Benefits and limitations of Linked Data approaches for road modeling and data exchange

Jakob Beetz[1] and André Borrmann[2]

[1] Chair of Computational Design, RWTH Aachen University of Technology, Germany
[2] Chair for Computational Modeling and Simulation, Technical University Munich, Germany

Abstract. The paper motivates the usage of Linked Data approaches by discussing the limitations of conventional monolithic data modeling facing the de-facto heterogeneity of information systems and data models in the diverse fields of digital built environment. It details the discussion by focusing on one specific sub-domain – the application of digital methods in road design, construction and operation which involves the exchange, management and querying of spatio-semantic data that typically stems from different data sources and involves diverse software systems and data models. It is argued that techniques of the semantic web significantly simplify the integration of these heterogeneous data models. In a case study, the German road data exchange standard OKSTRA is linked with the Dutch CB-NL and RWS object type libraries. Doing so, it is shown how nationally well-established and wide-spread standards can be integrated in order to perform cross-country querying of road data. As a basis for the cases study, OKSTRA was transferred into an OWL-based data format, resulting in the creation of okstraOWL. The paper discusses in detail strategies for realizing a semi-automated mapping with the Dutch CB-NL and RWS object type library, including a detailed analysis of the advantages and limitations of the different options. A major emphasis is placed on finding spatially related entities. Finally, the paper discusses the capabilities of linking the data sets by presenting a number of exemplary SPARQL queries for answering real-world questions that require the integrated analysis of data sets in different data models.

Keywords Linked Data, Infrastructure, Interoperability, Standardization, Road modeling.

1 Introduction

1.1 Motivation

For many years, the performance of the AEC industry has been drastically diminished by severe incompatibility problems between the software products employed in practice resulting in the loss of information whenever data is exchanged.

Research to overcome this issue started in the late 1980's. Back then, the development of an all-encompassing product model for every aspect of the building and construction sector seemed to be a viable solution. Based on this idea, the industry consortium buildingSMART International (bSI) has developed the Industry Foundation Classes (IFC) standard for the exchange of digital building models. Today, after 20 years of development, continuing improvement and extensions, the IFC model is the de-facto standard for data exchange in the building industry.

However, the complexity of the data model has already reached a critical limit: With more than 700 classes, thousands of attributes and a dense network of relationships between the classes [1], only a small number of experts is able to fully understand the functionality and behavior of the entire data model.

Nonetheless, even with this high level of complexity, many aspects of the construction domain are still not covered. This particularly applies to the infrastructure sector.

Although significant extension work regarding the infrastructure domain is currently undertaken in the context of the bSI Infra Room, the resulting international standard will not be able to consider all aspects of national or local legacy road information standards, nor will it cover all related information domains, such as traffic or accident data. Taking a broader perspective, it becomes clear that a singular data model will never be able to integrate all relevant data from all domains in all cultural contexts for all application scenarios.

At the same time, it has become common sense in the last years, that the continued extension of the IFC data model without proper modularization, disentanglement and strategic, high-level architectural adjustments would further increase its complexity and reduce its manageability [2].

Facing on the other hand the de-facto existence of a large set of legacy data models in the construction and operation domain, the fundamental concepts of Linked Data seem to provide a promising solution to the challenges described above. It is based on the idea of making use of existing data formats from the different domains mostly as-is, but enabling ways to relate them with each other through links between corresponding objects. The concept of Linked Data is opposing that of one all-embracing product model by respecting the existence of heterogeneous data sources.

Using the example of two national road data models and the international IFC standard, this paper presents how such an integration can be performed from a technical point of view, and subsequently discusses the advantages and limitations of this approach.

1.2 Linked Data

The notion of Linked Data is based on the idea to facilitate the connection, the alignment and ultimately the integration of heterogeneous data and information models by relying on a minimal set of common technologies and protocols: Known as the "Semantic Web Stack", a layered set of technologies is used as a

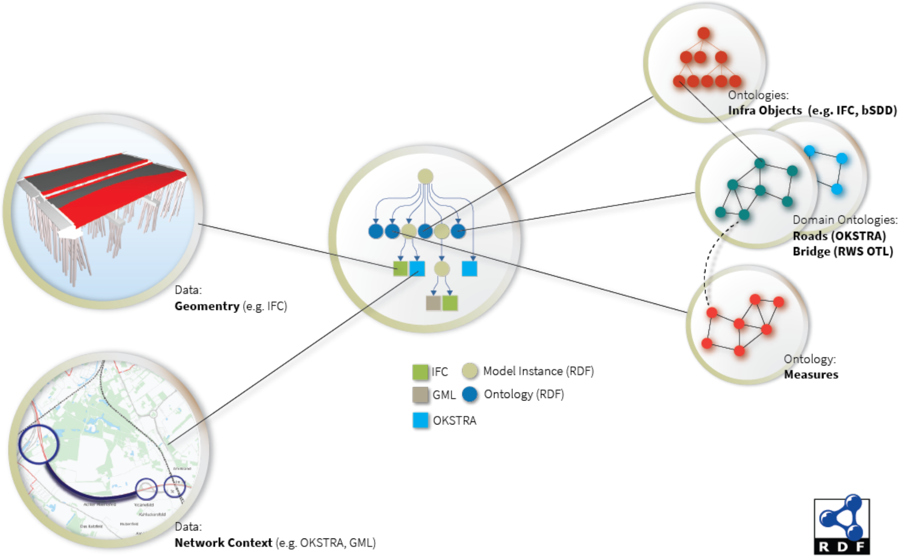

Fig. 1. Overview of the overall concept of multiple Linked Data vocabularies applied to legacy models

common platform to enable the creation of interoperable data models that can be connected across network boundaries.

Built on top of the common technological layers used for the World Wide Web such as HTTP, URIs and XML at the core of the Linked Data and Semantic Web Technology stack is the Resource Description Framework (RDF): In RDF data, information and knowledge is expressed in the form of statements. Each statement is composed of a 3-tuple of resources, referred to as Subject, Predicate and Object (S, P, O). Each of the three resources can either be identified by a Universal Resource Identifier (URI, including URL, URNs etc.) or can be a literal value, usually with a simple data type from XML Schema. By allowing arbitrary statements to be made about arbitrary resources residing at arbitrary locations, graph structures are composed of atomic statements that allow the expression of any data and information imaginable. While basic RDF offers simple constructs such as properties and lists, additional vocabularies from higher levels of the technological stack can be used that allow the uniform use of higher level concepts such as e.g. the notion of classes and subclasses, relationships and value range restrictions. The Ontology Web Language (OWL) vocabulary – stemming from the knowledge modeling domain – introduces formal logics, i.e. concepts that can be used to capture rules and knowledge to allow to make explicit inferences from implicit information referred to as reasoning on the data or the schema itself.

A major challenge for making Linked Data work, however, is the identification and definition of appropriate resources in models and data sets that should be linked together. Ideally, these resources represent the identical physical or logical items or concepts. However, due to the diverging granularity and diverging semantics of the involved data models, identical objects often do not exist. The paper will discuss in detail the different categories of matching problems as well as potential solutions.

1.3 Vocabularies, Concept Libraries and other modular data models

One of advantages of the Linked Data approach is the ability to share meta-models, vocabularies and data sets across different knowledge domains, tools and platforms using homogeneous infrastructures and protocols [3]. In contrast to other data modeling and interoperability standards, Linked Data has the advantage to be self-documenting, queryable and extendable and to support the notion of sharing and reuse of meta-models from the ground up. Using the built-in notion of statements, mappings between different vocabularies both for concepts as well as their relations can easily be added and processed with the same tool chain that is used for the concrete data sets itself. By interconnecting different meta-models and instance data sets, a machine-readable, global web of data has been growing in a decentralized fashion over the last years similar the network of information designed for the consumption by human readers known as the World Wide Web. This global data graph referred to as Linked Open Data Cloud currently consists of hundreds of vocabularies and classification systems that are interconnected. Even though not designed as a central hub, the Linked Data representation of the collaborative WikiPedia corpus, DBPedia [4] is currently the most interconnected data set and serves as a kind of nucleus for the Linked Open Data (LOD) Cloud. A prime example for useful vocabularies shared across engineering domains addresses the modeling of measures, units and quantities.

2 Existing data models for road infrastructure

While the road extensions of the IFC data model are still under development, there is a number of existing well-established data standards for representing and exchanging road data being used today.

2.1 InfraGML, CityGML

In the geospatial sector, two well-established domain models pertaining to infrastructural and urban artefacts including roads exist, that capture information valuable in infrastructural planning. The two standards are based on the

Geographic Markup Language (GML, ISO 19136), an XML Schema-based information model for the spatial geometric description of geographic information that is developed and maintained by the Open Geospatial Consortium (OGC). Interestingly, the GML model has its technological roots in RDF and lends itself very well for Linked Data approaches. This opportunity is increasingly used and numerous software implementations and products with dedicated support for efficient storage and spatial query of large data sets exist. Specialized standards for both geometric descriptions (Well Known Text, WKT, [5]) as well as queries (GeoSPARQL, [6]) have been introduced to further enhance the support for Linked Data.

2.2 LandXML

LandXML is an open information exchange standard focused on modeling terrain, alignments and roads including a dedicated, independent geometry description format. With more than 200 classes and more than 1800 attributes, it is an extensive model allowing a complete and fine grained description of road infrastructure. However, the model does not allow to describe design intent. The earlier versions 1.1 and 1.2 of the format are widely adapted in the road planning and construction domain, but the responsible standardization body LandXML.org has stopped its activities since 2012. An attempt to rescue and reuse the data model by the OGC has led to a first version 2.0 in 2016 which however has not been picked up by software implementers yet.

2.3 OKSTRA

The German road data model OKSTRA (Objekt Katalog Strassen- und Verkehrswesen) allows to describe road infrastructure on different granularity levels. It is a mandatory standard for the data exchange processes in all public road construction projects in Germany. It relies on the GML standard and provides capabilities for describing the shape of a road in compliance with established engineering approaches, i.e. by combining the horizontal and vertical alignment with a number of cross-sections. Apart from that it provides a rich set of semantics covering various aspects ranging from road design and condition rating to traffic and incident statistics.

With more than 1800 classes and almost 14,000 attributes, the extent of the model is enormous. To reduce complexity and allow modular configurations for different use case scenarios, the data model is divided into 41 sub-schemas. Similar to IFC, the original OKSTRA data model was modeled in the STEP EXPRESS language (ISO 10303-11, [7]). To better address contemporary industry practices and tool support however, since version 2.015 the data modeling language has been switched to UML and the defined data model is serialized as XMI [8].

2.4 IFC alignment

The Ifc-Alignment project [9, 10] is the first of the buildingSMART InfraRoom initiatives that has completed the full cycle of standardization and become part of the revision IFC 4.1. It is intended to serve as a common basis for future linear infrastructure extensions including IFC-Road, IFC-Railway and IFC-Bridge that can make common use e.g. of implicit parametric geometric descriptions of curves such as clothoids or Bloss curves. The alignment types have been adapted from existing data models such as LandXML and OKSTRA and integrated into the geometry descriptions of the IFC model.

2.5 RWS OTL, CB-NL

The Concept Library of the Netherlands (CB-NL) is an umbrella vocabulary intended as a bridge or pivot between different classification systems, vocabularies and data models used in the Netherlands. Currently it is comprised of 1557 concepts, among which 590 come from IMGeo (a Dutch CityGML extension), 491 from ETIM (electrotechnical equipment) 293 from NL/SfB (equivalent to Omniclass, Uniclass or DIN 276) and 57 from Rijkswaterstaat Object Type Library (RWS-OTL). The RWS-OTL is a modular ontology describing artefacts in the built environment as they are used and the Dutch Ministry for Infrastructure (RWS).

Overall, the RWS-OTL consists of 7482 concepts that are connected via 10681 object relationships and 22242 specializations (including restrictions, which also make use of subclassing mechanisms). Its current primary application is the use in the context of requirement engineering for infrastructural projects, where concepts and their properties from the RWS-OTL are used to add additional semantics to e.g. geometrical models represented in the form of IFC or GML models. For this, an intermediary format stemming from the COINS [11] standard is used that is currently on track for international standardization as ”Information container for data drop” (ISO/CD 21597-1): Here, concepts and properties from external vocabularies are linked with files providing 3D and 2D geometry of the facility under consideration.

3 Linked Data migration and integration strategies

To employ Linked Data strategies and realize the advantages described in Sec. 1.2 the following strategies can be identified:

1. **Conversion and migration**
2. **Semantic payloads in legacy formats**
3. **Semantic containers**

3.1 Basic considerations: Formal reasoning vs. querying

Among the main advantages of formal ontologies expressed in OWL is the applicability of reasoning mechanisms for checking the consistency and correctness of schema and instance information. Generic inference engines (reasoners) can be applied to prove that the instance data provided in the ABox fulfills the requirements and constraints of the schema information provided in the TBox, see 3.2.

3.2 Linked Data

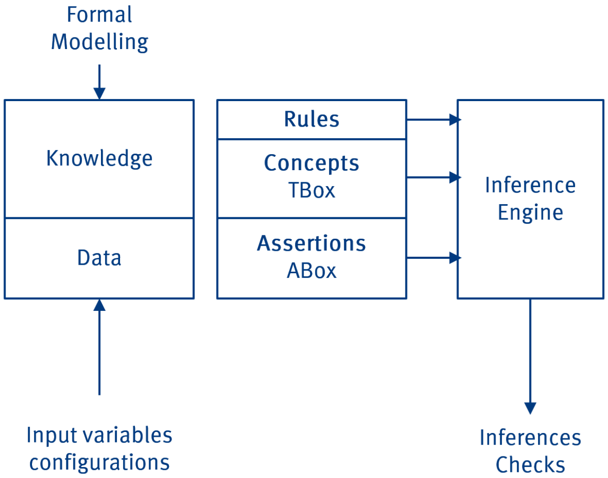

Fig. 2. Schematic depiction of inferences drawn from knowledge models (TBox), facts (ABox) and rules to derive new explicit knowledge from implicit facts

However, applying formal logic reasoning requires a very careful modeling of information and may involve significant computing resources: The effort of proving the consistency of ABox and TBox increases with the complexity of the data model and employed features of OWL for modeling constraints and boundary conditions. Also the extent of facts to be checked contributes to the required computing effort. Depending on the modeling approach taken, proves may not be computed in reasonable time or not be decided at all.

An easy workaround of the challenges involved with applying formal logic and reasoning is the employment of querying mechanisms such as the query language SPARQL. Applying a query language allows to detect schema–instance inconsistencies through the formulation of corresponding queries. Such queries can be processed also for very large data sets in reasonable time, but have the disadvantage of separating parts of the domain-specific logic and knowledge into external systems.

3.3 Challenges for employing Linked Data approaches

The greatest challenge for employing Linked Data approaches lies in the fact that data models with different origins and purposes typically implement different approaches in describing real-world objects and mental concepts. This results in differences in both coverage as well as granularity. While different types of traffic signs might be explicitly represented by individual classes in one data model, they might be subsumed by a more generic class in another one, or not modeled at all in a third one.

As 1-to-1 mappings of concepts can rarely be found in heterogeneous data models, the underlying semantics have to be carefully considered when defining links in order to avoid mis-matching and subsequent erroneous query results.

4 Case study: Linking OKSTRA with other data models

4.1 The OKSTRA data model

The OKSTRA standard is an object-oriented data model for representing road and traffic data. It was introduced by the German Ministry of Transport in the year 2000 as a means for harmonizing data exchange processes between software systems used for designing, constructing and operating roads. It has become a mandatory standard for exchanging road design data between roadway planners and the road authorities. Accordingly, OKSTRA is closely tight to the requirements of the German authorities. As road management in Germany is to a large extent under control of the 16 federal states, OKSTRA provides mechanisms for state-specific extensions.

Originally, OKSTRA was defined using the data modeling language EXPRESS. Starting with Version 2, the modeling language was changed to the more widespread Unified Modeling Language (UML). Accordingly, the schema is now represented in the XML metadata interchange format (XMI) and instance data is exchanged by XML documents.

OKSTRA consists of 41 sub-schemata addressing specific sub-domains of road design, construction and operation. This ranges from detailed design information including horizontal and vertical alignment as well as cross-sections over traffic signs and roadway condition data to traffic and accident statistics. OKSTRA is characterized by a very fine granularity and a high degree of detail. With more than 2.800 complex types and almost 14.000 attributes, the OKSTRA standard is among the most extensive road data models in the world.

A particularity of OKSTRA is its dynamic extension mechanisms. The 'key tables' work similar to the property sets of the IFC standard and allow to introduce extensions and modifications without the need for altering the schema. This provides flexibility, but at the same time reduces the possibilities for checking instance data for formal compliance with the data model. Another mechanism uses IDs to associate geometric objects with semantics provided by long lists of ID-to-concept mappings. These ”domain semantics lists” (Fachbedeutungslisten) differ from state to state. They increase the complexity and represent a significant challenge for implementing linked data approaches.

For representing geometric (spatial) information, OKSTRA makes use of the Geographic Markup Language (GML) which itself implements ISO 19107 Spatial Schema. Accordingly, a large part of geometric information is provided by means of the GML types gml:Point, gml:Curve and gml:Surface or gml:MultiPoint, gml:MultiCurve, gml:MultiSurface, respectively. This is completed by proprietary data types for describing more complex geometries as required for representing the alignment, for example. For describing the shape of a road, OKSTRA is basically providing 2D geometry for representing the horizontal and the vertical alignment as well as the cross-sections. This approach is similar to those implemented in LandXML, InfraGML and IFC-Alignment.

4.2 Making OKSTRA available for Linked Data

In order to make OKSTRA and its instances available for Linked Data techniques and procedures, three steps are necessary:

1. Transforming the schema from EXPRESS, XMI or XSD into RDF, RDFS or OWL
2. Transforming instance data from XML or SPFF into RDF/XML, NTriples or Turtle
3. Making the instance data available through Linked Data infrastructure (e.g. SPARQL Endpoints)

These three steps are discussed in the following sections.

4.3 Transformation of OKSTRA into okstraOWL

A number of research and development efforts can be found that have transformed data models from their underlying legacy schema modeling languages and formats such as NIAM, EXPRESS, or UML into Linked Data formats. The OntoSTEP initiative by NIST has proposed a generic transformation approach for EXPRESS-based models into OWL-DL [12]. Transformations of the IFC model from EXPRESS into OWL have been proposed in different variations in the past [13] [14]. A first version of the ifcOWL transformation has been standardized by the buildingSMART organization in 2016. The transformation approaches chosen for ifcOWL have been applied to OKSTRA and are discussed in the following sections.

As introduced earlier, the transformation process is executed in three general stages. The following sections introduce the conversion of the schema.

URI composition One of the decisive factors for the usability of Linked Data vocabularies is the readability and ease of navigation of the resources used. A number of best practices have been suggested by other research and development initiatives and the W3C. Next to names of classes and properties themselves, offering a de-referencable version of the schema contributes to the adoption and uptake. For this, both a machine-readable and a human-readable version should be provided online that are automatically presented to the consumer via HTTP content negotiation either as HTML documentation (human) or RDF/XML or Turtle file (machine).

Data types: boxed / unboxed When modeling in RDFS or OWL, two main types of data can be identified: Instances of objects and simple literal data types such as integer, float, string etc. In both the original EXPRESS schemata as well as in the current XMI/XSD format both types are specified further using constraints to clarify the semantics and intended use. For example, for measuring length (`Achselement.Laenge:Meter` in `S_Entwurf`), the data type `Meter` is used that is defined as a type of `REAL` (EXPRESS: `TYPE Meter = REAL` in schema S_Allgemeine_Objekte) or `xsd:double` as an extension of the basictypes.xsd definitions in GML. To increase interoperability however, only simple XML Schema types are allowed in RDF and RDFS.

Different approaches to model the data type Meter can be chosen. In okstraOWL, we adopted the approach that was implemented in ifcOWL:

Listing 1.1. Boxed data types for the length measures in meter

```
okstra:Double
     rdf:type owl:Class ;
     rdfs:subClassOf
            [ rdf:type owl:Restriction ;
              owl:allValuesFrom xsd:double ;
              owl:onProperty :hasDouble
            ] .

okstra:Laenge_Achselement
     rdf:type owl:ObjectProperty ;
     rdfs:domain okstra:Achselement ;
     rdfs:range okstra:Meter .

okstra:Meter
     rdf:type owl:Class ;
     rdfs:subClassOf :Double .

okstra:hasDouble
     rdf:type owl:DatatypeProperty ;
     rdfs:domain okstra:Double ;
     rdfs:range xsd:double .
```

Classes and names The translation of the OKSTRA class structure including multiple inheritance can be realized by a simple transformation of `<element xmi` ↪ `:type=uml:Class>` of the XMI schema into `owl:Class` definitions. In OWL, all classes must be declared as being disjoint, i.e. an instances can only be associated with exactly one class.

A critical question of representing large data models such as OKSTRA is the naming of classes. Although classes do have a unique name in OKSTRA, there is no unique name assumption (UNA) in Linked Data. For unambiguous identification, classes can be equipped with the UUID of the XMI model.

Listing 1.2. boxed data types for the length measures in meter

```
okstra:Achse
          rdf:type owl:Class ;
          rdfs:label Achse^^xsd:string@de ;
          rdfs:label Axis^^xsd:string@en ;
            xmi:id EAID_E4106DA8_B309_4197_A6A5_37570759D8B5^^xsd:string
```

Attributes and relations In contrast to 'locally' defined attributes of EXPRESS, XMI and XML Schema classes, definitions of properties and attributes in the form of `rdf:Property` and its subclasses `owl:DatatypeProperty` and `owl:ObjectProperty` are always 'global'. This is not without side effects: The attribute 'Laenge' (length) has different value ranges, types and measures in the context of different classes like meter or kilometer. A globally defined `rdf:Property` can only indirectly constrain the possible values using e.g. `owl:Restrictions`.

Key tables A notable feature of the OKSTRA model is the concept of the 488 core and even more regional 'key tables' (Schluesseltabellen) that require special attention due to the atypical data modeling style. Basically they represent lookup tables that can be used to add additional qualification to classes similar to indirectly modeled subclasses. For historic reasons they can either be modeled as references, embedded into the classes or captured in a compact form [15].

Aggregated Data / Collections Ordered aggregation data types such as arrays or lists cannot be directly translated into OWL. The reason is that RDF and OWL are based on sets of elements. The mapping of ordered data types onto RDF is cumbersome and requires an additional layer to become processable by OWL-supporting inference engines [16], [17]. Due to the low number of ordered collection data outside of geometrical elements which are covered in WKT, optional RDF lists have been adopted for okstraOWL. These can be kept in separate graphs that keep the main data files lightweight and allow DL-compatibility, but can be pulled into the graph on demand.

Integrating GML geometry With the transition from EXPRESS to UML/XMI in version 2.015, the previous geometry entities have been replaced by those of Geographic Markup Language (GML). As the GML format has built upon the Resource Description Framework (RDF) since its very beginning, a migration towards Linked Data is a straight-forward process.

In the approach implemented here, instantiated GML geometry is represented in RDF using "Well-known text" (WKT). The basic idea of WKT is to represent points, lines, linestrings etc. by means of literal values (strings), as otherwise a significant syntactical overhead would be required when using `rdf:list` elements. The following listing shows an example.

Listing 1.3. Usage of "well-known text" in RDF.

```
<sf:LineString rdf:about= "http://example.org/ApplicationSchema#
    ↪ EExactGeom">
     <geo:asWKT rdf:datatype= "http://www.opengis.net/ont/geosparql#
         ↪ wktLiteral">
         <![CDATA[
            <http://www.opengis.net/def/crs/OGC/1.3/CRS84>
            LineString((-83.4 34.0, -83.3 34.3))
         ]]>
     </geo:asWKT>
</sf:LineString>
```

4.4 okstraOWL instance data

In compliance with the conceptual foundations of Linked Data, okstraOWL instance data must fulfill the requirements of distributed networked environments. In contrast to conventional file-based approaches (as implemented by STEP Physical Files or XML instance files), the concept of a resource that can be identified and accessed by an URI plays a central role. Sending or downloading RDF files is only one of the many possible transportation forms.

Serialization The serialization formats available for RDF such as RDF/XML , TURTLE , N-Triples, Notation3 (N3) etc. are content-wise equivalent, but vary in their suitability for different purposes.

- Where a better readability is desired, TURTLE (.ttl) or N3 Notation should be used.
- Where fast import and export is required, N-Triples or N-Quads notations are recommended, as neither the XML Domain Object Model must be constructed nor prefixes resolved. The repetition of the URIs results in significantly increased file sizes, however the resulting files can be effectively compressed and easily archived.
- Through its standard XML structure, XML/RDF can be processed by standard XML (non-RDF) tools which are widely available.

For the reasons mentioned, the okstraXML schema has been published in both Turtle and RDF/XML notation.

Listing 1.4. Short example of the okstraOWL schema as RDF/XML

```
<owl:Class rdf:about="http://okstraowl.org/def/2017/okstraowl#Art_VES
    ↪ .03">
  <rdfs:label>Maximale Achslast</rdfs:label>
  <rdfs:subClassOf rdf:resource="http://okstraowl.org/def/2017/
      ↪ okstraowl#Schluesseltablle"/>
  <rdfs:subClassOf rdf:resource="http://okstraowl.org/def/2017/
      ↪ okstraowl#Art_VES"/>
</owl:Class>
```

4.5 Linking instance data

The creation of okstraOWL and the conversion of instance data provides the basis for linking OKSTRA data sets with other OWL-based data sets. As mentioned above, the focus of the case study lies on linking OKSTRA and CB-NL/RWS data sets.

As discussed above, the key for enabling the integrated analysis of data originating from different models is to find corresponding data items. Here we have to distinguish two matching concepts:

1. Matching on schema level, i.e. matching of objects of the same category or type. For example, RWS objects describing roads must be matched with OKSTRA objects describing road
2. Matching objects that represent the same identical object of the physical world, such as a road segment for example.

Matching in schema level can be achieved through an semi-automated matching based on the similarities of the taxonomy of both data models. A challenge for matching lies in the handling of different granularities of the data models involved, both with respect to the inheritance hierarchy as well as the aggregation relationships. A typical example for the former is the OTL class *Kruisingsconstructie* (engl. crossing construction) with its subclasses *Overbrug* and *Brug* as opposed to the OKSTRA class *Bauwerk* (engl. engineering construction) which does not have any subclasses. A typical example for the latter is the containment structure Weginfrasysteemdeel → Brug → HoofdraagConstructie → Brugdrek as opposed by the rather simple containment structure Bauwerk → Teilbauwerk in OKSTRA.

In the specific case of geometry-oriented data models, matching on instance level can be realized through spatial identification, i.e. objects that cover the same space are candidates for linking. To this end, spatial analysis functionalities can be applied, including identification nearest neighbors or test for spatial containment. In figure 3 the results of such linking processes between heterogeneous data sets are depicted showing school buildings near highways using queries on federated data sets that are introduced in the following section [18].

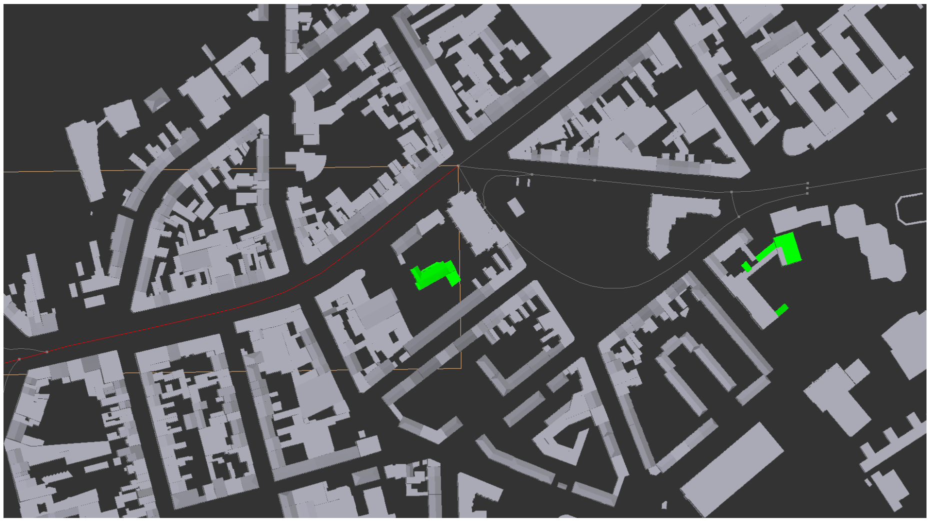

Fig. 3. Visualization of heterogeneous data sets in federated queries. Here schools (green) that are closer to highways sections (red) than 100 meter. [18]

4.6 Integrated analysis: Sample queries

Two main use case scenarios have been identified to test the added value and viability of Linked Data approaches for road data captured in the OKSTRA format:

1. **Federated Queries** In this use case, a heavy goods transport should be planned across borders of federal states. As a requirement, the road network structure should be queried across the individual data sets from the respective federal authorities and should span both network information pertaining to the topology as well as detailed information regarding e.g. material qualities of the pavement, alignment curvatures etc.
2. **International Mappings** For use cases where other highway models should be integrated, the capability of Linked Data approaches should be tested that would allow to extend data processing scenarios within uniform models as described in use case 1 to heterogeneous models like Dutch road models using the RWS-OTL.

Approaches of mapping different data sets that are 'lifted' to RDF have been implemented in the context of this research in [18]. An example is provided in listing 1.5: Here, instance data sets of the proposed okstraOWL Linked Data format have been transformed for an inner city area of the town Aachen. For the investigated area, both a CityGML LOD 2 model as well as OKSTRA data from the road authority of the federal state of North-Rhine Westphalia are available and have been transformed into RDF. The GeoSPARQL language extension that has been implemented in a number of triple stores with SPARQL endpoint allows to perform spatial analyses on the fly. In the example, all buildings with a building height greater than 20 meter (provided as an explicit attribute in the CityGML model) are selected in the target area that are within 50 meter range of one of they highway-sections (German: Abschnitt) within the city limits.

Listing 1.5. A short example query using okstraOWL data in combination with CityGML data in an RDF format to query for tall buildings near highways

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX gml: <http://www.opengis.net/gml:>
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 xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX bldg: <http://www.opengis.net/citygml/building/1.0>
PREFIX okstra: <http://schema.okstra.de/2016/okstra:>
PREFIX co: <http://citygmlinteokstra.tue.nl/>


SELECT DISTINCT ?building ?height
WHERE { ?building rdf:type bldg:building .
       ?building geo:hasGeometry ?ageo .
       ?ageo geo:asWKT ?awkt .
       ?building bldg:measuredHeight ?height .
              ?b geo:hasGeometry ?bgeo .
       ?b rdf:type <http://schema.okstra.de/2016/Abschnitt> .
       ?bgeo geo:asWKT ?bwkt .
       FILTER (geof:distance(?awkt, ?bwkt,uom:metre) < 50)
       FILTER (xsd:double(?height) > 20)
       FILTER (?ageo != ?bgeo) .
}
```

5 Summary, discussion and outlook

This paper has presented an approach for integrating diverging road data models using Linked Data concepts and techniques. The necessity of applying linked data approaches has been motivated by the fact that there is a large set of national data models in practical use, while new international data models are on the horizon, but not yet widely implemented. These data models are different in their purpose, structure, and extent. Experiences of the last decades have shown that it is practically impossible to create a singular data model that covers all relevant data from all domains and all nations. By applying linked data approaches it becomes feasible to analyze the data represented across different data models in a integrated manner.

The approach has been demonstrated by the case study on the basis of the German road data model OKSTRA. The goal of applying linked data approaches is to allow for cross-national analysis of road and traffic data. A pre-requisite for applying Linked Data approaches to OKSTRA is its conversion into an OWL-based data model. The involved challenges and the decisions taken have been discussed extensively in the paper. With the availability of resulting okstraOWL schema and the corresponding instance data sets it becomes possible to link corresponding data objects and perform cross data-model queries.

The case study has shown promising results. However, also the challenges and limitations become apparent. In order to make Linked Data approaches work effectively and generate correct query results a semantic alignment is required where semantically equal oder similar elements of both are associated with each other. For similar models the matching items can be identified using semi-automated matching methods. If the information coverage and granularity differs strongly between the models, a manual matching must be performed.

Acknowledgment

The research presented in this paper has been funded by the German Federal Highway Research Institute (BASt) under grant FE 01.0195/2015/AGB.

References

1. Amor, R., Jiang, Y., Chen, X.: BIM in 2007 - are we there yet. In: Proceedings of CIB W78 conference on Bringing ITC knowledge to work, Maribor, Slovenia (2007) 26–29
2. Laakso, M., Kiviniemi, A.: The IFC Standard – A Review of History, Development and Standardization. ITcon Journal of Information Technology in Construction **17** (2012) 134–161
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. Semantic services, interoperability and web applications: emerging concepts (2009) 205–227
4. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia – a crystallization point for the web of data. Web Semantics: Science, Services and Agents on the World Wide Web **7**(3) (Sep 2009) 154165
5. Open Geospatial Consortium, year=2015, m.: Geographic information – well-known text representation of coordinate reference systems
6. Open Geospatial Consortium: GeoSPARQL – a geographic query language for RDF data. (2012)
7. ISO10303-11:1994: Industrial automation systems and integration - Product data representation and exchange - Part 11: Description methods: The EXPRESS language reference manual. (1994)
8. Object Management Group: XML Metadata Interchange (XMI) Specification. (2015)
9. Amann, J., Singer, D., Borrmann, A.: Extension of the upcoming IFC alignment standard with cross sections for road design. In: Proc. of the ICCBEI 2015, Tokyo, Japan (2015)
10. Amann, J., Borrmann, A.: Embedding procedural knowledge into building information models: The IFC Procedural Language and its application for flexible transition curve representation. Journal of Computing in Civil Engineering **30**(4) (2016) 1–14

11. Willems, P., Schaap, H.: COINS 1.1 specs (Release 15/12/2014) - CoinsWiki
12. Krima, S., Barbau, R., Fiorentini, X., Sudarsan, R., Sriram, R.D.: Ontostep: OWL-DL ontology for STEP. National Institute of Standards and Technology, NISTIR **7561** (2009)
13. Beetz, J., van Leeuwen, J., de Vries, B.: IfcOWL: A case of transforming EXPRESS schemas into ontologies. AI EDAM **23** (2009) 89–101
14. Pauwels, P., Terkaj, W.: EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. Automation in Construction **63** (2016) 100–133
15. Hettwer, J., Portele, C.: T0006 ableitung von OKSTRA?-XML aus der EXPRESS-modellierung (2011)
16. Drummond, N., Rector, A.L., Stevens, R., G, M., Horridge, M., Wang, H., Seidenberg, J.: Putting OWL in order: Patterns for sequences in OWL. In: Proceedings of 2nd OWL Experiences and Directions Workshop, Athens. (2006)
17. Pauwels, P., Terkaj, W., Krijnen, T., Beetz, J.: Coping with lists in the ifcOWL ontology. In: 22nd EG-ICE International Workshop. (2015) 113–122
18. Zheng, Y.: Improving the interoperability of between city and infrastructure information. an integration of citygml and okstra data based on semantic web and linked data technology. Master's thesis, Eindhoven University of Technology (2017)