Technische Universität München
Lehrstuhl für Computergestützte Modellierung und Simulation
Ingenieurfakultät Bau Geo Umwelt

Rheinisch-Westfälische Technische Hochschule Aachen
CAAD – Design Computation
Fakultät für Architektur

FE 01.0195/2015/AGB:

Linked Data

Analysis of application possibilities of linked information (Linked Data) and ontologies and related technologies (Semantic Web) in the road sector

Summary of research results

Jakob Beetz
Julian Amann
André Borrmann

Commissioned and funded by the
Bundesanstalt für Straßenwesen (BASt)

Munich / Eindhoven, 06/11/2018

Contact details

Technische Universität München
Ingenieurfakultät Bau Geo Umwelt
Lehrstuhl für Computergestützte Modellierung und Simulation

Arcisstraße 21
80333 Munich
https://www.cms.bgu.tum.de

Prof. Dr.-Ing. André Borrmann
Tel: +49 89 289-23047
Fax: +49 89 289-25051
Email: andre.borrmann@tum.de

Julian Amann
Tel: +49 89 289-25062
Fax: +49 89 289-25051
Email: julian.amann@tum.de

RWTH Aachen
Fakultät für Architektur
CAAD – Design Computation

Schinkelstraße 1
52062 Aachen
https://caad.arch.rwth-aachen.de/

Prof. Dr.-Ing. Jakob Beetz
Telefon: +49 241 80 95235
Telefax: +49 241 80 92618
E-Mail: beetz@caad.arch.rwth-aachen.de

1 Scope of work

1.1 Project goals

In this project, methodological approaches and prototypes for linking the “Objektkatalog für das Straßen- und Verkehrswesen” (OKSTRA[1]) with other information models based on “Linked Data” and “Semantic Web” technologies were developed and tested for their applicability and practicability using exemplary data sets.

In this context, the integration and contextualisation of similar and complementary international and European information models as well as the critical exploration of the development potential of mutual connections in various applications were investigated.

Over the course of the project, it was examined to what extent OKSTRA could play a possible role as a future basic model for a pan-European model. Initially limited to bi-national (DE/NL) examples, a study of the technical feasibility of cross-border road and path planning was carried out as part of the project that demonstrates the possibilities and limitations of the use of Linked Data technologies.

In addition, findings and results were examined with regard to their possible significance for federal and transnational deployment, which allows the various national systems to be connected.

2 Methods

The central focus of the research approach was the transformation of the OKSTRA data model, previously defined in UML[2] into the Web Ontology Language (OWL). A number of different approaches to this transformation were available, and their respective advantages and disadvantages were documented in order to assist in determining which mapping to employ. Based on okstraOWL, Linked Data methods were used to connect the Dutch object library CB-NL or the Object Type Library of the Dutch Ministry of Infrastructure (RWS-OTL) to the OKSTRA catalogue. This mapping was tested on an instance level using real data sets and the overall approach was evaluated in terms of the ability to formulate and process cross-model queries.

2.1 Evaluation of work on Linked Data in the field of geoinformatics

A detailed overview of the state of the art of data processing in road construction and infrastructure was first compiled. This overview discusses the individual models used today in the fields of geoinformatics (GML, CityGML, InfraGML), of infrastructure and road planning (LandXML, IfcAlignment, IfcRoad) as well as existing models used in building information modeling (IFC, ifcOWL) that are suitable for, or are undergoing further development for networking using Linked Data principles.

Along with RDF and OWL, the SPARQL query language is one of the building blocks of the Semantic Web. SPARQL can be used to query RDF data records. The OGC standard GEOSPARQL[3] is used to map GML data as RDF/OWL data records and to query such information using SPARQL. It defines an ontology for geographic information based on the General Feature Model, Simple Features and Feature Geometry as well as a set of SPARQL filter functions. The latter include the use of topological relationships (Egenhofer model and RC8 model) as filters in a query. Of special note is also- the effective description of geometric information as Well Known Text (WKT). This very effectively reduces the overhead that would otherwise result if generic RDF were applied directly to very extensive geometric information. A prerequisite for the use of GEOSPARQL is the use of a so-called end-point that supports and can process such requests, especially with regard to geographical filter functions.

[1] We have elected not to use the registered trademark symbol in OKSTRA® for better legibility in the document.
[2] Or, up until a few years again in ISO 103030-11 STEP EXPRESS. The current UML-schema is made accessible as XML thanks to the open XMI standard and serves as the basis for the research described here.
[3] http://www.opengeospatial.org/standards/geosparql

2.2 Analysis of OKSTRA and CB-NL

For OKSTRA, the schemes *S_Entwurf* (S_Design) and *S_Bauwerk* (S_Building) were examined in detail and in particular the geometric description forms of line curves or terrain models. The analysis showed that OKSTRA provides an extremely comprehensive, detailed and fine-grained description of road objects. The geometric description forms, in particular, are well-suited to application in practice through their use of GML-compliant geometry elements. One limitation, however, is that only 2D geometries can be used. Routing elements are defined by specific OKSTRA objects and the OKSTRA catalogue is defined using UML and XMI as conventional data modeling techniques.

The "Conceptenbibiblioteek Nederland" (CB-NL) is defined as an ontology using the description language OWL and is designed as a central collection of concepts that can serve as a hub for existing and future vocabularies, classifications, ontologies and data models. The concepts it can model have a high degree of abstraction and are not intended as a data model for the comprehensive and detailed modelling of concrete building projects in civil engineering. The central area of application of CB-NL is in the linking of different vocabularies and data models in a common model. The core of the CB-NL model is designed as an anchor onto which relations to other vocabularies can be attached.

In contrast to the generic CB-NL, the Object Type Library (OTL) of Rijkswaterstaat (RWS-OTL) is designed to capture concrete properties with which, for example, functional tenders and requirements and basic system trees of a number of different types of infrastructure can be described. In the description of roadways, RWS OTL makes a strict distinction between functional and physical objects. The RWS OTL 7 distinguishes between 7 levels of detail. A description of the geometry is not included in RWS OTL nor in CB-NL but is recorded using suitable external data sets in the form of generic IFC or GML models and is indirectly semantically enriched using OTL concepts.

CB-NL serves as the basic ontology for linking different classification and labelling systems in the Netherlands. RWS OTL's main area of application is asset management, and road design does not feature significantly. The primary areas in which OKSTRA and RWS-OTL overlap are road networks, intersections, road structure (structural road properties) and signage. In all other areas there is little to no overlap.

In summary, one can conclude that OKSTRA and RWS-OTL differ considerably in their descriptive depth and detail.

2.3 Translation of OKSTRA into OWL

To represent and process the OKSTRA schema and instance data records based on it as linked data structures, the following steps were taken:

1. Transfer of the data schema (von EXPRESS, XMI, XSD in RDF/RDFS/OWL)
2. Transfer (exemplary) instance data currently available as XML, SPFF or WebServices into one of the RDF serialisation formats
3. Making available of both data structures as Linked Data Infrastructures (SPARQL end-points).

Each of these three main components of the representation of OKSTRA data as RDF/OWL can be achieved in different ways. Although there are numerous "best practices" from other disciplines ad R&D initiatives, there is to date no “gold standard" for the transformation. Every decision made during these main steps has consequences for the handling of the resulting information and infrastructures. The report discusses the basis of the various transformation design decisions that serve as a basis for the implementation of a Linked Data representation of OKSTRA.

As a result, the OKSTRA schema defined in XMI is mapped to OWL resulting in “okstraOWL”. For the description of geometry, the decision was made to adopt the GML/Geo-SPARQL approaches with WKT. With this variant, we adopt the approaches developed by the OGC for the economical and efficient representation of geometries as RDF graphs using the so-called "Well-Known Text" (WKT). Since ordered collections of e.g. coordinates, as used to describe lines, polygons, etc., can only be captured in RDF using corresponding rdf:list elements, the basic idea of WKT is to represent points, lines, polygons, etc. as literal values, i.e. as character strings in a compact microformat.

Based on the considerations and decisions regarding the mapping of OKSTRA to OWL, a corresponding conversion for the generation of the okstraOWL schema was made. To illustrate the principle, we looked at the partial schema *S_Entwurf* (S_Draft). To represent this as an OWL schema, the definitions of the basic schema *okstra-basis.xsd*, *okstra-types.xsd* as well as *S_Allgemeine_Geometrieobjekte* (S_General_Geometry_Objects) and *S_Allgemeine_Objekte* (S_General_Objects) also had to be converted. The resulting schema was documented in detail and made available as a dereferencable SPARQL endpoint on the BASt server, which makes it available for immediate use in SPARQL queries. In addition, a converter for instance data sets was developed, that makes it possible to convert OKSTRA instance data of the above schemas into corresponding RDF data sets.

2.4 Comparison and harmonisation of object definitions and their attributes

The different modeling possibilities of semantic links between several models were examined with regard to their applicability for interlinking okstraOWL↔CB-NL/OTL. Different ways to link the two models were investigated based on the developed OWL representation of the OKSTRA. Using linked data techniques (e.g. Silk Workbench[4]), a semi-automatic alignment of the models was undertaken by defining equivalence relationships (*owl:sameAs*, *rdfs:seeAlso*, *skos:closeMatch* etc.) between individual concept definitions and their attributes. In a second step, these relationships were manually checked and confirmed using suitable tools.

Given the existing data basis and in particular the focus of the translations on the sub-schema *S_Entwurf* with around 450 translated concepts and properties, the automated linking of individual concepts is helpful to start with but requires significant manual reworking and manual creation of specific links by experts to produce a complete and practical mapping. Exemplary links were made for individual scenarios of WP 6 in order to examine their possible suitability for application.

2.5 Enriching RWS data sets with OKSTRA content

To investigate the possible use of the models in their different reciprocal contexts, we examined the aforementioned mapping considerations using several data sets provided by BASt and RWS for selected scenarios. These included a network model of the main national and local trunk roads in North Rhine-Westphalia and the motorway network of the province of Limburg (NL), which link up at various points as shown in Figure 1.

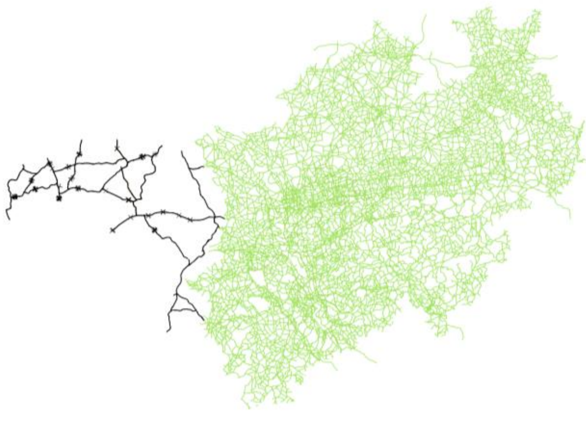

Figure 1 – Graphical overview of a common data space in the Netherlands (black) and Germany (green) using road data from different sources and in different data formats.

The graphical representation makes the different character and information content of both data sets immediately apparent: the NRW data comprises mainly topological network data and consists predominant-

[4] http://silkframework.org

ly of individual section elements represented as individual curve geometries in GML. The Dutch data set, on the other hand, is limited to a few individual routes, but has a higher level of detail and also contains, for example, extensive data on terrain, signage, lighting, cameras, etc. In contrast to the OKSTRA data, in which geometric representation and other, e.g. alphanumeric data are combined in a graph, the RWS-OTL data is composed of individual sections of COINS[5]-Containers. In order to us both different data sets from NRW and RWS-OTL together, each data set can additionally be enriched with the respective reciprocal vocabularies. Further details on the technical implementation are described in the report.

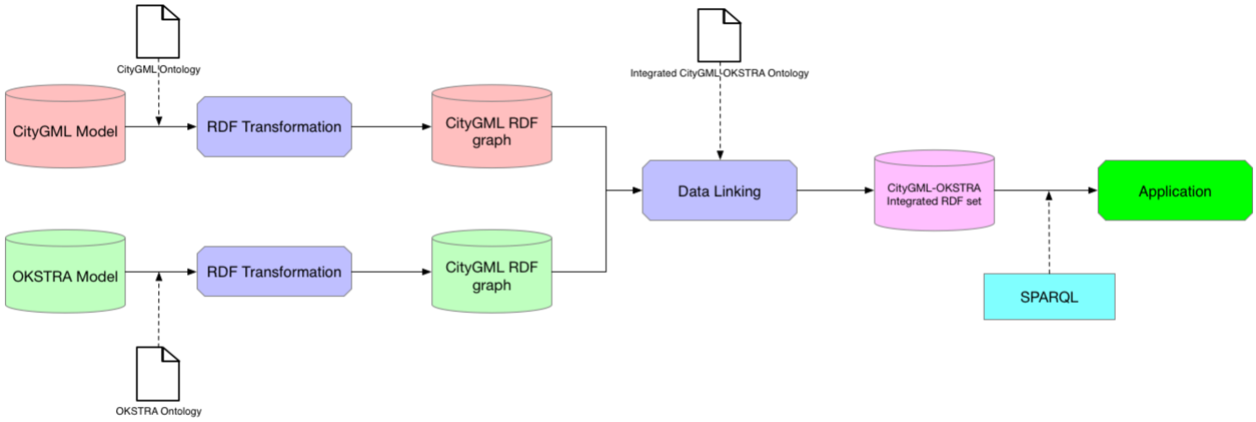

Figure 2 – Schematic representation of the link between okstraOWL data with CityGML data [Zheng 2017]

Another case study on linking heterogeneous data with okstraOWL data is the connection of 3D data in urban space and was carried out as part of a master's thesis[6]. Partial coverage models of the CityGML LOD 2 representation of NRW, with special focus on an inner-city section of Aachen[7], were combined with OKSTRA data of the state of NRW of B1, B1A and L136 and connected through spatial queries using GeoSPARQL. The basic technical structure of the case study is shown schematically in Figure 2.

The transformation of the CityGML data was done with an adapted version of an OWL representation of the CityGML metamodel based on the XSLT transformer by the University of Geneva[8]. Links between the different kinds of objects in both graphs were automatically generated, making use of the implicit spatial-geographical relationships between the two. Together with the multifaceted information contained in the CityGML model, such as the type of use or the height of the respective buildings, the resulting common OKSTRA/CityGML information space can be used to answer questions such as "Where are kindergartens and schools within a radius of 100 m close to a federal road?" Or: "Which buildings over 20 metres high are closer than 50 metres to a trunk road?"

The scenarios described here for linking and enriching information using okstraOWL illustrate the promising kinds of applications that can be undertaken flexibly and distributed with the help of widely-supported standard-compliant tools used in the processing of Linked Data. The clear advantage of Linked Data approaches over conventional database-based solutions lies in their simple management, ease of expansion, ability to be distributed and the flexibility creation of queries. The tests showed how structures in the semantically rich OKSTRA model can easily be used to semantically enrich other data and make it applicable across domain boundaries.

[5] COINS is a Dutch standard format for the transfer of planning data in infrastructure construction to the client and is conceptually and structurally similar to the Information Container Data Drop (ICDD), which is currently in the international standardisation process as ISO 21597

[6] Zheng, Yuan, 2017 "*Improving the Interoperability of Between City and Infrastructure Information. An Integration of CityGML and OKSTRA Data Based On Semantic Web and Linked Data Technology"* Technische Universiteit Eindhoven

[7] Section of ETRS89-UTM32 geo-coordinate system window (295000.0,5629000.0), (295000.0, 5630000.0), (296000.0, 562000.0) and (296000.0, 563000.0)

[8] http://cui.unige.ch/isi/onto//citygml2.0.owl

2.6 Other model applications

The availability of the OKSTRA model in okstraOWL format makes a whole range of applications possible that make use of the Semantic Web. In the first instance, a key aspect is the possibility of querying information using the SPARQL language which makes it possible to comprehensively analyse OKSTRA data using standardised tools. This includes all areas that OKSTRA covers, from the design phase to condition assessment and evaluation. The availability of okstraOWL makes it possible to simply to link objects from the *S_Entwurf* (design) schema to the *S_Strassenzustandsdaten* (roadway condition data) schema. Until now these objects have remained independently locked within their separate schemata. SPARQL also allows the inclusion of data sources from other domains (and other data models).

In addition to the ability to simultaneously query OKSTRA and CB-NL, there are many other possible applications in this context. These include, for example, the integrative formulation and answering of queries across data models, in which 3D city models in CityGML format can be evaluated together with OKSTRA data sets in order to answer questions such as identifying noise emissions. Further fields of interest for possible applications are the combination of accident, road condition and weather data to identify, for example, possible correlations between insufficient roughness of the road surface and the occurrence of accidents.

2.7 Prototypical implementation

The TUM Open Infra Platform is being developed at the Chair of Computer Aided Modeling and Simulation at the Technical University of Munich. The software is provided free of charge and can be downloaded from the department's website.[9] The software focuses on the promotion of open BIM approaches for infrastructure planning through the prototypical implementation of various open building data models. Standards such as OKSTRA, LandXML, InfraGML or IFC 4.1 are supported. The open availability of the software means that it can also be used as a reference implementation and help to accelerate the incorporation and support of certain standards in other software products.

Within the scope of this project, the TUM Open Infra Platform (OIP for short) was extended to provide export options for writing okstraOWL instance files. The OKSTRA class library (OKLABI) was used, which can be obtained free of charge from the OKSTRA website.[10] This is used to read in OKSTRA instance data. To write OWL-compliat files, the Raptor RDF Syntax Library[11] was used. By extending the OIP, okstraOWL instance files can be written as Turtle and RDF/XML files. All OKSTRA versions supported by OKLABI can be read, which covers all past OKSTRA versions. Exports of an OWL file, on the other hand, use OKSTRA schema 2.017 as its basis. For example, if an elevation plan is imported in OKSTRA format 1.014, it may contain the specific object *Tangentenfolge* which is no longer supported by the OKSTRA 2.017 standard. In this case, the old OKSTRA version is automatically upgraded to the new OKSTRA version and old domain objects are replaced by new ones accordingly.

3 Conclusions

Within the framework of this project, the data exchange standard OKSTRA for the description of road data, defined in XMI, was converted into a representation based on the Ontology Web Language (OWL). This means that Semantic Web methods and techniques are now also available for OKSTRA data sets. In particular, the Linked Data approach makes it possible to link data sets with other schemata or domains. This can be done by using the SPARQL query language for the integrative analysis of the data of various ontologies.

As shown in case studies, data from the Dutch road ontology CB-NL/OTL-RWS can be retrieved and analysed together with OKSTRA data. This makes a range of cross-border application scenarios possible, such as the planning of heavy load transports. Other applications of Linked Data in the road sector

[9] https://www.cms.bgu.tum.de/oip
[10] http://www.okstra.de/
[11] http://librdf.org/raptor/

include the integrated analysis of 3D city models in CityGML format with OKSTRA data or the linking of existing data with design data in OKSTRA format.

For the conversion of OKSTRA into okstraOWL, a multitude of different mapping options were available, the respective advantages and disadvantages of which are explained in detail in the report. Certain characteristics of the OKSTRA standard, such as the Fachbedeutungslisten (technical definition tables), make mapping more complex, but in principle a conversion that preserves the semantic structures is possible.

Although Linked Data functionalities using the okstraOWL are now available in principle, it became apparent in the course of the project (especially when working with real data sets) that the real challenge in linking different ontologies lies in the different semantic structure and granularity of the different data models. Methods of semi-automatic matching based on textual matches are also of limited help here. Instead, the user of the query mechanisms needs to have detailed knowledge of the semantics and structure of the ontologies involved and must create implicit or manual links based on this, the design of the query and the result they wish to achieve.

Despite the availability of Semantic Web and Linked Data technologies, the consistent, possibly pan-European use of street information databases therefore requires a fundamental harmonisation of data structures, especially with regard to their semantic structure and granularity.