastructure elements are explicitly linked to their corresponding operational points. In contrast, OSM does not provide this information directly; it must be inferred—such as in the query shown in Listing 6—by identifying all nodes contained within a relation tagged with `landuse="railway"`.

## 6. Conclusion

The purpose of this paper is to share our experience that identifying instance alignments in the railway domain can be a challenging task, particularly when the ontologies differ significantly in their modelling philosophies. One-to-one alignments are often not feasible, as certain concepts may be absent in one of the formalisms. Surprisingly, even when ontologies share the same underlying model for topology (e.g., RTM), alignment remains difficult due to the various ways in which data can be represented—such as differences in symmetry, granularity, or alternative location information (e.g., metric vs. GPS coordinates). This challenge persists even when the same ontology is used.

In our experience, alignments can be effectively achieved with a human-in-the-loop approach, where domain experts leverage a combination of geographical, topological, and railway-specific knowledge to determine corresponding infrastructure elements. However, manual alignment is a tedious and time-consuming process, especially when it must be repeated periodically due to changes in the railway infrastructure—although the railway network tends to be relatively stable.

Motivated by these challenges, our research aims to develop improved semi-automatic methods for integrating railway data from heterogeneous sources. We are working on an approach on combining topological and geographical reasoning with alignment techniques such as graph-based and constraint-based mapping algorithms.

## Declaration on Generative AI

During the preparation of this work, the authors used Sonnet 4.0 in order to: paraphrase and reword, improve writing style, and grammar and check spelling. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

- [1] D. Camarazo, A. Roxin, M. Lalou, Railway Systems' Ontologies: A Literature Review and an Alignment Proposal, in: Information Integration and Web Intelligence, volume 15342 of *LNCS*, 2025, pp. 311–319. doi:10.1007/978-3-031-78090-5\_27.
- [2] S. Verstichel, F. Ongenae, L. Loeve, F. Vermeulen, P. Dings, B. Dhoedt, T. Dhaene, F. De Turck, Efficient data integration in the railway domain through an ontology-based methodology, *Transportation Research Part C: Emerging Technologies* 19 (2011) 617–643. doi:10.1016/j.trc.2010.10.003.
- [3] R. Arp, B. Smith, A. Spear, Building Ontologies with Basic Formal Ontology, MIT Press, 2015.
- [4] S. Bischof, G. Schenner, Rail Topology Ontology: A Rail Infrastructure Base Ontology, in: ISWC 2021, volume 12922 of *LNCS*, 2021, pp. 597–612. doi:10.1007/978-3-030-88361-4\_35.
- [5] C. Rahmig, A. Richter, A railway simulation landscape creation tool chain considering openstreet-map geo data, in: SUMO2014 – Modeling Mobility with Open Data, 2014, pp. 167–175.
- [6] UIC, RailTopoModel – Railway infrastructure topological model, Standard, International Railway Solution IRS 30100:2016, International Union of Railway (UIC), 2016. Edition: 1 ISBN: 978-2-7461-2513-1.
- [7] A. Nash, D. Huerlimann, J. Schuette, V. P. Krauss, RailML – A Standard Data Interface For Railroad Applications, *WIT Transactions on The Built Environment* 74 (2004) 233–240. doi:10.2495/CR040241.
- [8] J. A. Rojas, M. Aguado, P. Vasilopoulou, I. Velitchkov, D. Van Assche, P. Colpaert, R. Verborgh, Leveraging Semantic Technologies for Digital Interoperability in the European Railway Domain, in: ISWC 2021, volume 12922 of *LNCS*, 2021, pp. 648–664. doi:10.1007/978-3-030-88361-4\_38.
- [9] H. Bast, P. Brosi, J. Kalmbach, A. Lehmann, An Efficient RDF Converter and SPARQL Endpoint for the Complete OpenStreetMap Data, in: Proceedings of the 29th International Conference on Advances in Geographic Information Systems, 2021, pp. 536–539. doi:10.1145/3474717.3484256.