Building a KG for the European Railway Agency: The story

Julián Rojas

International Workshop On Semantics And The Web For Transport
SEMANTiCS Conference
September 2022

GHENT UNIVERSITY

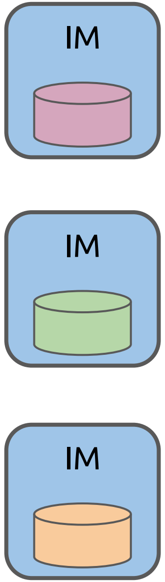
EUROPEAN UNION AGENCY FOR RAILWAYS

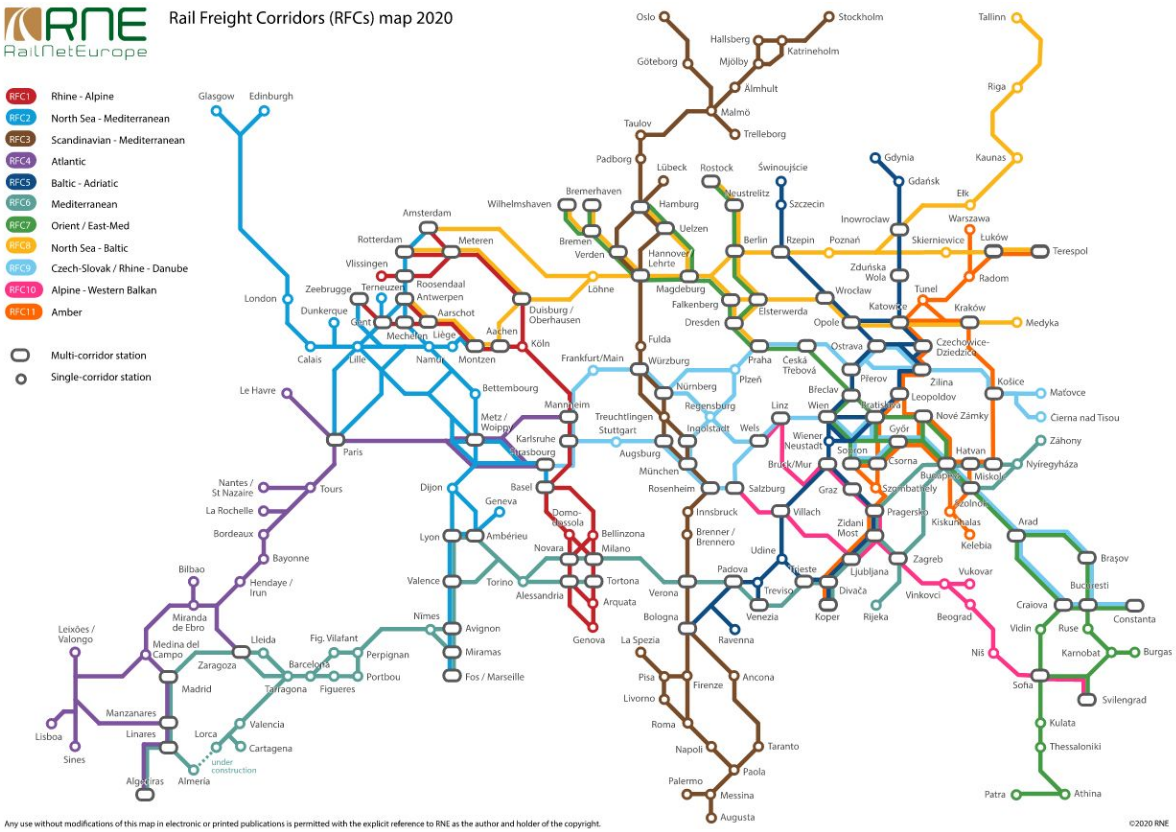
imec

Agenda

- Intro context and use case motivation
    - Who are ERA
    - Data sources
    - Use case
- Data exploration and modeling
    - Ontology definition (v1, v2 and beyond)
- Data Mapping
    - (YARR)RML mapping architecture
    - Challenging mappings (Conditionals, gradient profile)
- Data integration
    - Integration at mapping time vs integration at Querying time
    - Accurate routing (railML)
    - OSM (geo-shapes, others)
    - Challenging mappings/post-processing (interlinking, CONSTRUCT queries)
- RCC
    - Solving routes (OSRM vs client-side)
    - Compatibility assessment (SHACL-rules?)
    - Compatibility based on meteorological conditions
    -
- Conclusions and next steps
    - Quality assurance
    - IM independent generation (mappings, phase-out of RINF)
    - LDES-based generation
    - Shortest-path querying on KGs (Contraction Hierarchies)

European railway ecosystem





RU

RU

RU

The European Agency for Railways (ERA)

Mission: Moving Europe towards a sustainable and safe railway system without frontiers.

Tasks:

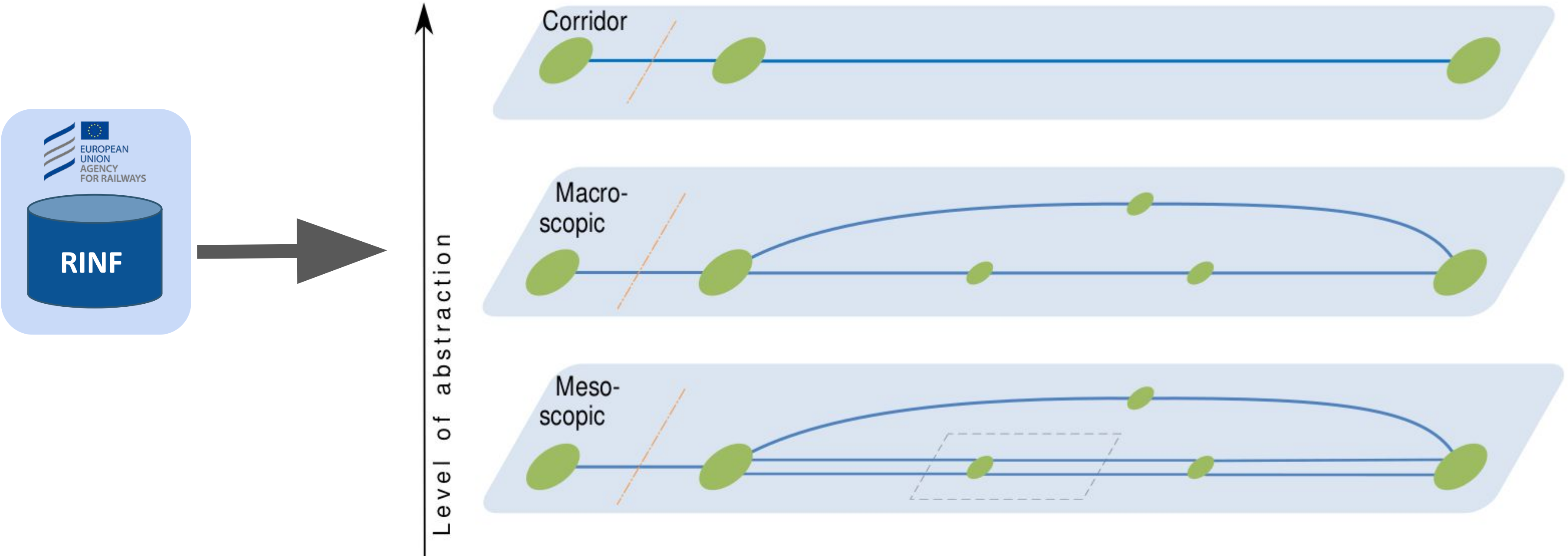
- Promote a **harmonised approach** to railway safety
- Devise the technical and legal **framework** in order to enable **removing technical barriers**, and acting as the system authority for ERTMS and telematics applications
- **Improve accessibility** and use of railway system **information**
- Act as the **European Authority** under the 4th Railway Package issuing vehicle (type) authorisations and single safety certificates, while improving the competitive position of the railway sector.

EUROPEAN UNION AGENCY FOR RAILWAYS

ERA aims on creating interoperability for stakeholders

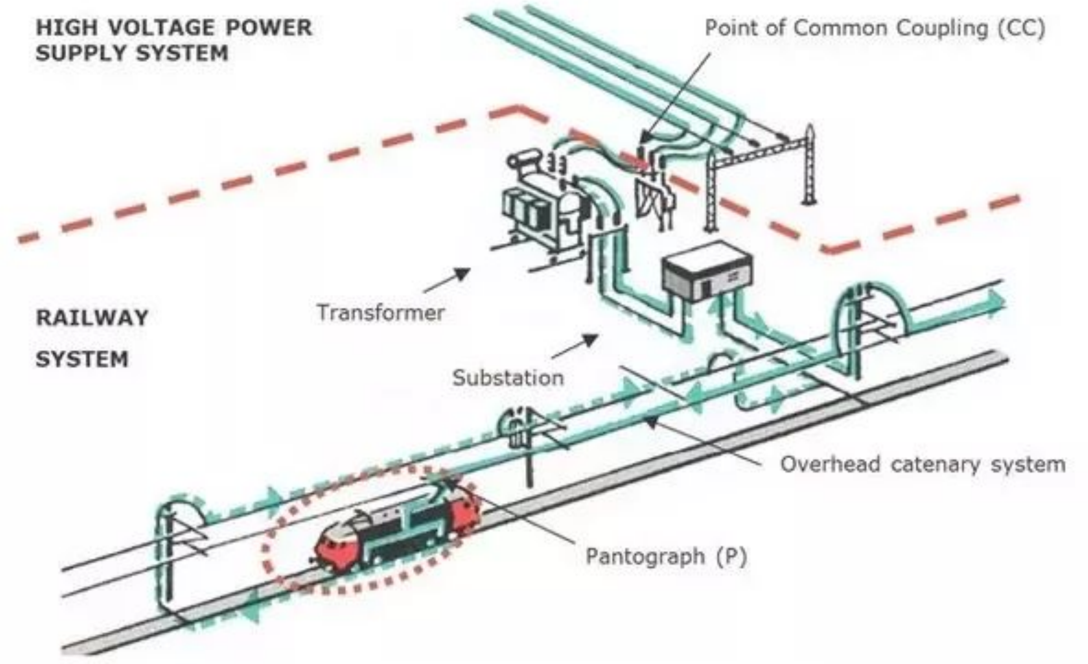
Register of Infrastructure (RINF)

RINF contains the main features of fixed installations related to the subsystems of **infrastructure**, energy and parts of control-command and signaling



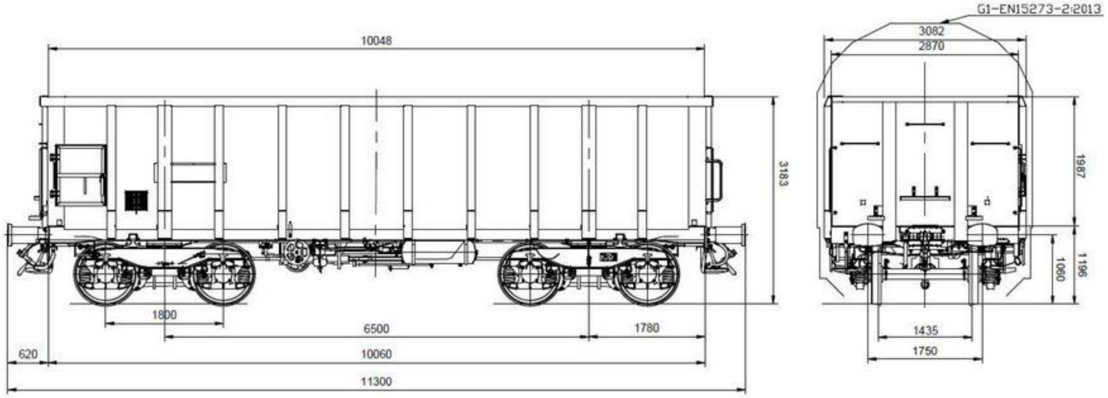
Register of Infrastructure (RINF)

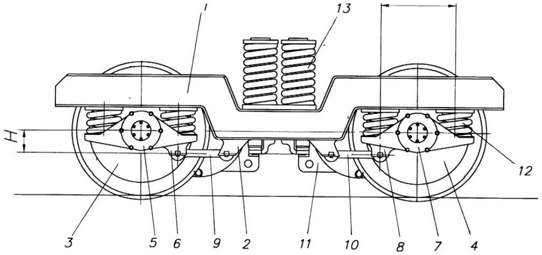
RINF contains the main features of fixed installations related to the subsystems of infrastructure, **energy and parts of control-command and signaling**



European Register of Authorized Types of Vehicles (ERATV)

ERATV publishes and keeps an up-to-date set of authorized types of vehicles including information that references their technical specifications.





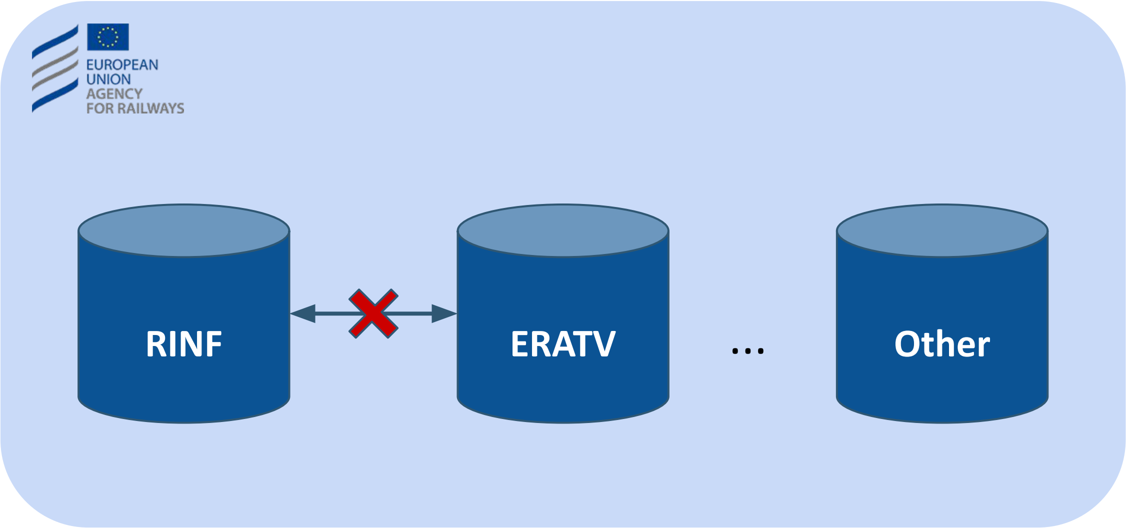
Use case: Route Compatibility Check (RCC)

"Before a railway undertaking uses a vehicle in the area of use specified in its authorisation for placing on the market, it shall check: …(b) that the vehicle is compatible with the route on the basis of the infrastructure register…"

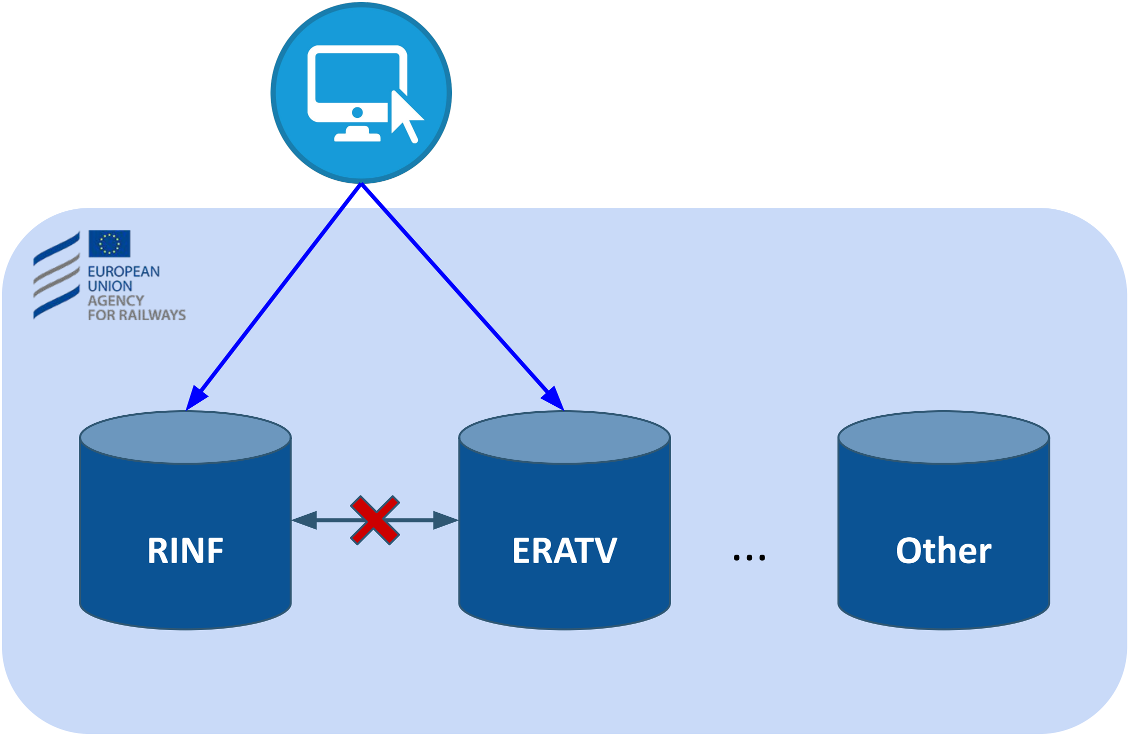
European regulation 2016/797

E.g.: What is the shortest **compatible route** with **vehicle C30-M** from Paris to Amsterdam?

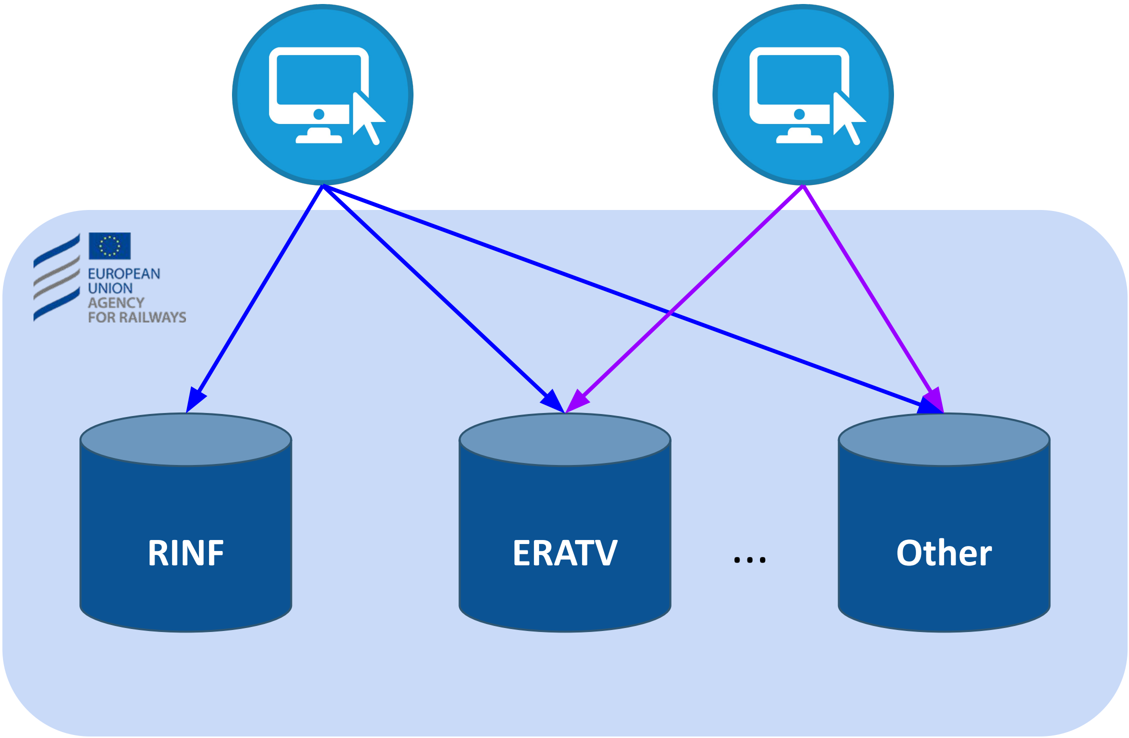
Data silos difficult interoperability



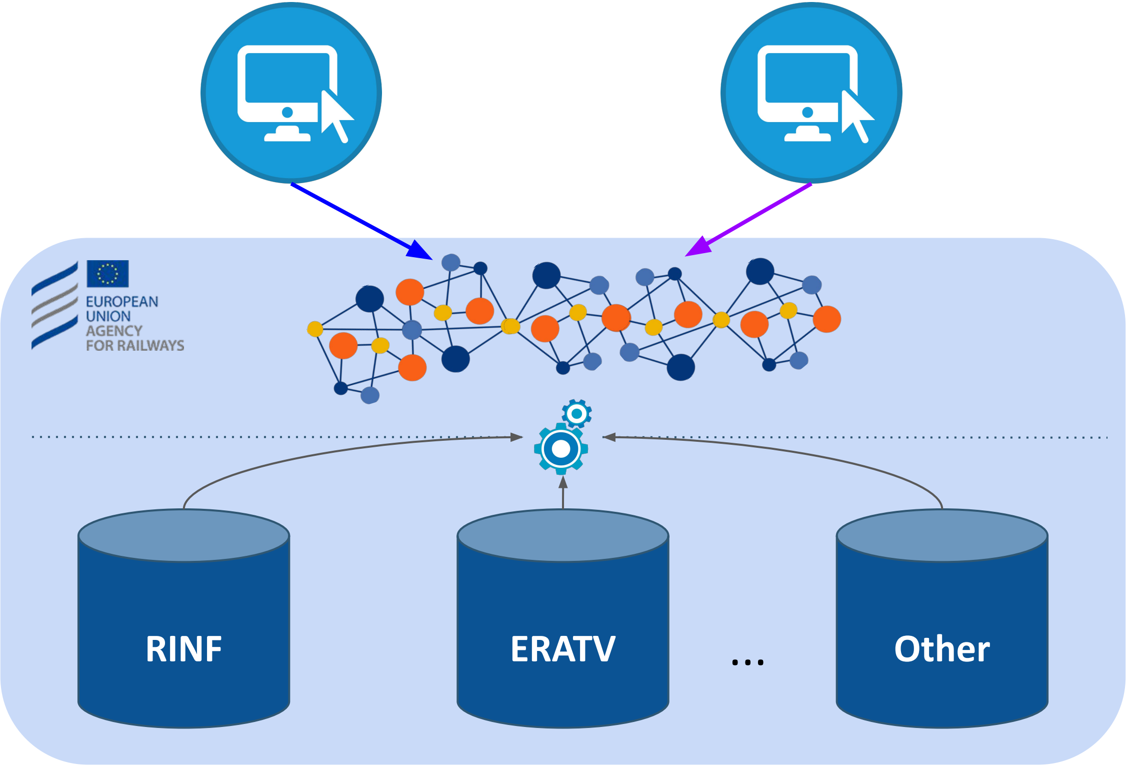
Data silos difficult interoperability and **force application-centric solutions** for data integration



Data integration hidden in application code is **expensive to maintain and reuse**



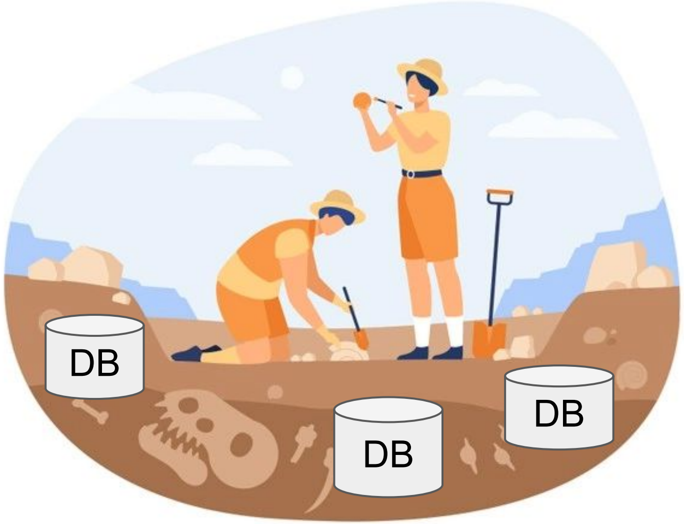
A **semantic layer** for Knowledge Graph-based interoperability that enables data-centric solutions



ACT I: Data exploration and modelling

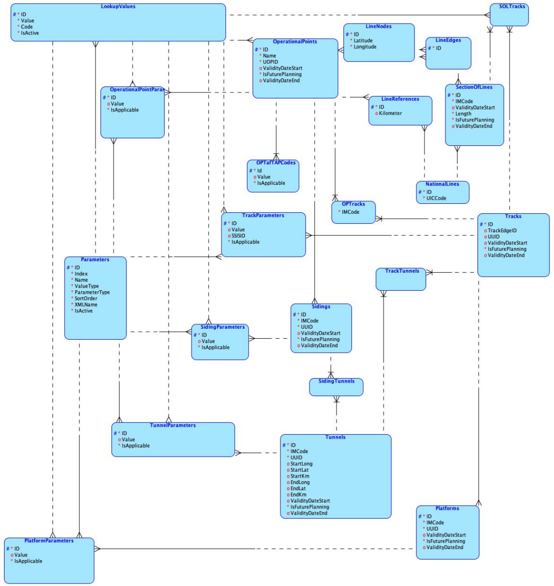
We started with some *data archeology* …

The documentation

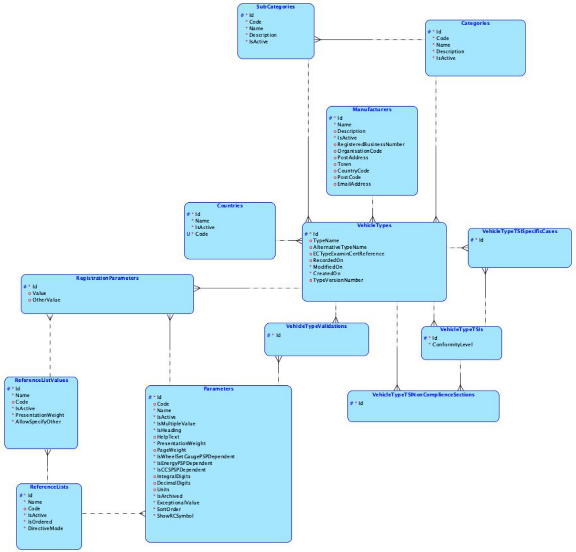


Entity-relation diagrams

And we went from this…

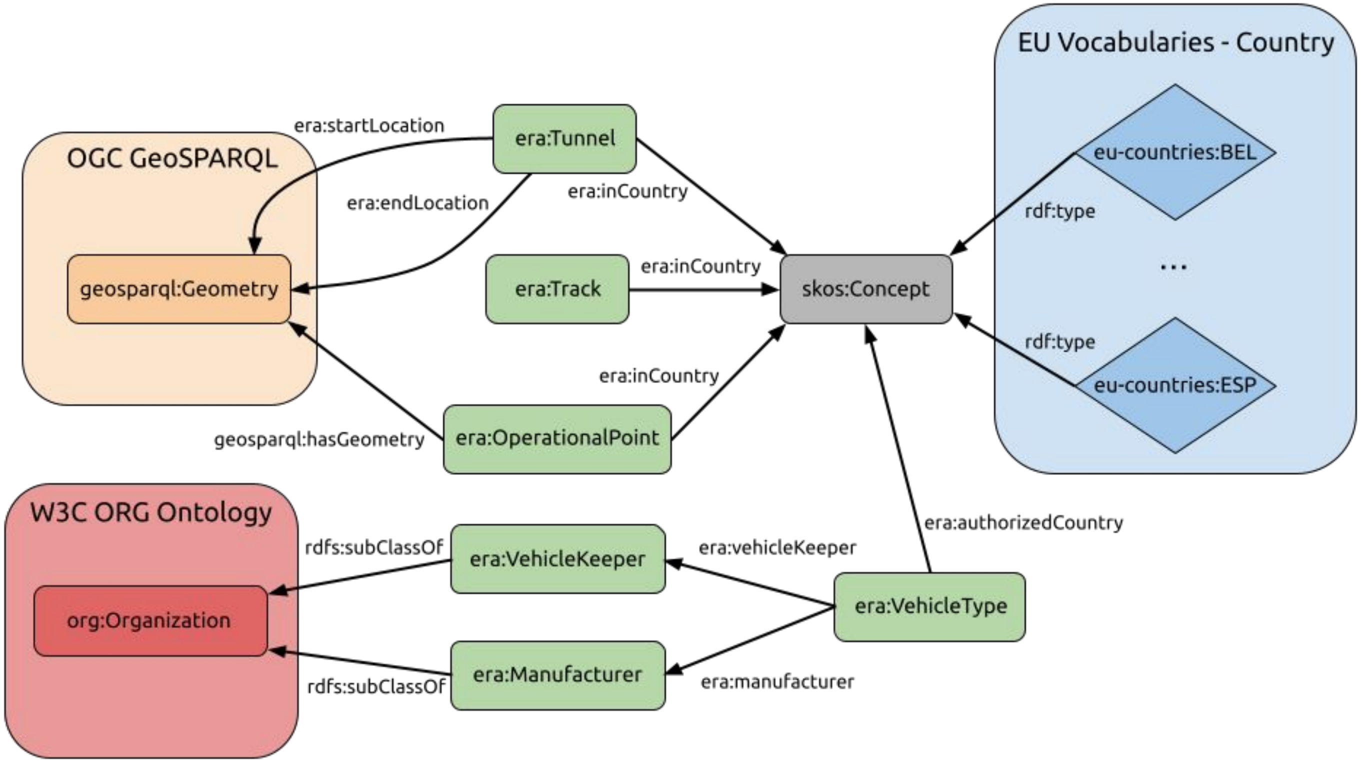


Main tables in RINF

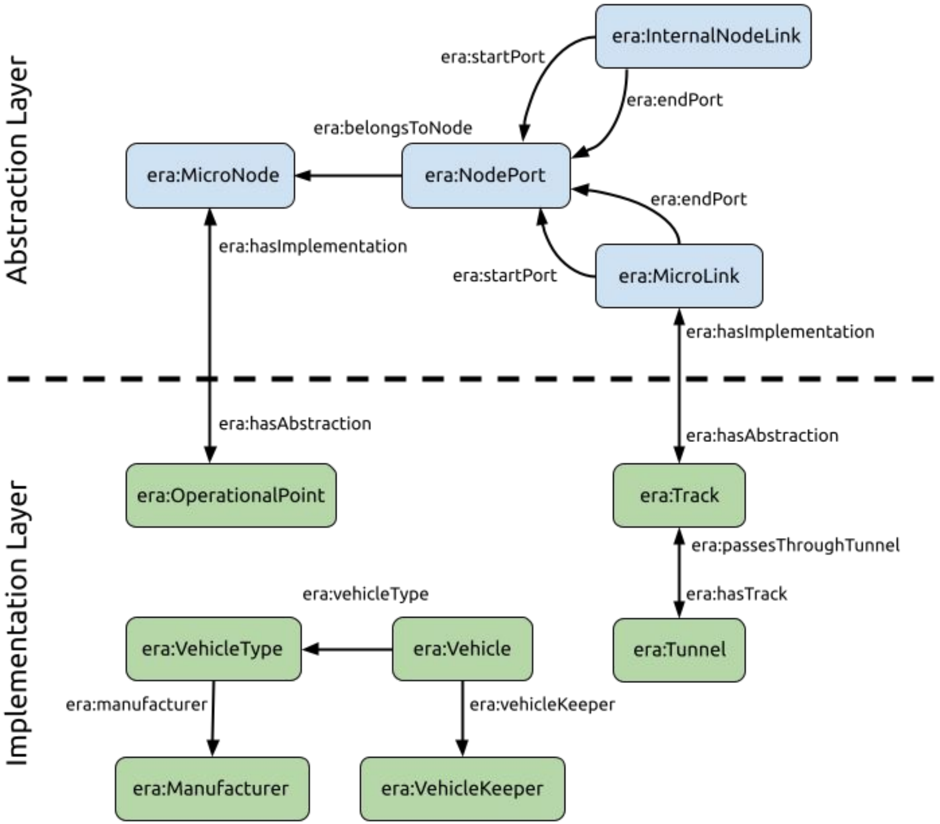


Main tables in ERATV

To this: The ERA Vocabulary v1

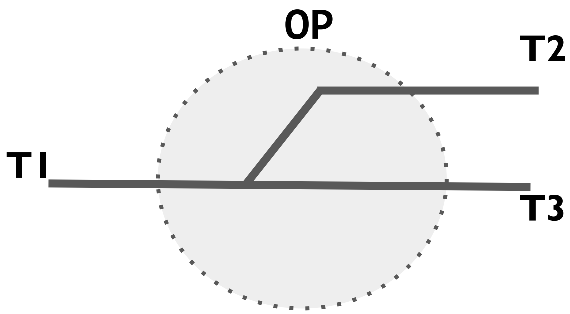


Topology model of the ERA Vocabulary v1

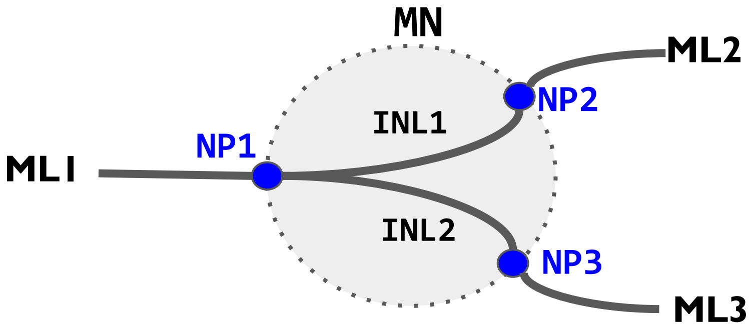


Topology model of the ERA Vocabulary v1

Real-world schematic scenario



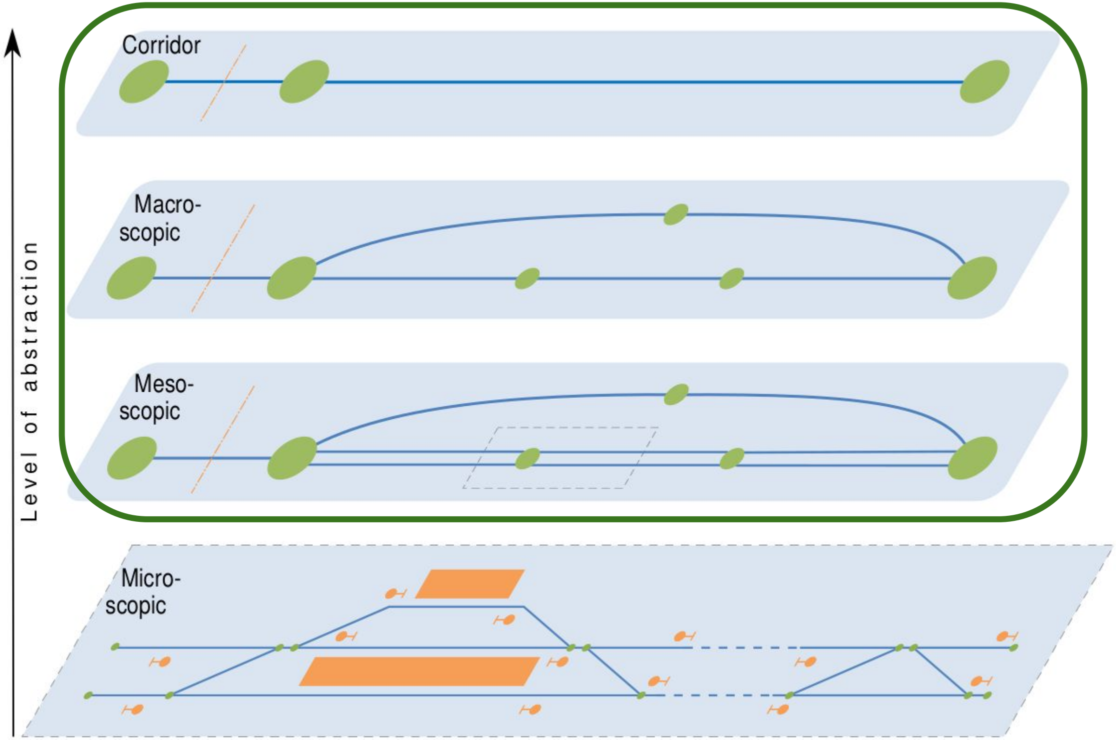
Modelled in the ERA Vocabulary



ERA Vocabulary:

MN: era:MicroNode
MLx: era:MicroLink
NPx: era:NodePort
INLx: era:InternalNodeLink

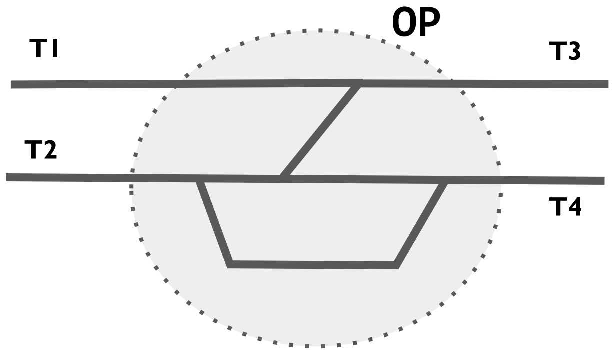
RINF's data model does not provide sufficient level of detail for topological description



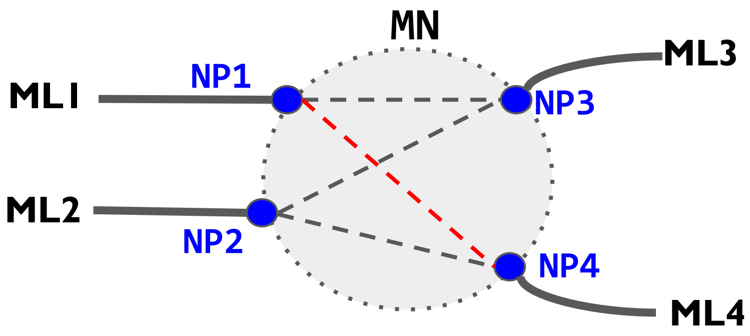
RINF reaches this level of detail

Full connectivity assumption → leads to incorrect topology

Real-world schematic scenario



Modelled in the ERA Vocabulary

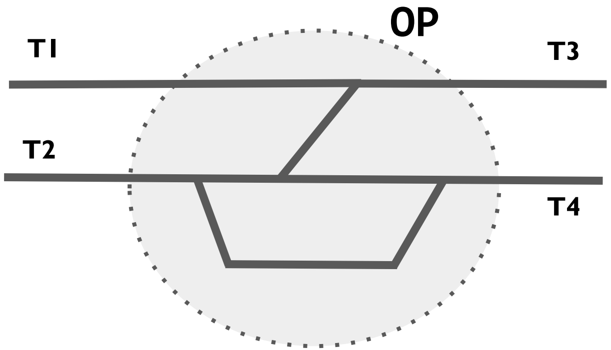


ERA Vocabulary:

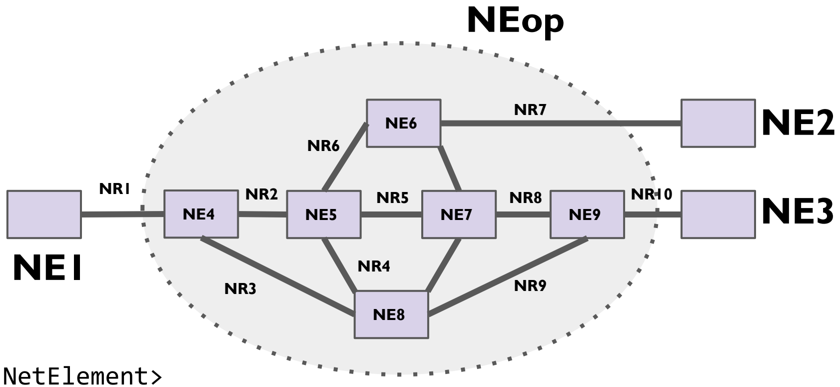
MN: era:MicroNode
MLx: era:MicroLink
NPx: era:NodePort
INLx: era:InternalNodeLink

RailTopoModel standard-based modelling (aka connexity graph*)

Real-world schematic scenario



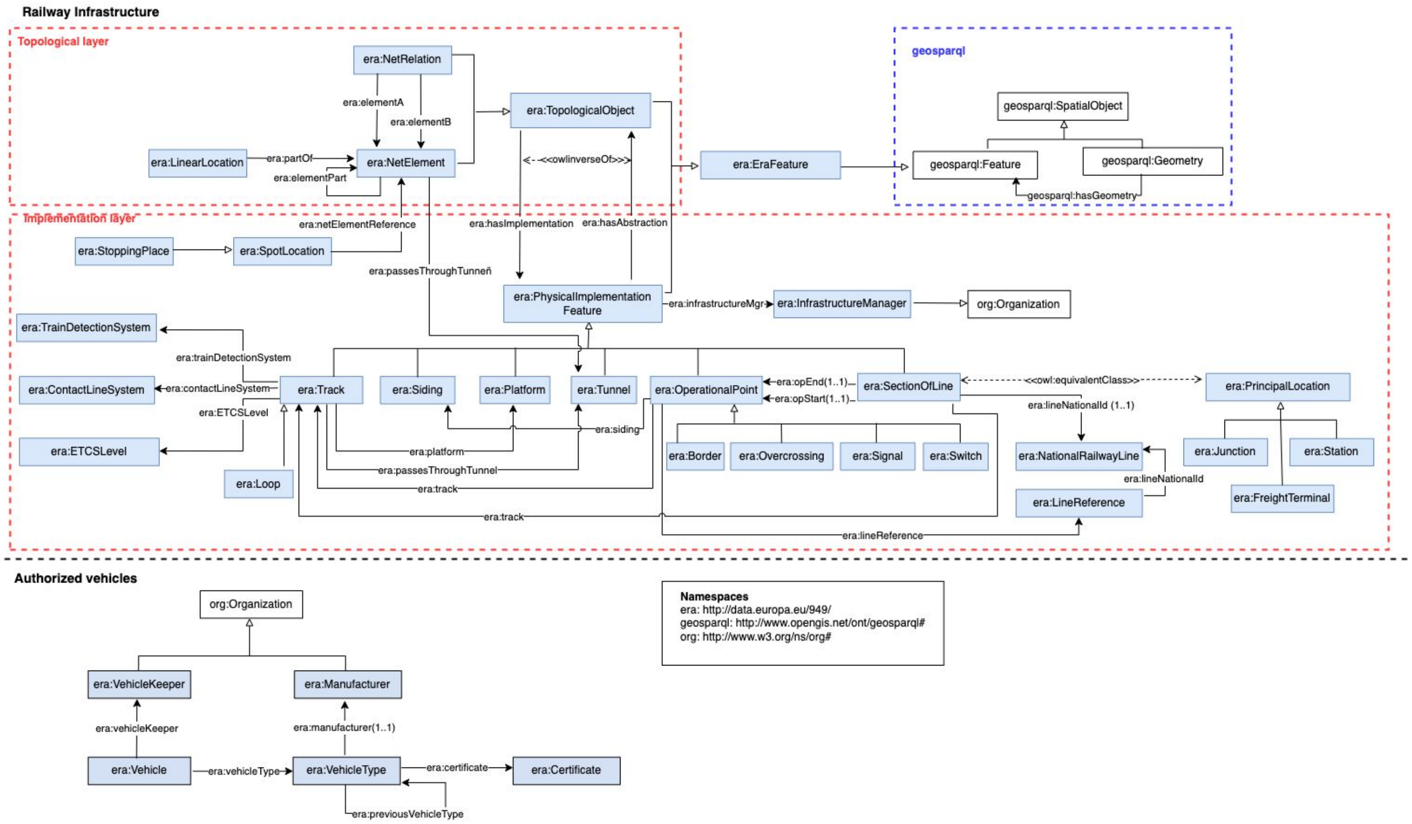
Modelled with RailTopoModel



NEx: <NetElement>
NRx: <NetRelation>

* Gély et al., A multi scalable model based on a connexity graph representation, 2010

Enter the ERA Vocabulary v2



Enter the ERA Vocabulary v2

Ontology Specification

EUROPEAN UNION AGENCY FOR RAILWAYS

ERA vocabulary. Version 2.5.2

This version:
https://data-interop.era.europa.eu/era-vocabulary/

Version:
v2.5.2 (released on 2022-09-05)

Publisher:
European Union Agency for Railways

Download serialization:
Format JSON LD | Format RDF/XML | Format N Triples | Format TTL

Browse SKOS thesauri:
Format HTML

Download SHACL shapes:
Format TTL

License:
License https://creativecommons.org/licenses/by/4.0/

Cite as:
European Union Agency for Railways (2022) ERA vocabulary. Version v2.5.2. Retrieved from https://data-interop.era.europa.eu/era-vocabulary/

6. Acknowledgments

back to ToC

Contributors:
The following individuals have contributed to the development process of this ontology, at different stages of the development process:

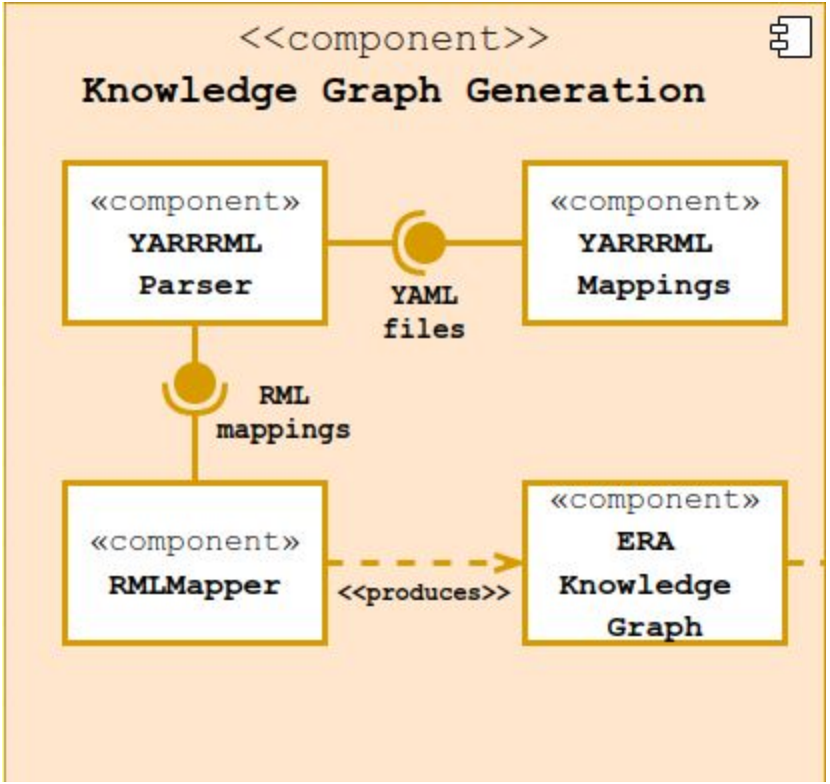
Julián Rojas, (Ghent University - imec)
Dylan Van Assche, (Ghent University - imec)
Ivo Velitchkov, (DG DIGIT)
Marina Aguado, (ERA)
Pieter Colpaert, (Ghent University - imec)
Polymnia Vasilopoulou, (ERA)
Edna Ruckhaus, (UPM)
Oscar Corcho, (UPM)
Wouter Beek, Triply

ACT II: Data mapping and integration

(YARR)RML-based mapping architecture

- A total of 775 YARRRML mappings have been written so far.

- Data sources include:
  - RINF and ERATV relational DBs
  - Norway and Sweden railML (XML)
  - ProRail connectivity data (CSV)
  - OpenStreetMap (XML)



The ERA **Knowledge Graph** in numbers

Live SPARQL endpoint: https://linked.ec-dataplatform.eu/sparql

~40 million triples

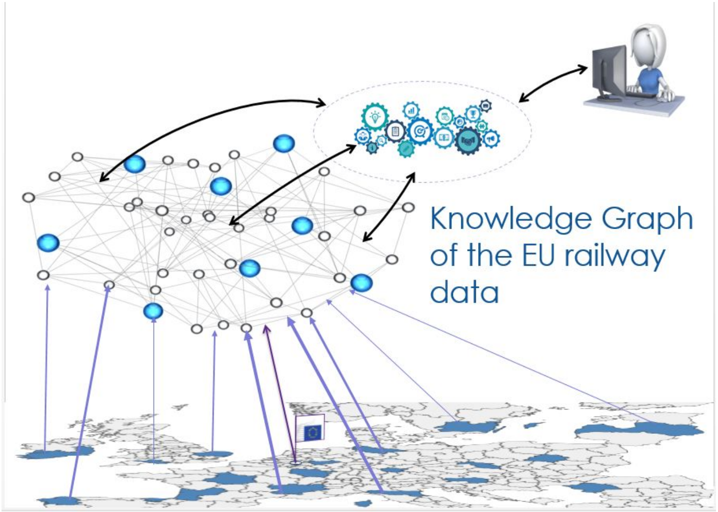
+270k Track segments described

+50k stations (aka Operational Points) described

+50k geo-referenced objects (lat/lon)

+2k Vehicle Types described

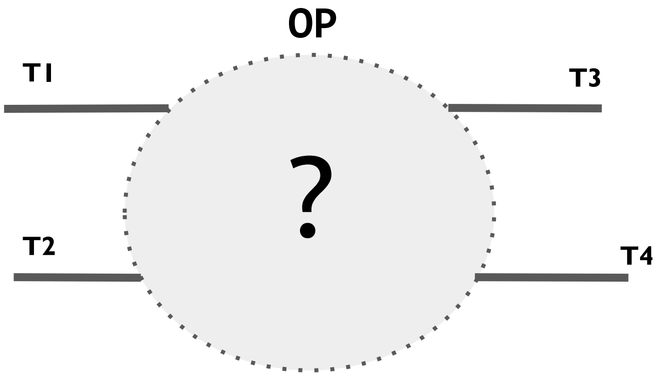
27 countries covered



+50 SKOS taxonomies: https://data-interop.era.europa.eu/era-vocabulary/skos/index.html

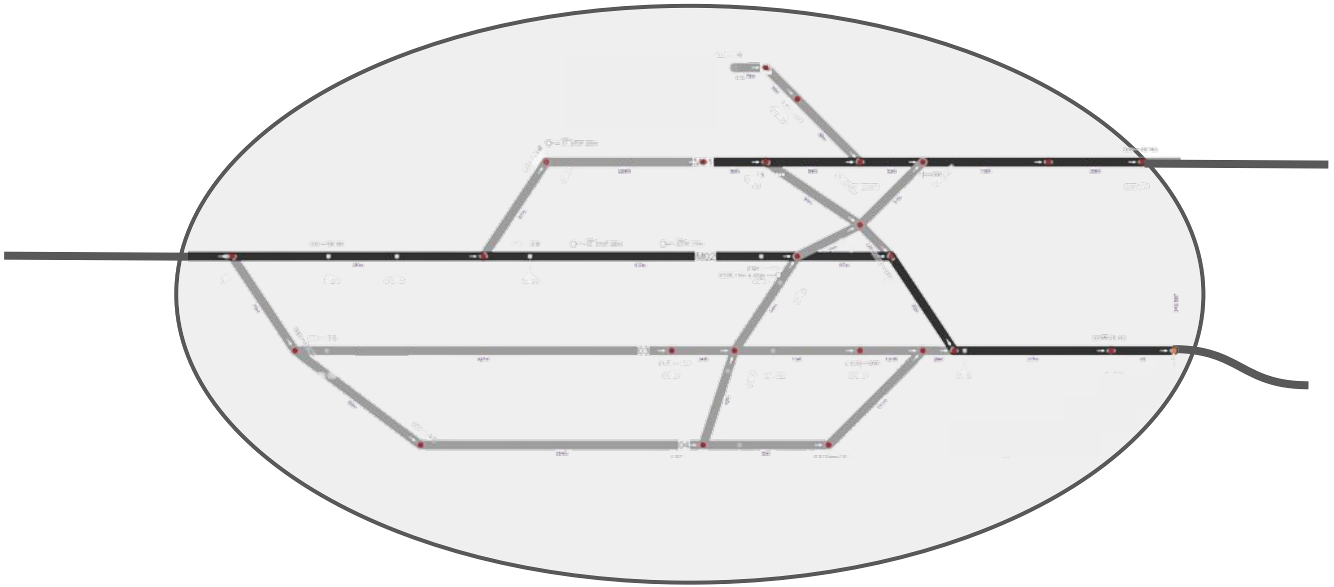
Data integration: micro-level topology with railML data

railML is a **railTopoModel-based** data model that envisions micro-level topology descriptions.



railML.org

For example the schematic topology* of **Dombas Station** in Norway



Dombas Station

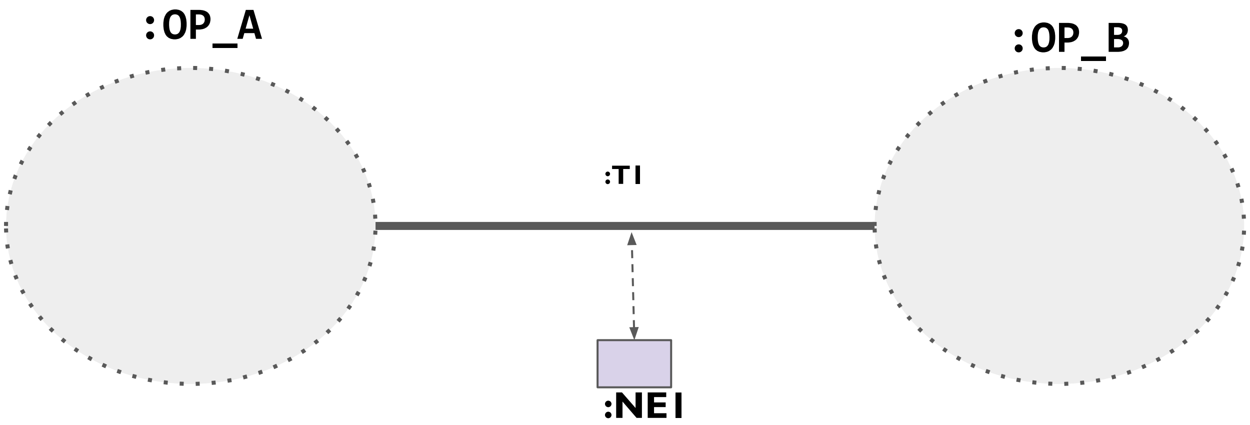
* taken from Norway's railML data

The main mapping challenges are:

- Interlinking RINF-based tracks with railML-based net elements
- Stitch back the topology graph from file based partition

The main mapping challenges are:

- **Interlinking RINF-based tracks with railML-based net elements**
- Stitch back the topology graph from file based partition



The main mapping challenges are:

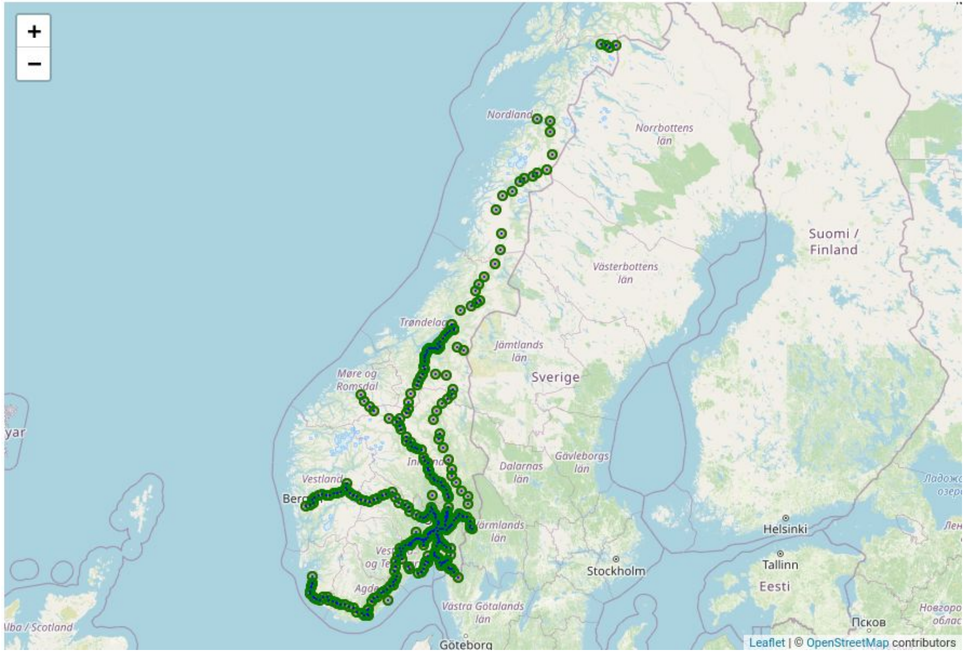
- **Interlinking RINF-based tracks with railML-based net elements**
- Stitch back the topology graph from file based partition

```
source:
    - [data/norway-railml/BRB.railml~xpath, "//netElement[not(@length)]"]
  s: "$(('http://data.europa.eu/949/topology/netElements/' || ancestor::railML/metadata/dc:title || '_' || @id\\))"
  po:
    - p: era:hasImplementation # Link by circumstantial evidence based on the Kilometer points of FROM and TO OPs
      o:
        - mapping: rinf-sections-of-line
          condition:
            - function: idlab-fn:listContainsElement
              parameters:
                - [idlab-fn:str, $(FROM_KM), o]
                - [idlab-fn:list, "$( let $ne_id := @id return (ancestor::infrastructure//operationalPoint[
./@netElementRef = //netRelation[ ./elementA/@ref = $ne_id ]/elementB/@ref ]//@measure\\))", s]
            - function: idlab-fn:listContainsElement
              parameters:
                - [idlab-fn:str, $(TO_KM), o]
                - [idlab-fn:list, "$( let $ne_id := @id return (ancestor::infrastructure//operationalPoint[
./@netElementRef = //netRelation[ ./elementA/@ref = $ne_id ]/elementB/@ref ]//@measure\\))", s]
  graph: http://era.europa.eu/knowledge-graph/banenor
```

The main mapping challenges are:

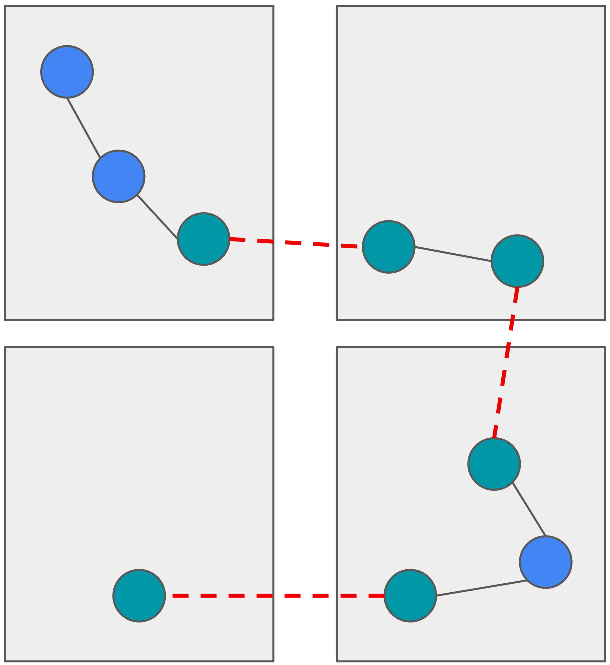
- **Interlinking RINF-based tracks with railML-based net elements**
- Stitch back the topology graph from file based partition

328 links created out of 375 possible ones (87%)



The main mapping challenges are:

- Interlinking RINF-based tracks with railML-based net elements
- **Stitch back the topology graph from file based partition**



The main mapping challenges are:

- Interlinking RINF-based tracks with railML-based net elements
- **Stitch back the topology graph from file based partition**

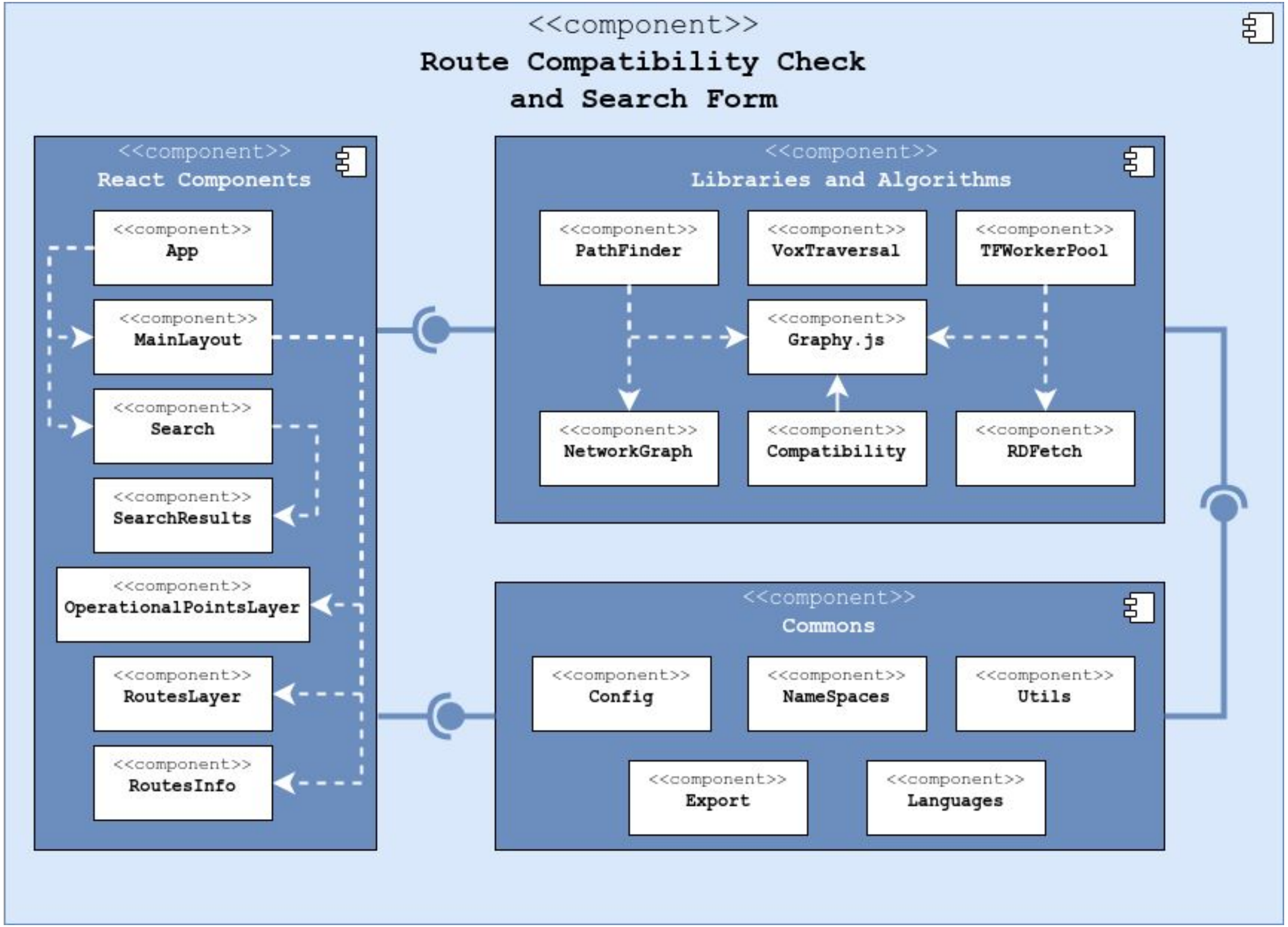
```
CONSTRUCT {
    ?bordNe era:linkedTo ?ne;
            era:partOf ?borderMesoNe;
            era:length ?length;
            geosparql:asWKT ?wkt.
} WHERE {
    ?border a era:Border;
            era:hasAbstraction ?bordNe.

    ?bordNe era:length ?length;
        ^era:elementPart ?borderMesoNe.

    ?mesoNr era:elementA|era:elementB ?borderMesoNe;
            era:elementB|era:elementA ?opMesoNe.
    ...
```

ACT III: Route compatibility check

The app:



<<component>>
Route Compatibility Check and Search Form
<<component>>
React Components
<<component>>
App
<<component>>
MainLayout
<<component>>
Search
<<component>>
SearchResults
<<component>>
OperationalPointsLayer
<<component>>
RoutesLayer
<<component>>
RoutesInfo
<<component>>
Libraries and Algorithms
<<component>>
PathFinder
<<component>>
VoxTraversal
<<component>>
TFWorkerPool
<<component>>
Graphy.js
<<component>>
NetworkGraph
<<component>>
Compatibility
<<component>>
RDFetch
<<component>>
Commons
<<component>>
Config
<<component>>
NameSpaces
<<component>>
Utils
<<component>>
Export
<<component>>
Languages

EUROPEAN UNION AGENCY FOR RAILWAYS

Route Compatibility Check | Search | English

Route Compatibility Check

From: Almere Centrum - NLALM

+ Via point

To: Dronten - NLDRON

11-054-0007-9-001 - TRAXX2E MS-KF4-D-A-B-NL-CZ-SK-H

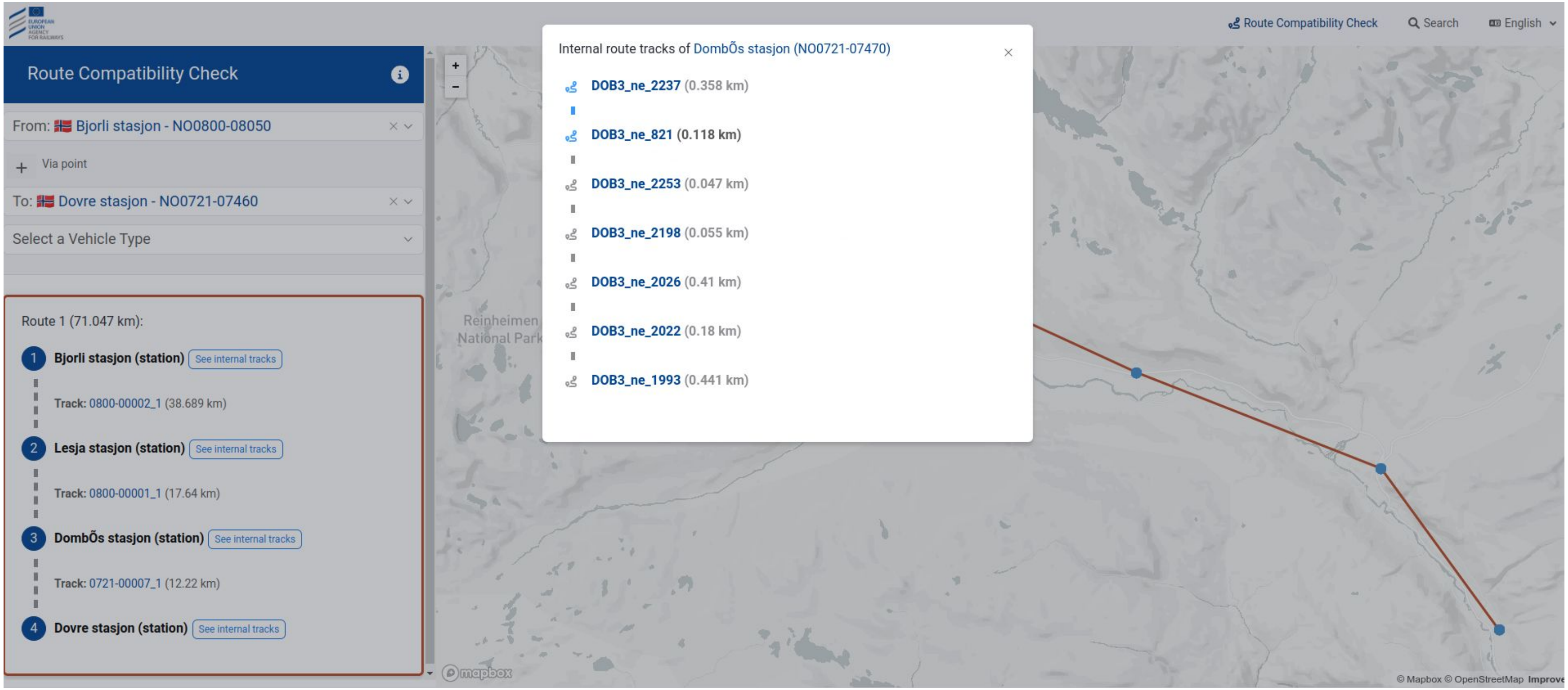
Route 1 (45.466 km):

1 **Almere Centrum (station)**

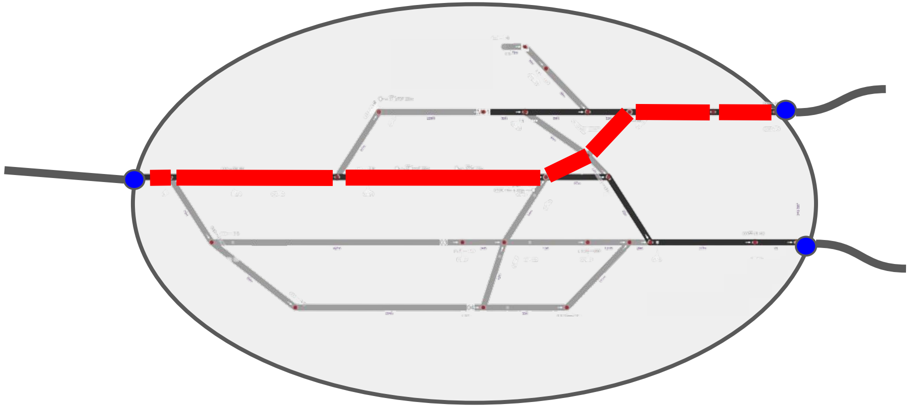
Track: FE
Vehicle Type: 11-054-0007-9-001

Properties	Compatible	Track	Vehicle
• Gauging	NO	NL1	G1
• Type of train detection system	YES	track circuit	• track circuit • wheel detector
• Existence of trackside hot axle box detector (HABD) • axle bearing condition monitoring	NO	false	Detectable by line side
• Rail			

The route calculated through Dombas station



The route calculated through Dombas station



Dombas Station

Internal route tracks of DombÕs stasjon (NO0721-07470)

DOB3_ne_2237 (0.358 km)

DOB3_ne_821 (0.118 km)

DOB3_ne_2253 (0.047 km)

DOB3_ne_2198 (0.055 km)

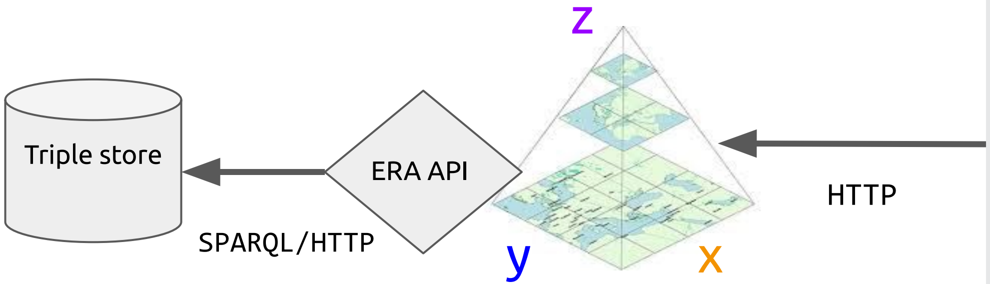
DOB3_ne_2026 (0.41 km)

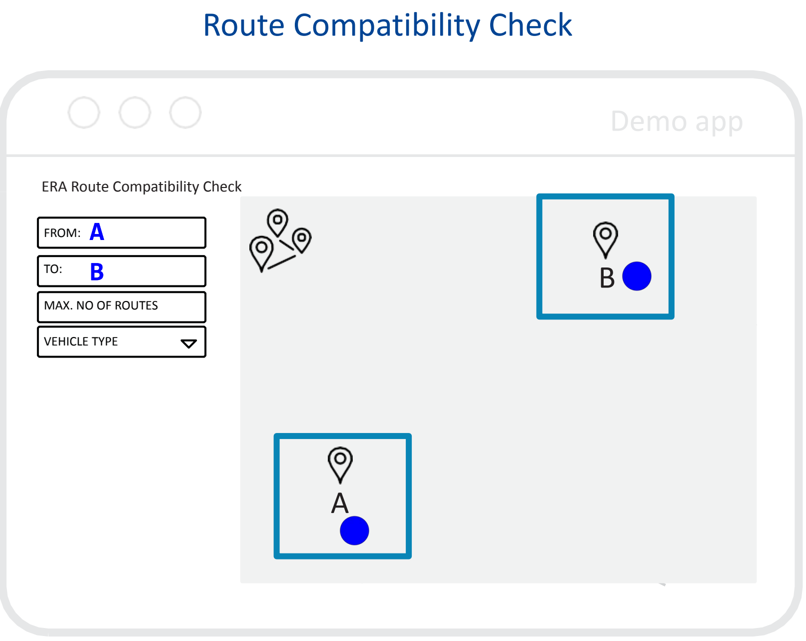
DOB3_ne_2022 (0.18 km)

DOB3_ne_1993 (0.441 km)

Route planning process flow

Step 1: Selection of **FROM** and **TO** Operational Points.





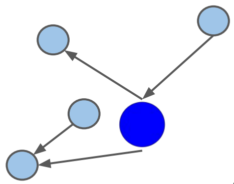
Fetching of (topological) data tiles related to FROM and TO operational points, e.g.:

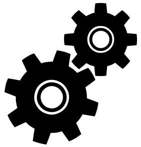
https://data-interop.era.europa.eu/ldf/sparql-tiles/abstraction/10/524/343

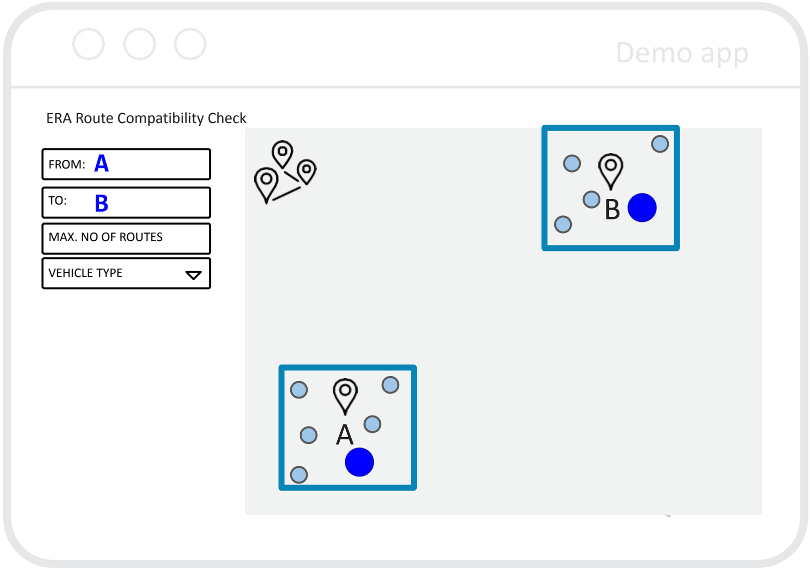
Route planning process flow

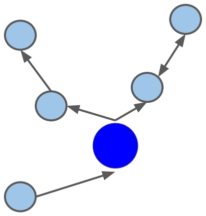
Step 2: Execution of shortest path algorithm (A*) and dynamic fetching of additional data tiles.

Route Compatibility Check



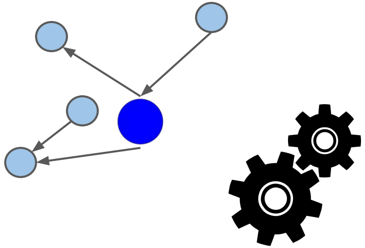


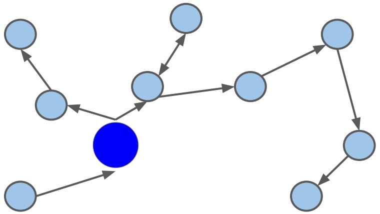




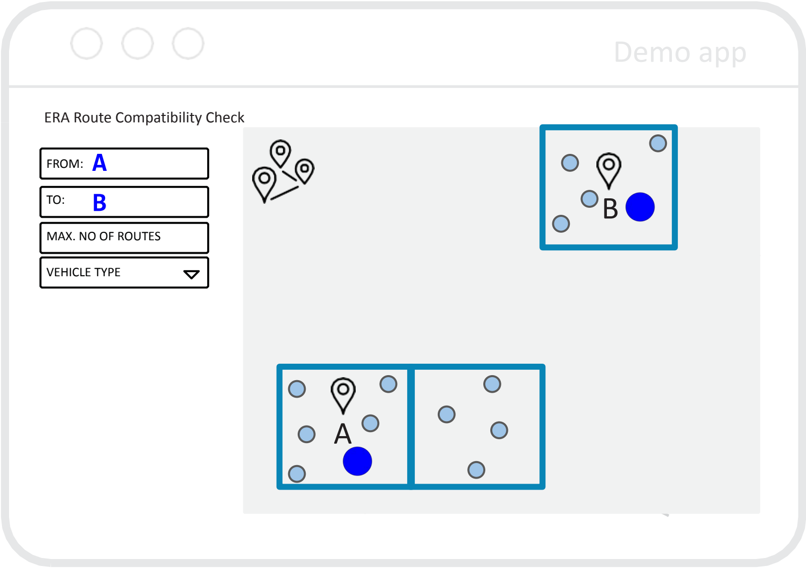
Route planning process flow

Step 2: Execution of shortest path algorithm (A*) and dynamic fetching of additional data tiles.



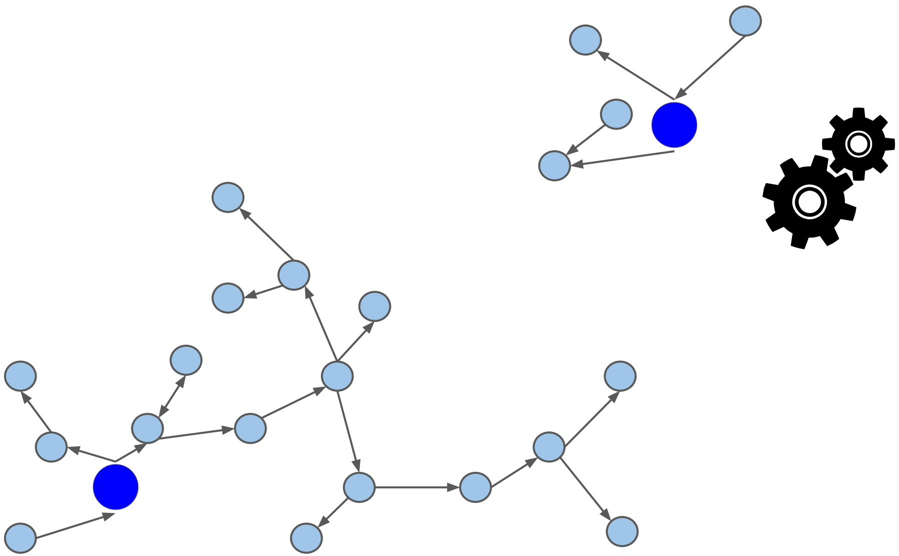


Route Compatibility Check

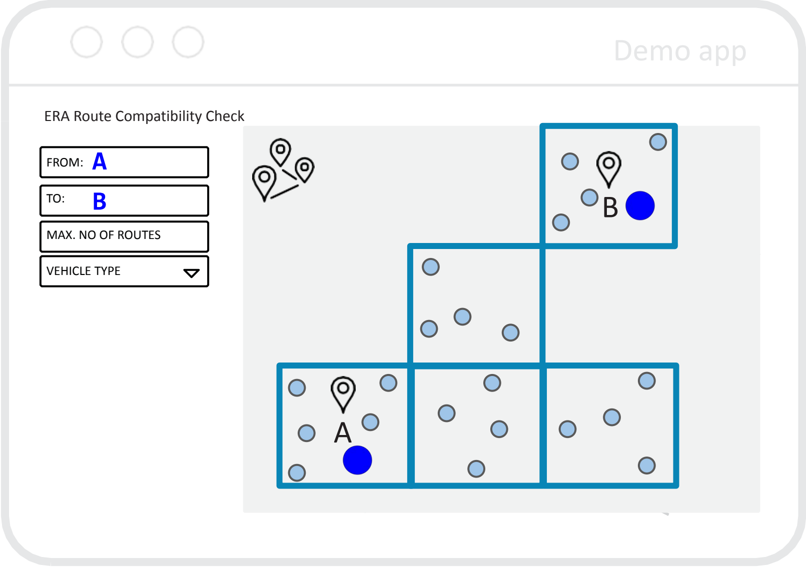


Route planning process flow

Step 2: Execution of shortest path algorithm (A*) and dynamic fetching of additional data tiles.

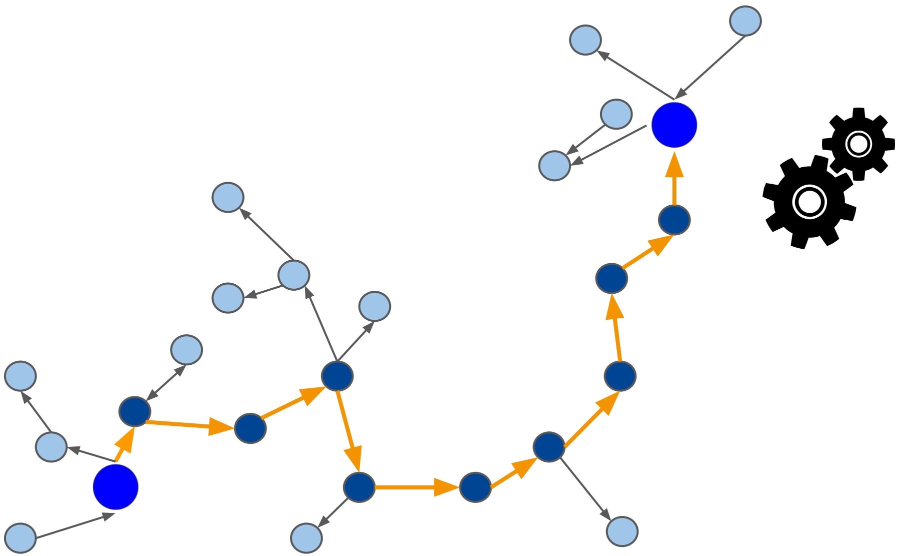


Route Compatibility Check

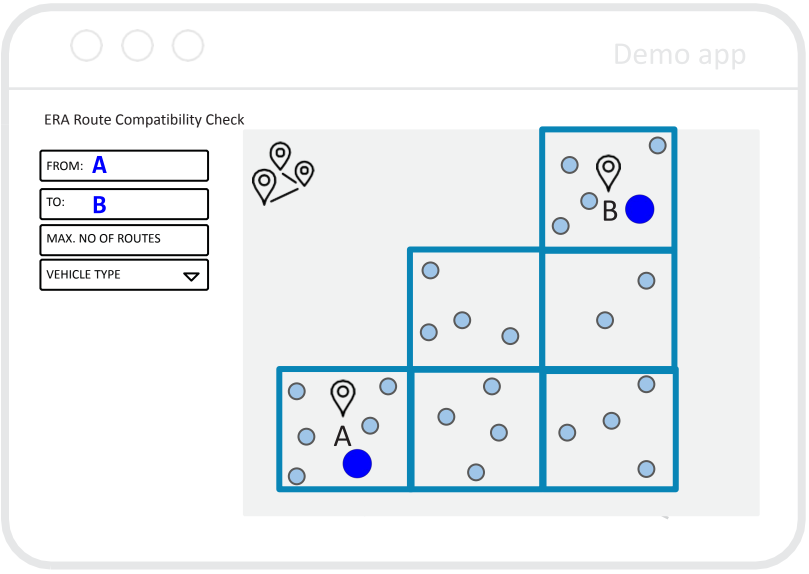


Route planning process flow

Step 2: Execution of shortest path algorithm (A*) and dynamic fetching of additional data tiles.

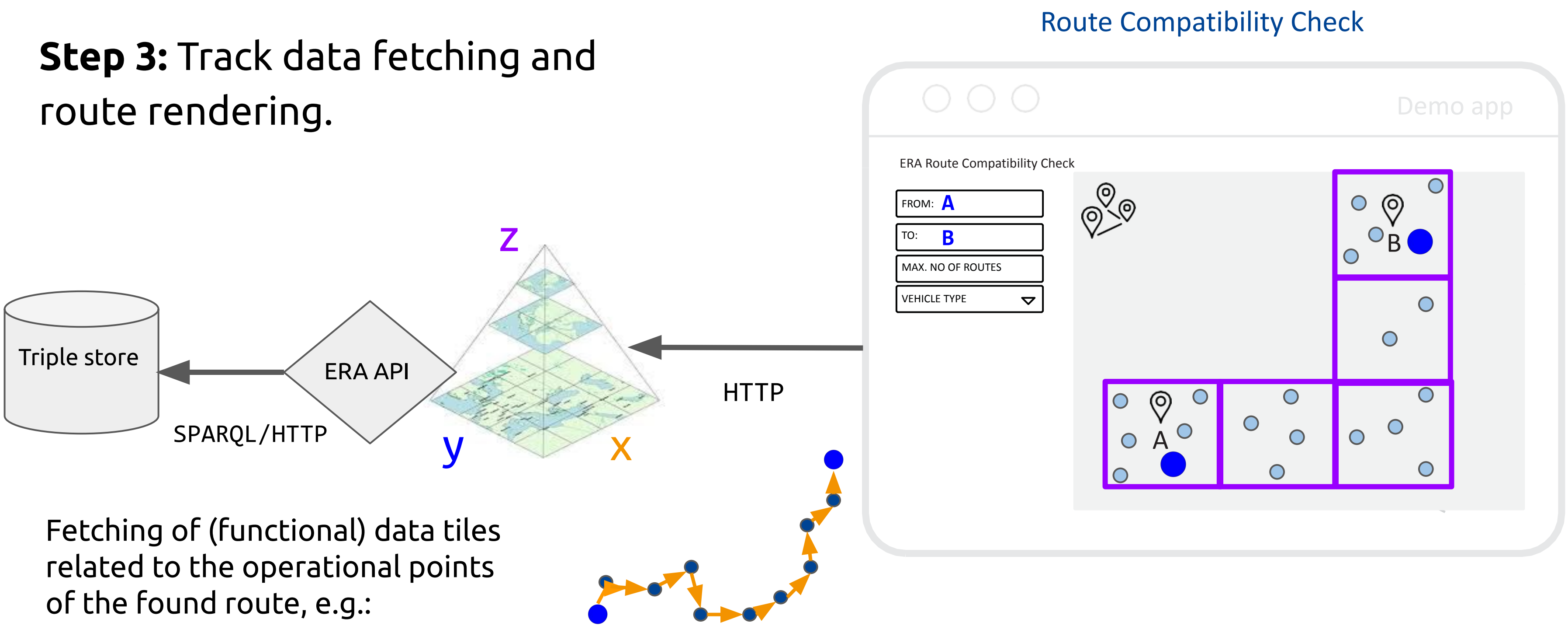


Route Compatibility Check



Route planning process flow

Step 3: Track data fetching and route rendering.

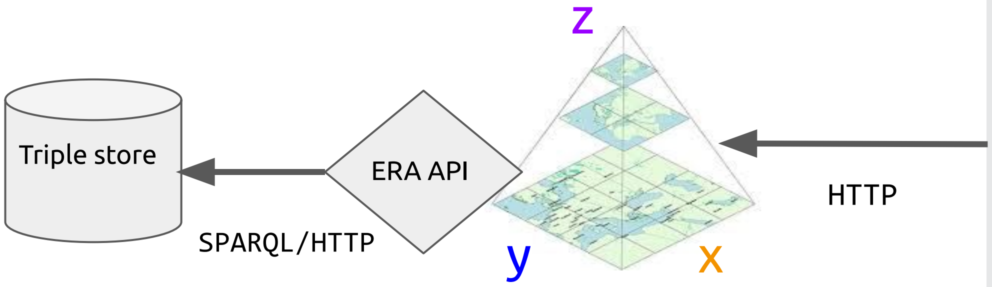


Fetching of (functional) data tiles related to the operational points of the found route, e.g.:

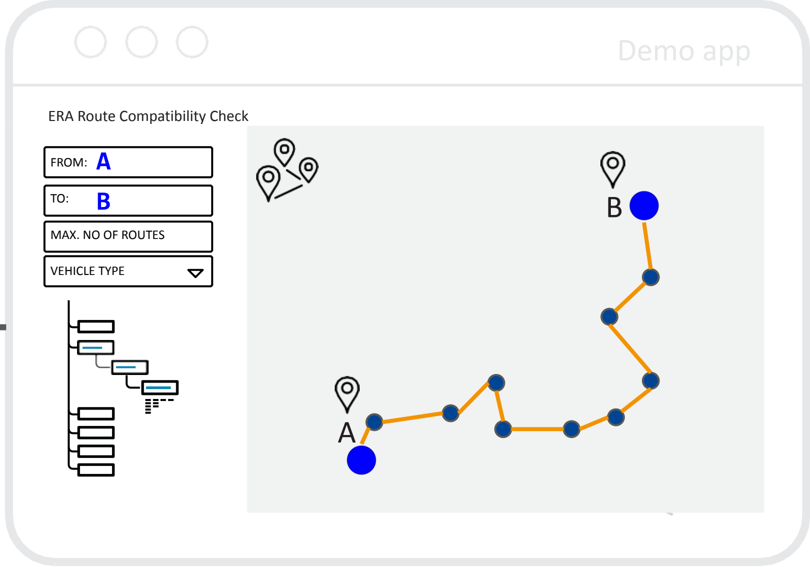
https://data-interop.era.europa.eu/ldf/sparql-tiles/implementation/10/524/343

Route planning process flow

Step 3: Track data fetching and route rendering.



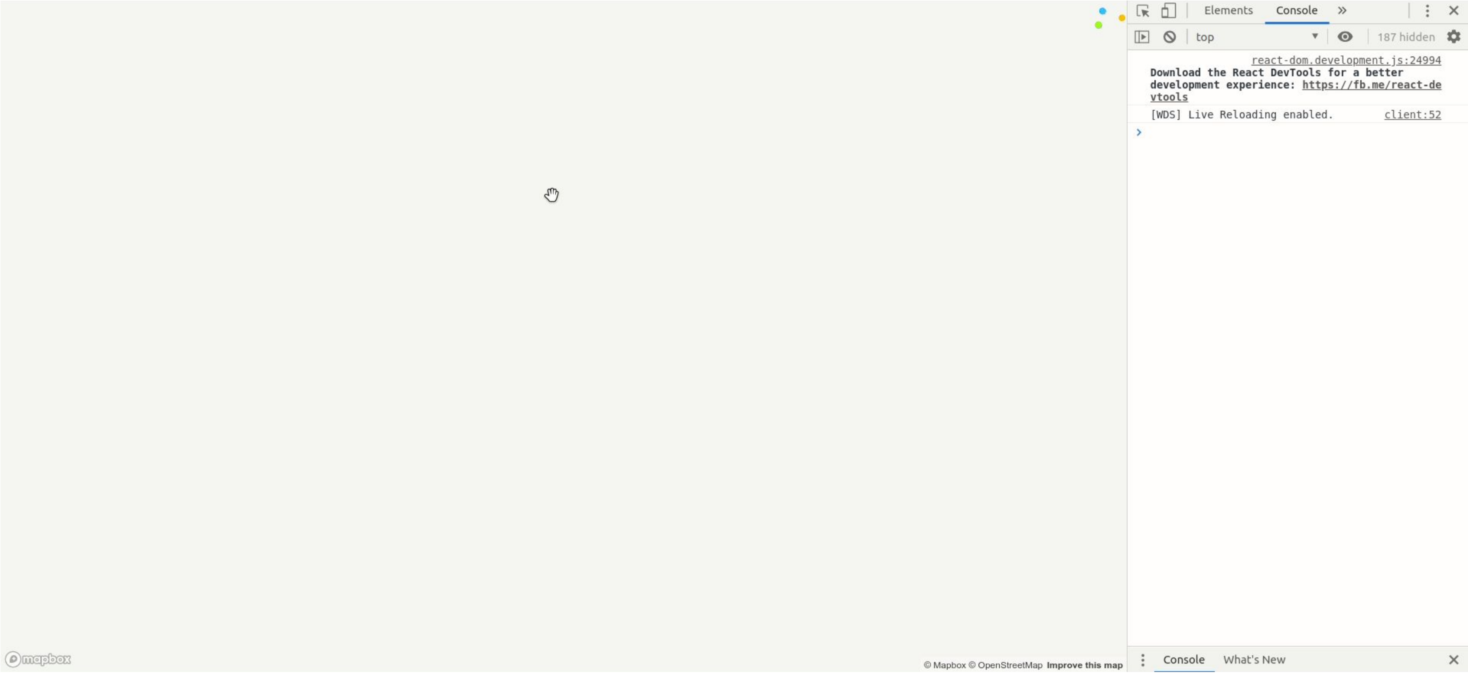
Route Compatibility Check



Fetching of (functional) data tiles related to the operational points of the found route, e.g.:

https://data-interop.era.europa.eu/ldf/sparql-tiles/implementation/10/524/343

Route planning process flow



Pros and Cons of this approach

Pros

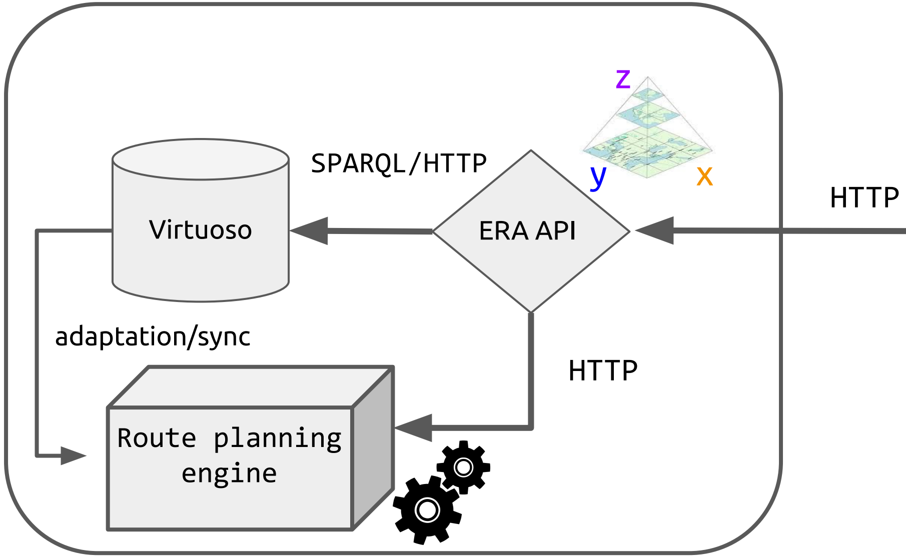
- Low server-side costs
  - High-cacheability of query responses
- Client-side autonomy for implementing customized route calculations

Cons

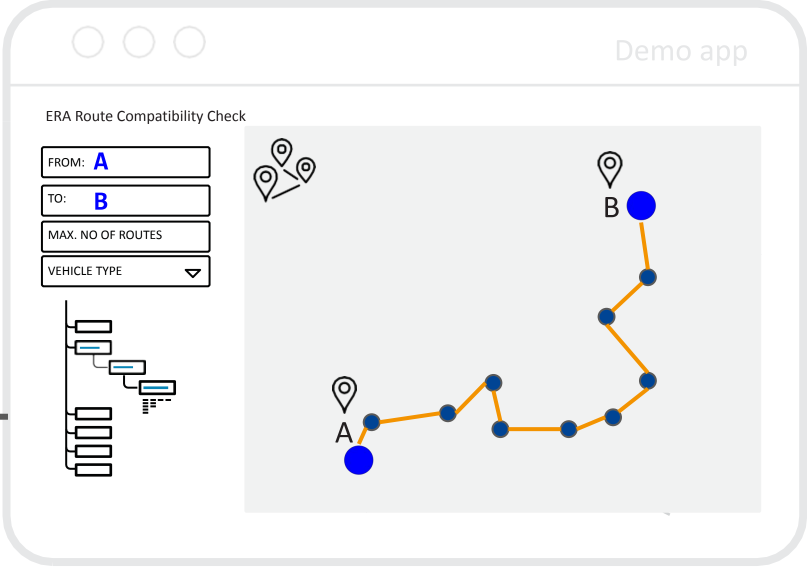
- Severe performance issues on long-distance routes

Potential solutions

Fully server-side route planning with dedicated engine

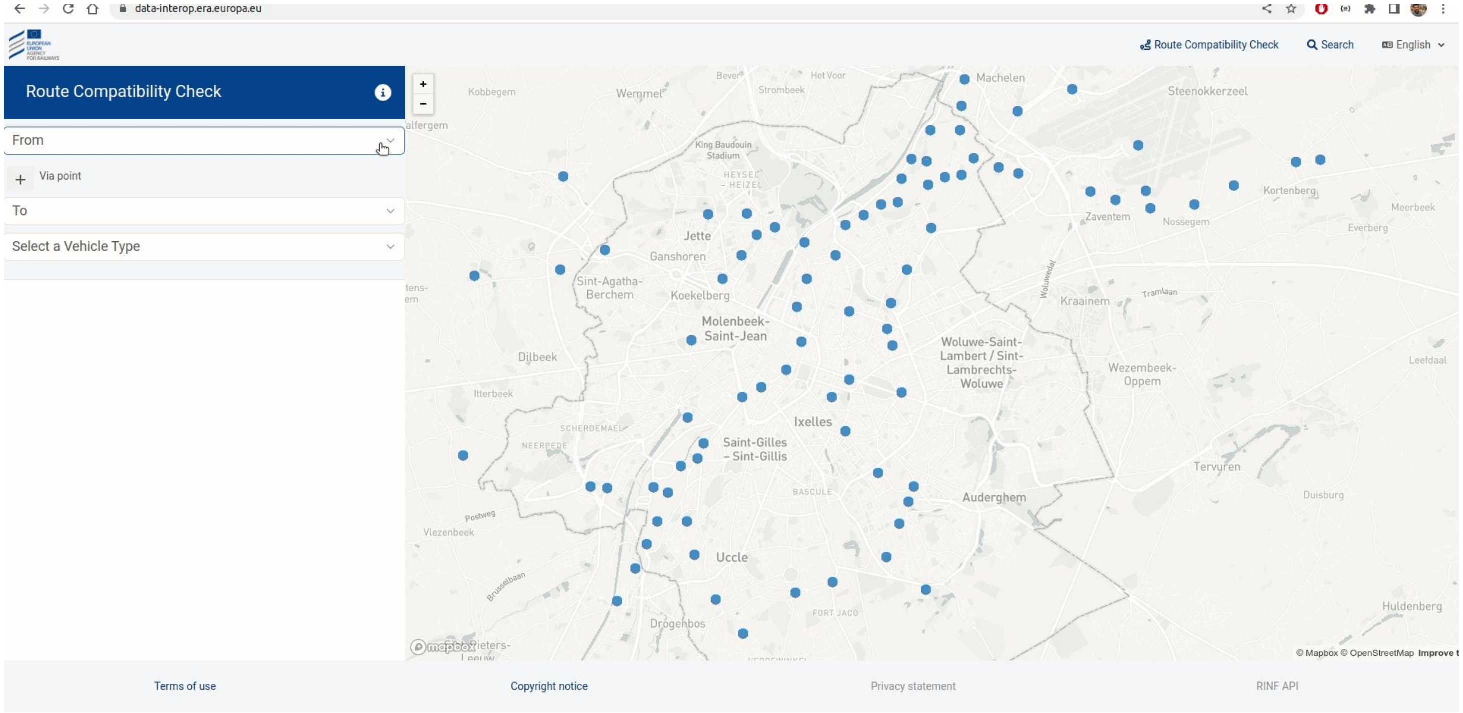


Route Compatibility Check



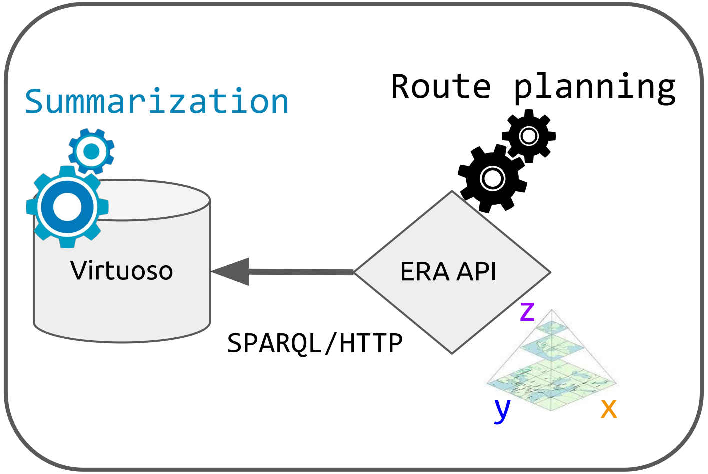
Potential solutions

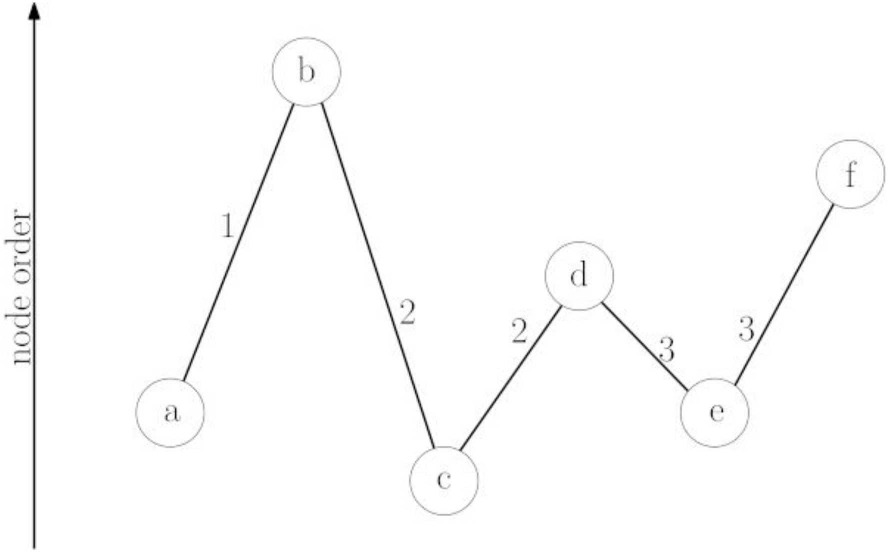
Fully server-side route planning with dedicated engine (OSRM)



Potential solutions

Summarization techniques for route planning (e.g. Contraction Hierarchies)



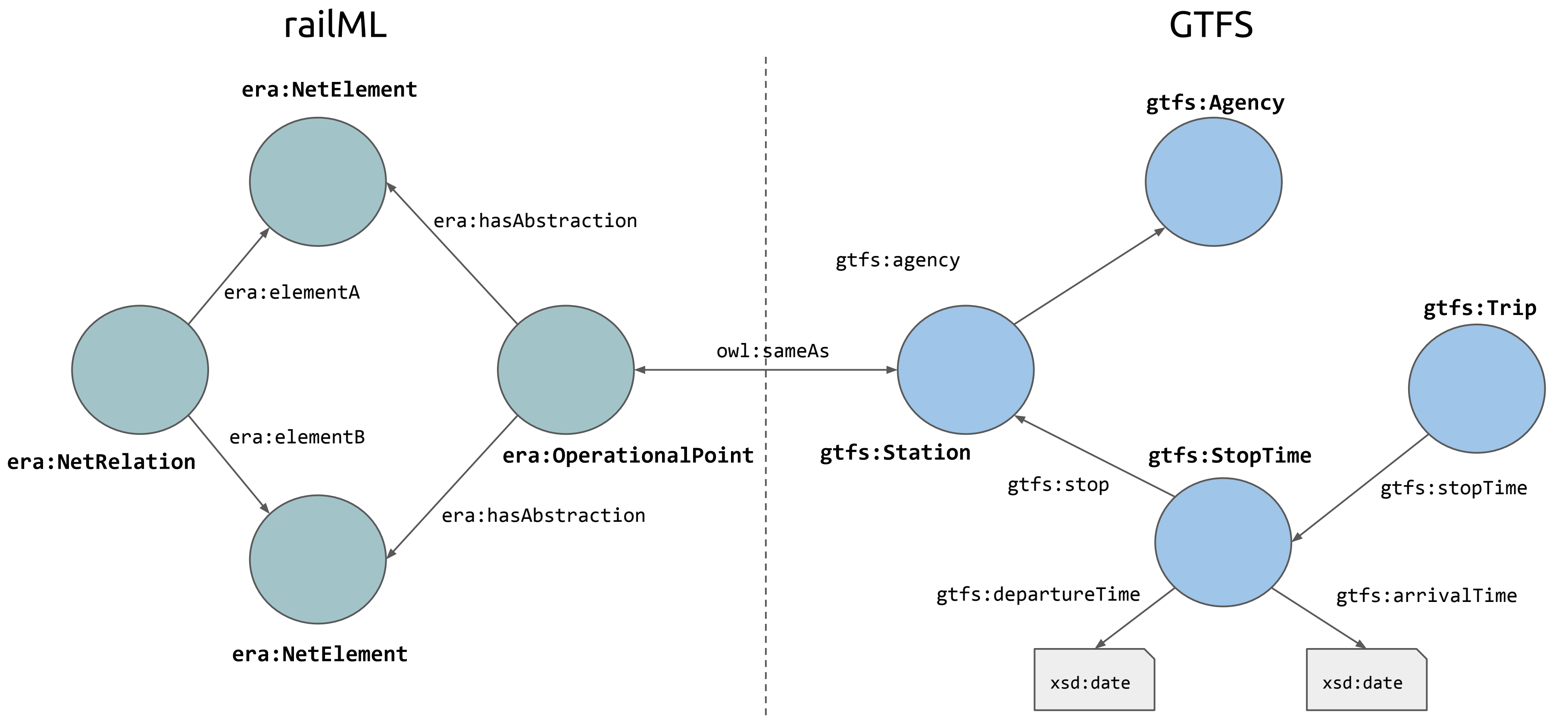


What does the future hold?

Next steps…

- SHACL-based quality assessment
- SHACL rules-based compatibility assessment
- OSM integration for geometry extraction
- Summarization techniques for optimizing route planning queries
- Schedule data integration (e.g., GTFS, NeTEx)
- LDES-based replication architecture

Ontology-based interoperability example for railML and GTFS

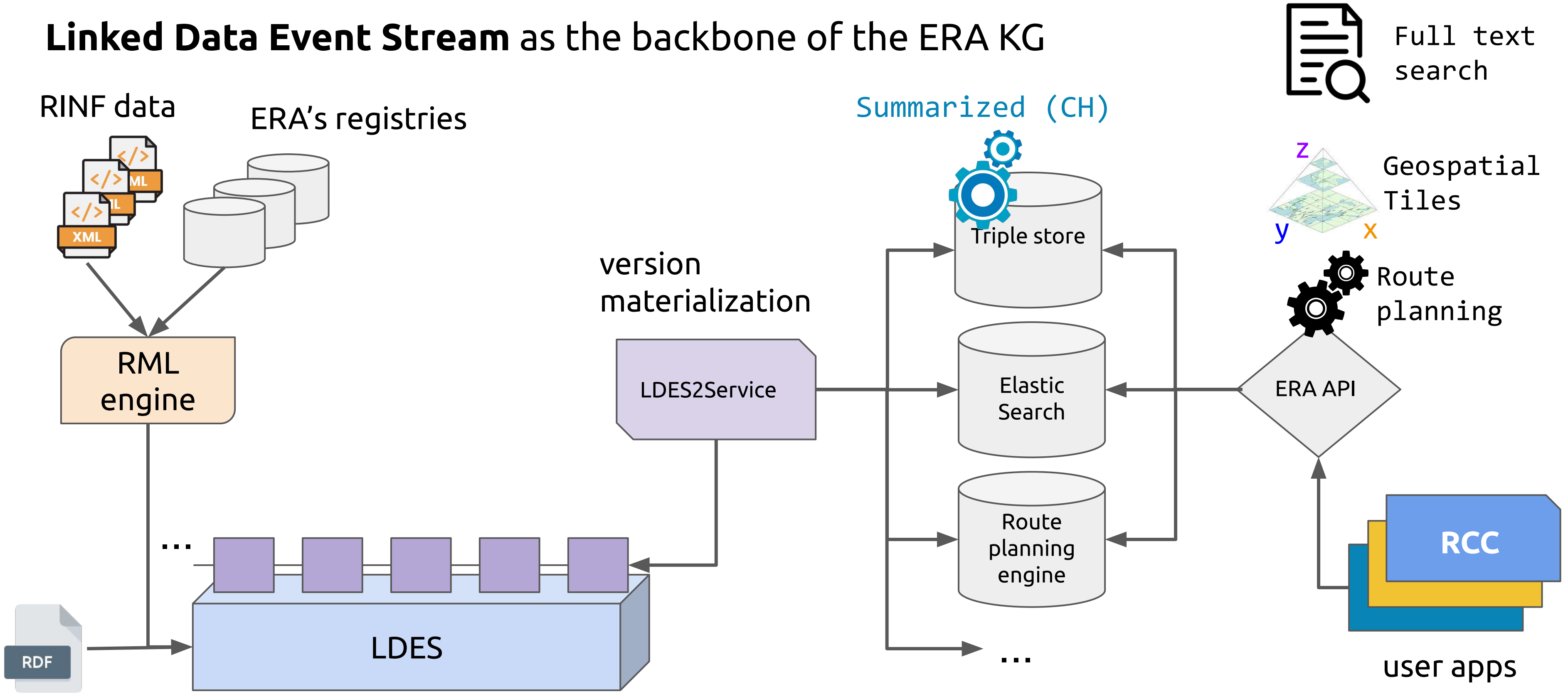


era: http://data.europa.eu/949/

gtfs: http://vocab.gtfs.org/terms#

LDES-based replication and sync

Linked Data Event Stream as the backbone of the ERA KG



Takeaways

- ✓ Thanks to this work ERA set Knowledge Graphs as the default setting for future data endeavors.
- ✓ Knowledge graph technologies can fulfill the requirements of transport data use cases (although there is room for improvement)
- ✓ Balance the trade-offs of data integration (mapping time vs querying time) and query execution (client- vs server-side)

Building a KG for the European Railway Agency: The story

Julián Rojas
Ghent University - imec
@julianr1987
julianandres.rojasmelendez@ugent.be

International Workshop On Semantics And The Web For Transport
SEMANTiCS Conference
September 2022