Interpreting Heterogeneous Geospatial Data using Semantic Web Technologies

Timo Homburg[1], Claire Prudhomme[1] Falk Würriehausen[1], Ashish Karmacharya[1], Frank Boochs[1], Ana Roxin[2], Christophe Cruz[2]

{timo.homburg, claire.prudhomme,
falk.wuerriehausen, ashish.karmacharya,
frank.boochs}@hs-mainz.de
[1]Mainz University Of Applied Sciences
Lucy-Hillebrand-Straße 2
55128 Mainz, Germany

{ana-maria.roxin,
christophe.cruz}
@u-bourgogne.fr
[2]Université de Bourgogne
9 avenue Alain Savary
21000 Dijon, France

Abstract The paper presents work on implementation of semantic technologies within a geospatial environment to provide a common base for further semantic interpretation. The work adds on the current works in similar areas where priorities are more on spatial data integration. We assert that having a common unified semantic view on heterogeneous datasets provides a dimension that allows us to extend beyond conventional concepts of searchability, reusability, composability and interoperability of digital geospatial data. It provides contextual understanding on geodata that will enhance effective interpretations through possible reasoning capabilities. We highlight this through use cases in disaster management and planned land use that are significantly different. This paper illustrates the work that firstly follows existing Semantic Web standards when dealing with vector geodata and secondly extends current standards when dealing with raster geodata and more advanced geospatial operations.

Keywords: Heterogeneity, Interoperabiltity, SDI, CIP, GeoSPARQL, R2RML, Semantification

1 Introduction

When working in a geospatial context, one is confronted with several historically grown data sources which may be important for the current task to be solved. Integrating heterogenous datasets into a database ecosystem of some kind remains a necessary step before working on the actual task the data was determined for. To a certain degree, this process can be automized, yet can be seen as inflexible, as the data is often integrated and used for a limited set of usecases. Once the requirements of stakeholders change slightly, a redesign of at least some parts of the database system could be necessary. Simplifying the process of data integration, increasing its flexibility to deal with changing requirements and possibly extracting implicit information from integrated data could be a desirable goal.

2 State Of The Art

This section will discuss related work and the State Of The Art in integration of heterogenous geospatial data.

2.1 Related work on geospatial data integration

The work on geospatial integration focuses primarily on the syntactic standardization of geospatial data. Works like the OpenGIS Geography Markup Encoding Standard (GML) that expresses the geographic features serve as an open interchange format for transactions on the internet [8]. The issue has also been addressed at a policy level. Most countries have National Spatial Data Infrastructures (NSDIs) which are SDIs at national levels that provide standardized frameworks for sharing geospatial data. Now, regional efforts are also being made to harmonize these NSDIs under a common regional SDI umbrella. An example could be the Infrastructure for Spatial Information in European Community (INSPIRE) [21] at a European level. The root problem in all these efforts is that they there are not have enough constructs to express semantics [28]. However, the current upsurge of Semantic Web technologies has fueled the implication of semantics in geospatial data. The general tendency of this implication is to combine data from different sources semantically to provide a unified semantic view [18] through mostly a global schema that maps definitions to the schema definitions at the local data sources. Projects like GeoKnow[1] apply similar strategies to combine, structure and expose geospatial data onto the web [17], [15] . We extend this tendency by integrating distributed data sources and laying a foundation for geospatial inference.

2.2 The Semantic Web Technologies

The term “Semantic Web“ is coined by Tim Berners-Lee in his work to propose the inclusion of semantics for better enabling machine-people cooperation for handling huge information that exists in the web [6]. It is basically an extension to the current web supported through the standards and technologies standardized by World Wide Web Consortium (W3C)[2]. We present a few important technologies that define the Semantic Web in this paper.

2.2.1 Web Ontology Language The term ontology is being used for centuries to define an object philosophically. The core theme of the term remains the same in the domain of computer science; however the approach in defining it has been modified to adjust the domain. Within the computer science domain, ontology is a formal representation of knowledge through the hierarchy of concepts and the relationships between those concepts. In theory an ontology is a formal, explicit specification of shared conceptualization [13]. OWL or the Web

---

[1] http://geoknow.eu/Welcome.html

[2] https://www.w3.org

Ontology Language is a family of knowledge representation languages to create and manage ontologies. The World Wide Web Consortium (W3C) has standardized OWL to model ontologies. The standardization of OWL has sparked off the development and/or adaption of a number of reasoners like Pellet[3].

In recent years a number of efforts on developing geospatial ontologies for adding semantics in spatial data have been witnessed. Geospatial ontologies take their domain and range of concepts as geospatial objects, relations and features[3]. Recent studies like [4], [11], [12] implement through geospatial semantic expressions in the Web Ontology Language[4] (OWL) and/or Resource Description Framework[5] (RDF). The Geonames ontology[6] adds possibilities to add geospatial semantics in the world wide web and also describes the relation between toponyms. Likewise, the LinkedGeoData project developed the LinkedGeoData ontology knowledge base[7] that lifts Open Street Map (OSM)[8] data to be presented in a Semantic Web infrastructure [24]. It is a lightweight ontology build through first conversion of OpenSteetMap data into RDF and then interlinking to DBPedia[9], GeoNames, and other datasets. Moreover, it supports multi-lingual class labels from various sources. The GeoKnow project applied LinkedGeoData knowledge base to set up benchmarks within its use cases. It also contributed in updating and modifying the the knowledge base.

2.2.2 Query and Reasoning

SPARQL is an RDF query language for querying Resource Description Framework RDF triples of which OWL is syntactically aligned. Therefore, SPARQL is able to query OWL ontologies. As a query language, SPARQL is “data-oriented” in the sense that it only queries the information held in the models; there is no inference in the query language itself. SPARQL uses FILTERS to limit the solutions to only those which are returned true with the expression.

The OGC GeoSPARQL standard is the spatial extension to SPARQL, proposed by OGC. It is the initiative taken by OGC in collaboration with the SPARQL W3C working group to define vocabularies for representing geospatial data in RDF. These standards implement the SPARQL query language to process geospatial data. Moreover, it accomodates qualitative spatial reasoning and systems based quantitative spatial computations[1]. Qualitative spatial reasoning tests the binary spatial relations between features and do not usually model explicit geometries. In the meantime, quantitative systems transform the query into a geometry based query that evaluates computational geometries between features[9].

---

[3] https://www.w3.org/2001/sw/wiki/Pellet

[4] https://www.w3.org/TR/owl-guide/

[5] https://www.w3.org/RDF/

[6] http://www.geonames.org

[7] http://linkedgeodata.org/About

[8] http://www.openstreetmap.org

[9] http://dbpedia.org

The Semantic Web Rule Language (SWRL) is a rule language to infer ontology knowledge bases. SWRL has the form, antecedent → consequent, where both antecedent and consequent are conjunctions of atoms written a1 ∧.... ∧ an. Atoms in rules can be of the form C(x), P(x,y), Q(x,z), sameAs(x,y), differentFrom(x,y), or builtIn(pred, z1, ..., zn). For instance, the following rule asserts that one's parents' brothers are one's uncles where parent, brother and uncle are all individual-valued properties.

$$parent(?x, ?p) \wedge brother(?p, ?u) \rightarrow uncle(?x, ?u) \tag{1}$$

Limited work has been conducted to include semantic on spatial geofeatures and operations into SWRL. Nevertheless, spatial extensions through spatial built-ins were proposed in the research work [16]. This work defines how spatial semantics could be used within SWRL to trigger spatial inferencing.

3 Semantification

We define semantification as the process of interpreting given geospatial data using semantic technologies. A general description of a semantification process can be stated as follows:

1. Determining a describing set of owl classes for a corresponding data set
2. Determining a unique identifier for instances of the describing class
3. If possible inferring types for to be extracted features
4. If possible inferring restrictions for extracted features and classes
5. Interlinking aforementioned elements to other domains

3.1 Schema-assisted semantification

A schema-assisted semantification can take place using either a database or a file that contains a certain schematic description of its content. In a GIS context this involves databases like POSTGIS and a substantial amount of geospatial data given in an XML dialect, mostly as KML or GML files. However, from a semantic web perspective it can make sense to not only integrate data directly related to a geospatial object, but rather to associate possible non-geospatial entities linked to a geospatial object. We will illustrate this thought in our usecases section.

3.1.1 Semantification of relational databases Interpreting relational databases using a semantic web layer, has been extensively studied and W3C[10] recommendations such as R2RML[11] and Direct Mapping[12] have emerged as well as in the case of SPARQLify[13] been further developed. While R2RML proposes using a

---

[10] World Wide Web Consortium

[11] https://www.w3.org/TR/r2rml/

[12] https://www.w3.org/TR/rdb-direct-mapping/

[13] http://aksw.org/Projects/Sparqlify.html

manually defined mapping profile to map database tables to preexisting classes in ontologies, the Direct Mapping approach delays the interlinking part to be manually specified in the ontology after having interpreted database tables using their given designations. In our approach we are allowing currently employing R2RML and hope to add automated support for direct mapping including interlinking in the future.

3.1.2 Semantification of XML files with a given schema (XSD) A few projects like Redefer [14] , Ontmalizer [15] and XS2OWL [26] have been attempting to convert given XSD schemas to RDF, in order to simplify the interpretation of given file formats. Many of those projects perform well on some XSD files, but have shortcomings in terms of completeness. This comes to little surprise as [16] points out, a general transformation process is not trivial and because of ambiguities sometimes errornous. Alternatively, rather than transforming XSD schemas to RDF, projects like GeoKnow[17] have been providing one XSL transformation file per file format on the data files themselves. This approach compared to a schema extraction is less complicated, usually much more feasible but cannot encompass property cardinalities, the class hierarchy and available codelist references. In our approach we rely on a preextraction of XSD schemas using a modified XSL script inspired by the previously mentioned approaches to create a local ontology [7] for every file format we needed to integrate. After the creation of local ontologies, data files matching the preextracted schema can be imported effortlessly into the ontologies class structure. In addition to XSD transformations, codelists of XSD schemas[18] need to be taken into account. Those are often not included in the schema descriptions and therefore need to be merged independently, as they often contain the most valuable describing information (e.g. classification of an object). With our extraction mechanism we are able to gain a class hierarchy, available properties including their ranges and domains, defined restrictions on a per class basis, as well as codelist representations as enumerations along with additional information contained in comments.

3.2 Semiautomated Schemaless semantification

File formats without a distinctive schema such as shapefiles or schemaless GML files lack relations to other classes in the to-be-built ontology. On the example of a shapefile containing data about tree locations in Germany we would like to illustrate our semantification approach for schemaless data:
In table 1 the contents of the shapefile ”Tree.shp” are given. It includes the geometry column, a tree type description column, an integer column and a column including city names. This information enables us to automatically infer the

---

[14] http://rhizomik.net/html/redefer/
[15] http://www.srdc.com.tr/projects/salus/blog/?p=189
[16] http://www.ieee-icsc.org/ICSC2011/slides/XSD2OWL_Patterns_ICSC2011.pdf
[17] http://geoknow.eu/Welcome.html
[18] Example Codelists of INSPIRE: http://inspire.ec.europa.eu/codelist/

ID	Geometry	Feature1	Feature2	FeatureN
Tree1	POINT(..)	Oak	4	Mainz
Tree2	POINT(..)	Box	3	Koblenz

Table 1: File Tree.shp

representation shown in figure 1 of the file in RDF with the following optional pieces of information provided by the user:

- A unique identifier for individuals from one of the columns. If none is provided a unique identifier will be generated.
- An optional column defining subclasses of the class described by the files name

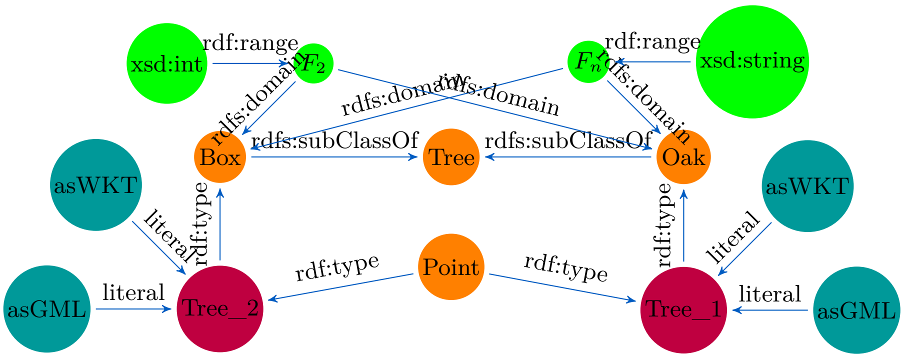


Figure 1: Semantification Example

This representation is consistent, yet far from satisfying, as we have no semantic interpretation of the integer column and the column representing the cities. Semantic interpretations of a column containing text such as the city column could be matched using two ways:

- The column description can possibly be matched to a class in a preexisting ontology using a tool like BabelNet[20]
- Attribute values may as well be matchable using BabelNet

Each of those approaches will produce a set of possible preexisting and known concept classes a user can choose from, therefore enabling a user to semiautomatically interlink schemaless data to the ontology infrastructure. In the case of columns representing numbers only, the user should not only choose a concept but, if possible also a corresponding unit provided by the QUDT[19] ontology. On import, values of units are converted by our system to a Base SI Unit in order to guarantee a degree of comparison between values in GeoSPARQL queries.

---

[19] Quantities, Units, Dimensions and Data Types Ontology: http://www.qudt.org/

3.2.1 Semantification and querying of raster data Raster data is represented by a (thematic) raster of pixels, each of which correspond to an associated value or the NODATA value. Raster data can therefore be interpreted as an owl:Class named after the files' designation/a user assigned concept having exactly one DataProperty describing the value or if metainformation is given an ObjectProperty respectively. Rasterdata is not natively supported by GeoSPARQL but needed in many geospatial applications.
One of our project partners' typical flood simulation cases aims to find out for concerned elements at risk of being flooded and identifying still usable rescue paths from certain areas at certain flood altitudes. In this usecase we are following the latter approach and want to find out which roads are still usable after a flood of x meters has struck the city of Cologne. We are provided with the following data:

- A shape file of the roads (LineStrings) in Cologne
- A GEOTIFF file including the flood altitudes above ground level in our simulation

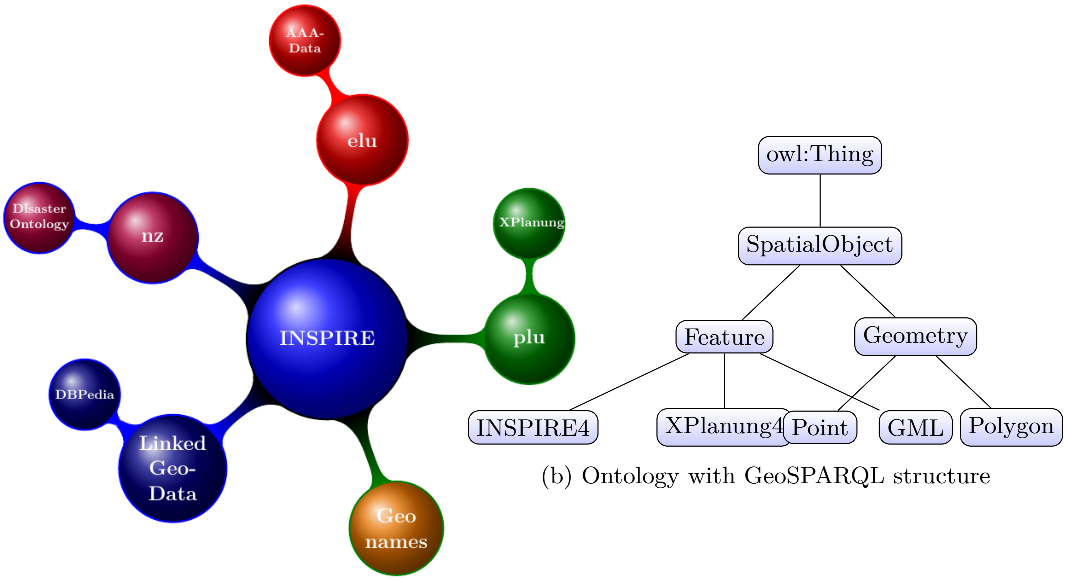
To solve this usecase, the roads are first splitted into segments of a certain length. Following this for every segment bounding boxes of width 1m are clipped on the raster data set to find out if the altitude of the flood increases 1m. In this case the road segment and therefore its corresponding road is considered as unaccessible. Clearly, on a semantic level there is currently no standard that provides access to both raster and vector data to accomodate a usecase as highlighted above.

3.3 Extending GeoSPARQL

The GeoSPARQL definition [5] defines a region connection calculus for the geospatial web. However, it does not include geometry constructors and geometry manipulation functions common in other spatial infrastructures. Support for raster data is also not present in its current specification. Clearly, our usecase demands such functionality to solve aforementioned problems. To our knowledge there exists no such system in the semantic web. We therefore are in the process of extending the ARQ query processor implementation using the Java Topology Suite library to support most of common functions available in comparable implementations such as POSTGIS. To support raster data we polygonize the desired region of the raster using the GDAL polygonizing library, as comparable other implementations do. On completion of this approach we want to provide a semantic platform that is capable of interpreting and dealing with both raster and vector data and combinations of them.

3.4 Interlinking

Each of the mentioned semantification methods lead to a local ontology which class tree structure is to be integrated in a global broker ontology to be accessible. For the creation of the geospatial part of the global broker ontology we follow the specifications of GeoSPARQL[20], thereby subclassing each local ontology under the class SpatialObject, as illustrated in 2b. Interlinking of generated local ontologies is a major challenge to be solved in this regard. Our goal is to create or reuse a unified vocabulary covering all concepts that have been provided us by the data sources. This will allow not only expert users familiar with the local ontologies to access geospatial data, but also expose the nowadays mainly inaccessible world of spatial data definitions to the semantic web community. In the absence of experts from our side we currently apply a semi-automatic interlinking approach to match data sources with concepts by the importing users (which we expect to have expertise on the data they import). However we intend to further investigate the quality of matching concepts in an automized fashion as described in section 5.1. In contrast to [7] we would like not only to use concept matching and attribute matching but furthermore focusing on matching concepts of geometries in related data sources.



(a) Current Broker Ontology Structure

(b) Ontology with GeoSPARQL structure

However, even with manual inteerlinking by experts alone we were able to create an ontology infrastructure that can be seen in figure 2a . This ontology infrastructure currently contains ontologies of data structures near to INSPIRE and will possibly be extended in the future. Thereby INSPIRE will not be the center of our infrastructure but more likely one of many ontologies to be considered.

---

[20] http://www.opengeospatial.org/standards/geosparql

3.5 Management of Data Sources

Semantification is useful to interpret given data sets and to infer a class structure in OWL. However, depending on the scale of the system it might not be possible to store all datasets in one database (triple store) all of the time. We are therefore investigating caching mechanisms to be used to only store data that is needed for the current requests received by the system. To utilize caching we need up-to-date references as annotations to classes and properties in the ontology linking to the actual data sources. A source annotation contains the path to the data source, a timestamp of last access and metainformation on how to interpret the data given. The given annotations are created on import or when an update request is triggered.

3.5.1 Query Rewriting for Performance Improvements In a previous publication, Tschirner et. al [25] proposed to use a ETL[21] process to access WFS services via a semantic layer and a annotated ontology as previously described. Depending on the SPARQL query given by the user we are taking this approach further to not only support WFS services, but rather supporting spatial databases through R2RML, as well as files queryable by OGC CQL or another query language. Depending on a to be refined caching algorithm to improve performance, we created our system to selectively access data by rewriting queries defined by the users to the underlying query languages and fetching data depending on demand and caching data as deemed useful by our system.

4 Case Study of Semantification

Based on the goals in the context of interpreting distributed geospatial data with semantic web, individual cases of semantification are presented in the following section.

4.1 National Strategy for Critical Infrastructure Protection

Based on the National Strategy for Critical Infrastructure Protection (CIP) in Germany [2], this case study highlights the need of integrated heterogenous geospatial data for the purposes of environment and energy. In this context we investigate flood and disaster management cases and the protection of critical infrastructures. As described in [10], disaster management is a cycle of four steps: Mitigation, Preparedness, Response, and Recovery. In addition to the real world application concept presented in the part 3.2.1 with raster data, we will present a use case of heterogeneous data interpretation in the field of disaster management. It is based on a flood situation and aims to show how different data can be combined in order to provide relevant information for preparing the response to a disaster (corresponding to step 2 in disaster management).

---

[21] Extract Transfer Load

Use Case: Our use case is based on a flood risk estimation of Saarland, a region in Germany. When a flood happens, people who are exposed, have to be evacuated to a safe area and transport casualties to a hospital. One of prioritary facilities are schools. In order to organize a rescue plan of schools in a flood area, we search to determine for each school in a flood area, the hospitals which are not flooded and in a radius of 15 Km. Thanks to the INSPIRE geoportal, three GML files have been retrieved from WFS services:

- A file containing hospitals[22] with information about their location, provided services, and number of beds.
- A file containing the positon of schools[23] .
- A file containing the position of flood risk areas[24] .

Using our approach exposed previously, these files are interpreted and gathered in one semantic model which can afterwards be queried. The interpretation of these data sets allows to obtain a quick answer due to GeoSPARQL queries. The GeoSPARQL query corresponding to the research of unexposed hospitals which have a maximum distance of 15Km from each school present in a flood area. This flood area is defined by a radius around a point of flood risk. This GeoSPARQL query is presented in Listing 1.1:

Listing 1.1: Exposed schools query

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX school: <http://geoportal.saarland.de/arcgis/services/Internet/
    Staatliche_Dienste/MapServer/WFSServer#>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX dangerzone:<http://geoportal.saarland.de/arcgis/services/Internet/
    Hochwasser_WFS/MapServer/WFSServer#>
PREFIX hospital:<http://geoportal.saarland.de/arcgis/services/Internet/
    Gesundheit/MapServer/WFSServer#>

SELECT DISTINCT
?dangerzone ?school_name ?hospital_name
WHERE {
?school a school:Schulen_SL .
?school geo:hasGeometry ?school_ind .
```

---

[22] http://geoportal.saarland.de/arcgis/services/Internet/Gesundheit/MapServer/WFSServer?SERVICE=WFS&REQUEST=GetFeature&typeName=Gesundheit:Krankenhaeuser

[23] http://geoportal.saarland.de/arcgis/services/Internet/Staatliche_Dienste/MapServer/WFSServer?SERVICE=WFS&REQUEST=GetFeature&typeName=Staatliche_Dienste:Schulen_SL

[24] http://geoportal.saarland.de/arcgis/services/Internet/Hochwasser_WFS/MapServer/WFSServer?SERVICE=WFS&REQUEST=GetFeature&typeName=Hochwasser_WFS:Betr_EW

```
?school_ind geo:asWKT ?school_geo .
?school school:SCHULNAME ?school_name .
?dangerzone a dangerzone:Betr_EW .
?dangerzone dangerzone:OBJECTID ”10” ^^<http://www.w3.org/2001/
    XMLSchema#int> .
?dangerzone geo:hasGeometry ?dangerzone_geoind .
?dangerzone_geoind geo:asWKT ?dangerzone_geo .
?hospital a hospital:Krankenhaeuser .
?hospital hospital:KRANKENHAU ?hospital_name .
?hospital geo:hasGeometry ?hospital_geoind .
?hospital_geoind geo:asWKT ?hospital_geo .
FILTER(geof:sfWithin(?school_geo, geof:buffer(?dangerzone_geo, 3500, uom:
    meter)))
FILTER(!geof:sfWithin(?hospital_geo,geof:buffer(?dangerzone_geo, 3500, uom:
    meter)))
FILTER(geof:sfWithin(?hospital_geo,geof:buffer(?school_geo, 15000, uom:meter)
    ))
}
```

The result of this query is presented in table 2. It depicts the following situation on a map of the area in figure 3.

Riskzone	**Exposed School**	**Next unexposed Hospital**
Saubach	Erich-Kästner -Schule	Marienhausklinik Wadern
Saubach	Grundschule St. Michael	Marienhausklinik Wadern
Saubach	Berufsbildungs -zentrum Lebach	Marienhausklinik Wadern
Saubach	Gemeinschafts -schule ERS Lebach, Theetal Schule	Marienhausklinik Wadern

Table 2: Queryresult

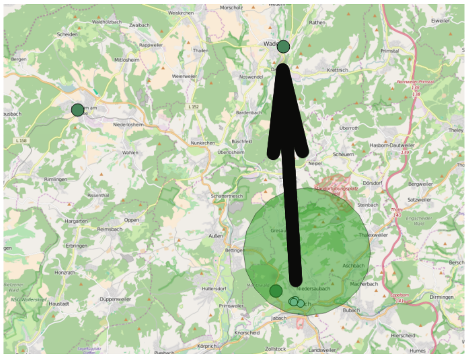


Figure 3: Result visualized (Hospitals in green, schools in light blue, riskarea in transparent green)

The interpretation of heterogeneous data is the base of a system for disaster management. This use case shows the relevance to retrieve and interprate heterogeneous data. In our future work, we will use this base to create a more complex and sophisticated system allowing to generate automatically new information thanks to rules.

4.2 Distributed Data in Spatial Data Infrastructures

4.2.1 Comprehensive Management The need to use heterogeneous and distributed data becomes evident particularly in the current developments of SDI's and already existing geoportals on the Internet. Users should be able to create a simple search for spatial data while the usability and usage conditions should be quickly recognizable [23]. Established procedures for functional spatial data exchange among and across different levels of government are operational, such as in land use planning data management. A considerable share of planned land-use information in Germany specified in XPlanGML[25] addressed by the INSPIRE Annex III theme land-use [23]. In addition to the work in [19,27], the use of ontologies in the context of German land-use planning and INSPIRE, should be presented.

4.2.2 Rule Management Particularly needed are accepted rules for the provision of local government datasets, as well as for services which are able to process comprehensive spatial data based on different, even municipal, spatial data themes. Integrated data concepts and data models can help ensure cooperation even in a heterogeneous environment of organization units. In-depth analysis of data structures, therefore, is imperative. The "Technical Guidance" for INSPIRE Transformation Networking Services (TNS) [14] not only explains the functional requirements for the INSPIRE regulations but also explains the implementation rules. The needed mapping and transformation rules in this case are those between any local SDI source data model and the common INSPIRE target data model, here defined in the Annex III specification Land use. By using semantics, our mapping rules can be described with (s)ubject, (p)redicate, (o)bject in a RDF-triple, see table 3.

(s)ubject	(p)redicate	(o)bject
<http://www.xplanung.de/xplangml/4/0#XP_Plan>	<http://www.w3.org/2002/07/owl#equivalentClass>	<http://inspire.ec.europa.eu/schemas/plu/4.0#SpatialPlan>
<http://www.xplanung.de/xplangml/4/0#name>	<http://www.w3.org/2002/07/owl#equivalentObjectProperty>	<http://inspire.ec.europa.eu/schemas/plu/4.0#officialTitle>
<http://www.xplanung.de/xplangml/4/0#BPlan_1000>	<http://www.w3.org/2002/07/owl#sameAs>	<http://inspire.ec.europa.eu/schemas/plu/4.0#infraLocal>
<http://www.xplanung.de/xplangml/4/0#BP_Objekt>	<http://www.w3.org/2002/07/owl#equivalentClass>	<http://inspire.ec.europa.eu/schemas/plu/4.0#ZoningElement>

Table 3: XPlanung4 to INSPIRE relations

Table 3 shows an excerpt of transformation rules for eqivalentClass, equivalentObjectProperty and sameAs relations with attributes of the land-use in national XPlanGML and European INSPIRE systems as URI. We can represent the corresponding mapping rules in an ontology graph, created with our prototype of the ontology mapping tool. It can be described in accordance with Tschirner

---

[25] http://www.iai.fzk.de/www-extern/index.php?id=680&L=1

et al. [25] as a step to integrate GML-/XSD- Source into the RDF/OWL- language. Rules are presented in two distinct forms, by using the ontology through restriction axioms and rules for the equivalenc.

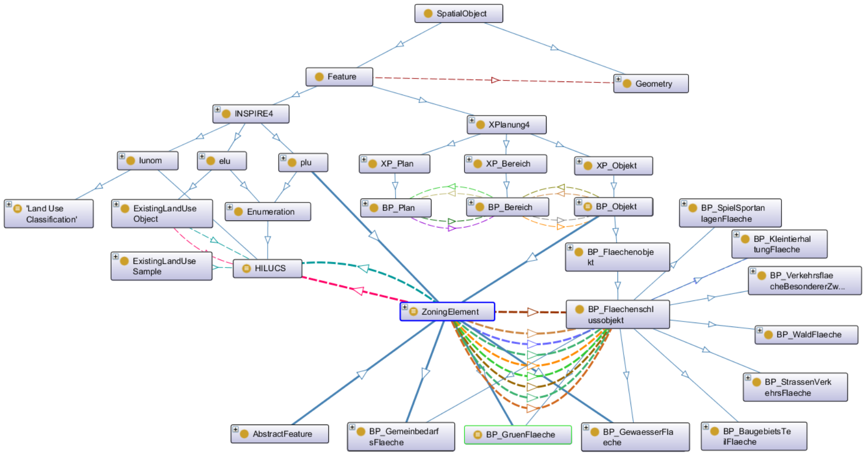


Figure 4: Owl Graph of interlinked INSPIRE and XPlanung4

Fig.4 demonstrates the graphical overview of some local and INSPIRE zoning element vocabularies and how they are interlinked in the ontology. The feature types of the local-level source data model had to be mapped to the INSPIRE target data model by using semantic rule definition.

5 Conclusion

The implementation of knowledge management in this application field can help to support interoperability for the benefits of the local level. The benefits of the approach can be found in the unique flexibility of an ontology instead of some feature manipulation, in the standard-based application and the versatility of use. In that way ontologies can be used to model extensive knowledge about the data, for rule definition, and other objectives in order to reduce the increased complexity for the local government. The automated service function focuses on those data quality elements and processes that can be fully automated (e.g. logical consistency). The semi-automatic service focuses on those data quality elements and measurements that require some human intervention (e.g. data generalisation). The implementation of knowledge management in schema interlinking application fields can help to support interoperability for the benefits at the local level. The feature types of the local level source data model are interlinked to the INSPIRE target data model. Based on the model of Waters et al. [22], the steps from "Feature Transformation" to "Data Generalisation" will be described. The supported Quality steps and Rules are those between any local SDI source data model and the common INSPIRE target data model, defined in the Annex specifications.

1. Transform local schemas and features to common INSPIRE data specification framework
2. Integrate INSPIRE data into ”virtual” collaborative datasets
3. Ensure cross-border consistency across neighbouring data
4. Reduce data complexity, for effective use at alternate scales and real worls scenarios

For all cases, the process results of transformation, data integration, consistency, generalisation of local sources. It can be expected that the requirement of easier usage of different data models can be met if the study of the rules can significantly reduce the cost of a user interpretation of heterogeneity. This can be achieved by semi-automated mapping of the class descriptions in the data collection process.

5.1 Future Work

Our future work will focus on extending not only the data we are about to access but also on insuring the data's quality and enriching it with inferrable knowledge.

5.1.1 Automated Concept Matching In the future we want to fully automize the process of matching concepts to relational geospatial datasets. This process has been described in section 3.1.2 We therefore want to investigate natural language processing methods and methods to match the spatial objects we are integrating using a reference source. We hope to achieve a very accurate ranking of plausible concepts for each dataset and will test the reulting ranking using a reference data set and appropriate metrics.

5.1.2 Application of reasoning Through the application of reasoning techniques we will be able to infer class structures and individuals from existing data, that can be useful in several use cases. For this purpose we may need to extend an appropriate reasoning language similarly to extending GeoSPARQL to support more spatial functions. In section 4.1 of disaster management we could profit from reasoning by creating a reasoning rule defining us the relation between related hospitals with exposed schools through the SWRL statement presented below.

$$\begin{aligned} Hospitals(?h) \wedge Schools(?s) \wedge RiskArea(?r) \\ \wedge semGIS : buffer(?r, ?br, 2) \wedge semGIS : within(?br, ?s) \\ \wedge semGIS : buffer(?s, ?bs, 15) \wedge semGIS : disjoint(?br, ?h) \\ \wedge semGIS : within(?bs, ?h) \rightarrow isSuitableHospital(?h, ?s) \end{aligned} \tag{2}$$

The resulting class suitableHospital could then be further reused to solve related issues.

5.1.3 Quality Assessment and Data Aggregation Lastly, we want to focus on integrating further data sources into our system and assess the quality of data provided using to-be-defined metrics. Having evaluated our data basis in this way, our system should be able to select ”good” and available data sources preferably on a request basis to give the best possible result under the current circumstances (missing, currently unaccessible data sources, etc.).

6 Acknowledgements

This project was funded by the German Federal Ministry of Education and Research[26]

References

1. Geosparql - a geography query language for rdf data. `http://www.opengeospatial.org/standards/geosparql`, accessed: 17.03.2016 2.2.2
2. National Strategy for Critical Infrastructure Protection (CIP Strategy) 4.1
3. Socop/geospatial ontologies. `http://ontolog.cim3.net/cgi-bin/wiki.pl?SOCoP/GeospatialOntologies`, accessed: 16.03.2016 2.2.1
4. Arpinar, I.B., Sheth, A., Ramakrishnan, C., Usery, L., Azami, M., po Kwan, M.: Geospatial ontology development and semantic analytics. In: Transactions in GIS. pp. 1–15. Blackwell Publishing (2005) 2.2.1
5. Battle, R., Kolas, D.: Geosparql: enabling a geospatial semantic web. Semantic Web Journal 3(4), 355–370 (2011) 3.3
6. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American 284(5), 34–43 (2001) 2.2
7. Bizid, I., Faiz, S., Boursier, Patriceand Yusuf, J.C.M.: Advances in Conceptual Modeling: ER 2013 Workshops, LSAWM, MoBiD, RIGiM, SeCoGIS, WISM, DaSeM, SCME, and PhD Symposium, Hong Kong, China, November 11-13, 2013, Revised Selected Papers, chap. Integration of Heterogeneous Spatial Databases for Disaster Management, pp. 77–86. Springer International Publishing, Cham (2014), `http://dx.doi.org/10.1007/978-3-319-14139-8_10` 3.1.2, 3.4
8. Bossomaier, T., Hope, B.A.: Online GIS and spatial metadata. CRC Press (2015) 2.1
9. Consortium, O.G.: Ogc geosparql - a geographic query language for rdf data. Tech. rep. (2011) 2.2.2
10. Coppola, D.P.: Introduction to International Disaster Management. Elsevier (2011) 4.1
11. Cruz, I.F.: Geospatial data integration. Chicago: ADVIS Lab, Dept. of Computer Science, University of Illinois (2004) 2.2.1
12. Fonseca, F.T., Egenhofer, M.J., Agouris, P., Câmara, G.: Using ontologies for integrated geographic information systems. Transactions in GIS 6(3), 231–257 (2002), `http://dx.doi.org/10.1111/1467-9671.00109` 2.2.1
13. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. 5(2), 199–220 (Jun 1993), `http://dx.doi.org/10.1006/knac.1993.1008` 2.2.1

---

[26] https://www.bmbf.de/en/index.html

14. Howard, M., P.S.S.R.: Technical guidance for the inspire schema transformation network service. version: 3.0, ec jrc contract notice 2009/s 107-153973 (2010) 4.2.2
15. Jens LEHMANN, Spiros ATHANASIOU, A.B.A.G.R.G.G.D.H.J.J.L.G.A.C.N.N.M.A.S.C.S.M.W.P.W.V.Z.: Managing geospatial linked data in the geoknow project. The Semantic Web in Earth and Space Science. Current Status and Future Directions 20, 51 (2015) 2.1
16. Karmacharya, A.: Introduction of a spatial layer in the Semantic Web framework: a proposition through the Web platform ArchaeoKM. Ph.D. thesis, Le2i – Laboratoire Electronique, Informatique et Image, University of Bourgogne (2011) 2.2.2
17. LEHMANN, J., ATHANASIOU, S., BOTH, A., ROJAS, A.G., GIANNOPOULOS, G., HLADKY, D., LE GRANGE, J.J., NGOMO, A.C.N., SHERIF, M.A., STADLER, C., et al.: The geoknow handbook (2015) 2.1
18. Lenzerini, M.: Data integration: A theoretical perspective. In: Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. pp. 233–246. PODS '02, ACM, New York, NY, USA (2002), http://doi.acm.org/10.1145/543613.543644 2.1
19. Müller, H., Würriehausen, F.: Computational Science and Its Applications – ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24-27, 2013, Proceedings, Part III, chap. Semantic Interoperability of German and European Land-Use Information, pp. 309–323. Springer Berlin Heidelberg, Berlin, Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-39646-5_23 4.2.1
20. Navigli, R., Ponzetto, S.P.: Babelnet: Building a very large multilingual semantic network. In: Proceedings of the 48th annual meeting of the association for computational linguistics. pp. 216–225. Association for Computational Linguistics (2010) 3.2
21. Rase, D., Björnsson, A., Probert, M., Haupt, M.: Reference data and metadata position paper. inspire rdm pp v4-3 en. european commission, joint research centre (2002) 2.1
22. Robin Waters, Matt Beare, R.W..M.M.: Schema transformation for inspire. International Journal of Spatial Data Infrastructures Research 6, 1–22 (2011) 5
23. Specifications", D.T.D.: D2.3: Definition of annex themes and scope v3.0., european union, (2008) 4.2.1
24. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: Linkedgeodata: A core for a web of spatial open data. Semantic Web Journal 3(4), 333–354 (2012), http://jens-lehmann.org/files/2012/linkedgeodata2.pdf 2.2.1
25. Tschirner, S., Scherp, A., Staab, S.: Semantic access to inspire. In: Terra Cognita 2011 Workshop Foundations, Technologies and Applications of the Geospatial Web. p. 75. Citeseer (2011) 3.5.1, 4.2.2
26. Tsinaraki, C., Christodoulakis, S.: Xs2owl: a formal model and a system for enabling xml schema applications to interoperate with owl-dl domain knowledge and semantic web tools. In: Digital Libraries: Research and Development, pp. 124–136. Springer (2007) 3.1.2
27. Würriehausen, F., Karmacharya, A., Müller, H.: Computational Science and Its Applications – ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30 – July 3, 2014, Proceedings, Part II, chap. Using Ontologies to Support Land-Use Spatial Data Interoperability, pp. 453–468. Springer International Publishing, Cham (2014), http://dx.doi.org/10.1007/978-3-319-09129-7_34 4.2.1
28. Zhao, T., Zhang, C., Wei, M., Peng, Z.R.: Geographic Information Science: 5th International Conference, GIScience 2008, Park City, UT, USA, September 23-26, 2008. Proceedings, chap. Ontology-Based Geospatial Data Query and Integration, pp. 370–392. Springer Berlin Heidelberg, Berlin, Heidelberg (2008), http://dx.doi.org/10.1007/978-3-540-87473-7_24 2.1