Linked Open Data in Spatial Data Infrastructures

Editors: Marcell Roth & Arne Bröring (52°North)

Date: 30.01.2013

Contents

1. Introduction

The term „Linked Open Data" (LOD) is a set of concepts and best practices for providing and connecting structured data on the Web. These concepts are increasingly used for the provision of environmental data. In the INSPIRE context, environmental data should be provided via Web services with open interfaces. In addition to such Web services, the LOD techniques offer the opportunity to not only use technical services for data access, but also to organize the data itself on the Web and to use the Web as a global database. From the power of this concept, and the speed this new technology is entering practical usage, there is a high relevance to the objectives of the GLUES project. The combination of LOD and Spatial Data Infrastructures (SDIs) can improve issues relevant to SDIs.

In the geographic information domain the number of data sets available as Web services increased in the last decades [1]. However the distributed data sets encoded in various formats remain isolated data islands. Individual features included in the data sets cannot be accessed by clients via URIs. There is also a lack of links to related data entities. Hence combining standardized geospatial data from various Web

sources yielding information about entities in the real world is challenging. Publishing geospatial data as Linked Data helps to solve these issues and offers many advantages. It enables and encourages linking between spatial objects in various data sets and supports users to construct sophisticated queries to filter information. Data sets can be associated with non-spatial data which may lead to creative applications. In addition geospatial data can easily be retrieved and reused by a larger group of users outside of SDIs which are unaware of the Web service standards of this domain.

The remainder of this paper is organized as follows. In the next section, we describe the Linked Data principles, review the standards it is built on and introduce related applications combining Linked Data and SDIs. Section 3 presents a concept of how to publish geospatial data as Linked Data and shows examples of how to query them. An implementation which converts SDI Web service metadata and its provided content to Linked Geo Data is presented in Section 4. Finally, we conclude this paper with a discussion about the benefits and open challenges with the use of Linked Data in geographic information standardization.

2. The Web of Data

Linked Open Data is a solution to create shared and structured information spaces [2] which include links between related information stating the nature of the connection. Its purpose is to create and connect related data on the Web with typed links being able to query the Web as a single database. People do not connect the documents from different Web sources which are often nested in HTML Web pages. They directly link the data entities. This approach leads to the transition from data islands to a global database and often refers to the term "Web of Data".

2.1. Linked Data Principles and Standards

To realize the "Web of Data", published data has to follow the Linked Data principles introduced by Tim Berners-Lee in his Web architecture note on *Linked Data* [3]: (1) use URIs as names for things, (2) the data is Web addressable via dereferencable URIs, (3) data is encoded in the machine-readable Resource Description Framework (RDF), (4) data can be queried with the RDF query language (SPARQL), and (5) data are interlinked via RDF links. RDF links enable to navigate from one data entity within one data set to related data items within other sources. In addition to his principles, Tim Berners-Lee proposes to adhere to the REpresentational State Transfer (REST) architectural style that governs the Web. For the implementation the Hypertext Transfer Protocol (HTTP) is used mainly as an application protocol which includes the use of its stateless methods (GET, POST, PUT, DELETE) to interact and manipulate the Web resources.

The RDF data interchange model, specified by the W3C, is based on the idea of making statements about resources. These statements take the form of subject-predicate-object expression (also called triples). A list of triples can be viewed as a directed graph, where the nodes are the subjects and objects and the predicates are the edges. Any of the three values in a single triple can be represented via a URI. The subject of a triple is the name of a thing which is described. The predicate is the property which indicates the relationship between the subject and the object. The object can either be a resource

(represented via a URI) or a literal value associated to the subject by the predicate. Schema languages such as RDF-Schema (RDFS) for the encoding of taxonomies, the Web Ontology Language (OWL) for the encoding of ontologies (i.e. formal representation of knowledge), and rules encoded with the Rule Interchange Format (RIF) can define classes and properties that help the understanding of statements. The statements can be stored in RDF repositories known as *triplestores* which can be built on top of a database or a file system.

The SPARQL query language can be used for querying those repositories, if they are exposed with a Web endpoint that implements the SPARQL protocol. SPARQL supports users in formulating more sophisticated queries. Information of relations between various RDF documents can be retrieved by querying across different sources. It also provides capabilities to easily combine information from different sources by merging two sets of triples into a single RDF graph [4]. Thus, new information can be created from the resulting dataset. SPARQL is useful for querying of relations that are explicitly represented in data, but implicit relationships, such as geospatial relations, cannot easily be queried.

The specification of the GeoSPARQL [5] standard is an important progress towards combining Linked Open Data and SDIs which aims to address such issues. It is a vocabulary for representing geospatial data in RDF which is consistent with OGC's Simple Features Model [6]. In addition, GeoSPARQL is a function library for SPARQL which improves the spatial query capabilities of the Web of Data by supporting geometrical or topological functions. Geometrical functions like "intersection" return a new geometry or another datatype, while topological tests, such as "crosses", return Boolean values.

2.2. Linked Geospatial Data Projects

In the recent years, Linked geospatial data has recently received attention, as the Geo Web Community is concerned with adding a spatial and temporal dimension to the Web of Data leading to many approaches trying to combine LOD techniques with SDIs. Some spatial data producers such as the Ordnance Survey – UK, the National Geographic Institute of Spain, or the U.S. Geological Survey are influenced by the Linking Open Data project[1] which publishes various open datasets based on the principles of Linked Data. These data providers have published their datasets as Linked Data enabling other users to make RDF links to these datasets to assert the location of resources. In addition, Auer et al. [7] have published the crowd-sourced OpenStreetMap[2] data collection as Linked Data and have tried to link the data items to equal resources of the DBpedia [8] data set.

Other approaches produce RDF descriptions of geospatial data which are based on OGC encoding standards. For example, the active OGC Sensor Web Community convert measure and sensor data to RDF and make them accessible for a broad audience outside of SDIs through LOD-endpoints [9] [10]. In [11], Observations & Measurements (O&M) encoded sensor data is interlinked with other LOD-repositories. A proxy which is able to convert O&M observations retrieved from Sensor Observation Service (SOS) instances and provides the data via a RESTful interface is proposed in [12]. Another *proxy-*

---

[1] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

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

based solution for translating OGC Web service data according to Linked Data principles is presented in [13]. The proxy is wrapping INSPIRE Download services realized as Web Feature Services (WFS) and acts as a SPARQL-endpoint to allow custom queries to filter the INSPIRE data.

3. Linked Geospatial Data

Linked Open Data and Spatial Data Infrastructures share the same basic principles: to make data easier to discover, connect, and use. To combine both data worlds, a report on Linked Data was presented to the 32nd ISO/TC 211 plenary held in Delft, Netherlands, on May 2011 [14]. The report collected benefits and challenges and included the main recommendation to proceed with the use of Linked Data in Spatial Data Infrastructures to enable the linkage to other domain data (e.g. health). Therefore, this section presents patterns for publishing Linked Geospatial Data and shortly introduces tools which can be used for implementing such applications. In addition, we show with an example how a combination of various datasets might look like and illustrate the powerful query capabilities on such a dataset.

3.1. Patterns for Publishing Data as Linked Data in a SDI

Publishing Linked Geospatial Data requires adoption of the basic principles outlined in Section 2.1. For doing so, Figure 1 shows the most common strategies for publishing geospatial data as Linked Data in the form of workflows, from structured geospatial data or textual content (e.g. metadata) through to Linked Geospatial Data published on the Web. The chosen strategy depends on the amount and nature of the structured geospatial data (e.g. sensor data, or feature data).

Structured data stored in spatial databases can be converted through the use of database-to-RDF wrappers such as Triplify [15]. It is a front-end implementation which converts the results of SQL queries on relational databases into RDF. Another tool for publishing relational databases on the Web of Data is D2R Server[3].

Queryable structured geospatial data accessible through an OGC Web service or an API can be exposed as Linked Spatial Data with a wrapper developed according to the Web service or API specification. Geospatial data can also be directly converted into RDF. The output of this process can be either RDF triples stored in a RDF store or static RDF files. The latters can be served up on the Web using a classic Web server (e.g. Apache) or loaded in a triplestore with a Linked Data interface. If the triplestore does not offer such an interface, Pubby[4] can be used to provide a Linked Data interface for the SPARQL endpoint. It provides a browser interface to view the RDF triples in the repository in an intuitive style.

Where spatial data is embedded in text (e.g. metadata or Twitter[5] posts), entities can be retrieved by passing the text through information retrieval tools. The output can be structured and merged with

---

[3] http://d2rq.org/d2r-server

[4] http://wifo5-03.informatik.uni-mannheim.de/pubby/

[5] https://twitter.com/

existing RDF data. Examples of such entity extractors are DBPedia Spotlight[6], Calais[7], or Ontos[8]. These tools look for things or types in your text and try to annotate them with the Linked Data URIs of entities referenced in the documents.
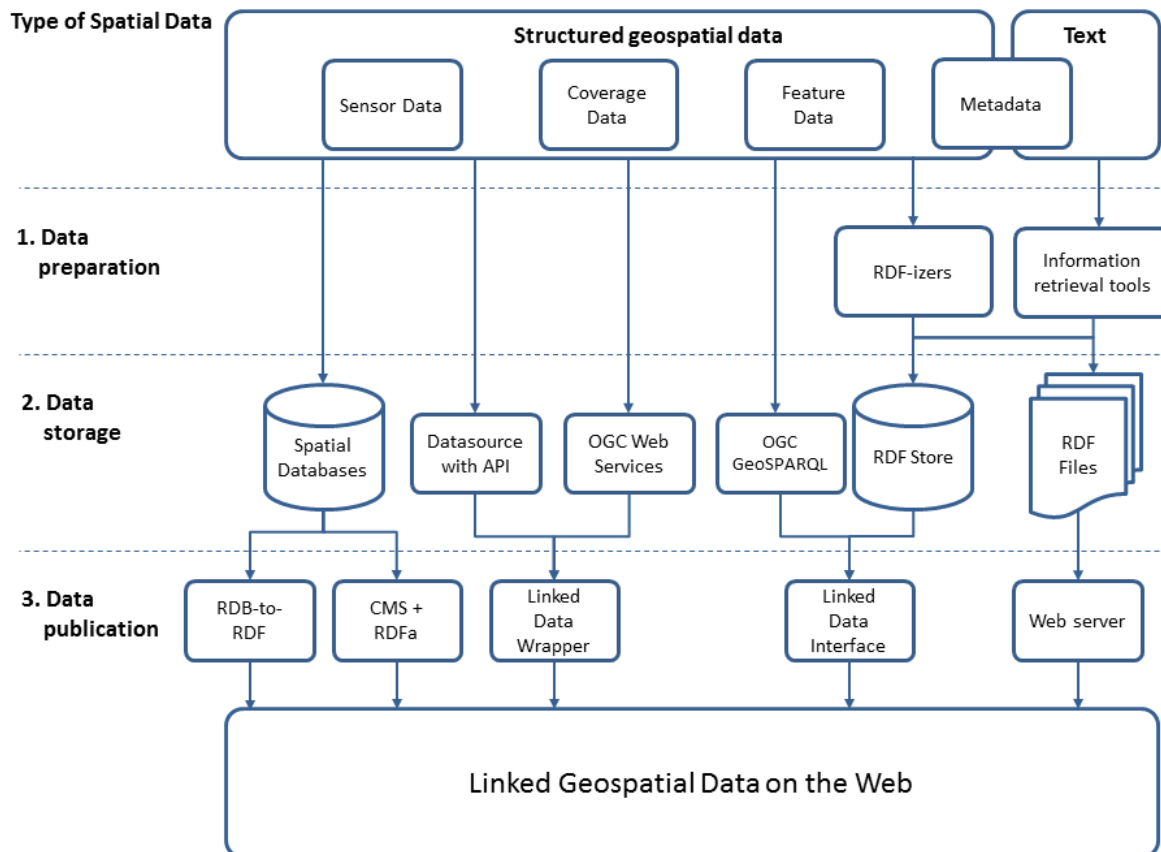

Figure 1: Linked Spatial Data publishing strategies. Adapted from Figure 2 in [16].

As mentioned before there is an increasing amount of structured spatial data in the Web of Data. To provide geospatial querying and reasoning capabilities in RDF stores, the data providers have to define both a vocabulary (i.e. ontology) for representing the spatial entities in RDF, and query predicates to discover these data entries. However, each spatial data producer has specified different vocabularies to express their data. Hence, the data can be only exploited in one single implementation and it is nearly impossible to combine it with other related data entries of another source. Therefore, for the cross-domain spatial exploitation of the data, we propose to either enrich the existing datasets with

---

[6] http://spotlight.dbpedia.org/

[7] http://www.opencalais.com/

[8] http://www.ontos.com/

GeoSPARQL representations or to use the GeoSPARQL ontology directly when implementing an application from scratch. It enables that the data can be properly queried and indexed in RDF stores. To take advantage of GeoSPARQL encoded Linked Spatial Data, it is necessary that these RDF stores provide a SPARQL endpoint which understands the GeoSPARQL ontology and provides support for performing spatial queries. However, the most of the existing triple stores cannot be utilized for spatial queries. An exception is e.g. Parliament[9], a triple store implementation that supports the GeoSPARQL specification. It allows GeoSPARQL data to be indexed and offers a query engine that supports GeoSPARQL queries. The triple store does not include a query processor, but is paired with a query processor such as Sesame[10] or Jena[11]. The latter should be used to support named graphs and temporal and geospatial indexes.

3.2. Geospatial data encoded in RDF

In the following, we present an example how to encode a dataset with spatial context in RDF and how to interlink it with sensor data published as Linked Data. First, each dataset is represented by a concept map which is a diagram representing the RDF triples as a graph. Concepts expressed by the same vocabulary are shown in the same color. In a second step, we show the corresponding RDF representations. Here, we use the more readable RDF serialization format N3 [17].

---

[9] http://parliament.semwebcentral.org/

[10] http://www.openrdf.org/index.jsp

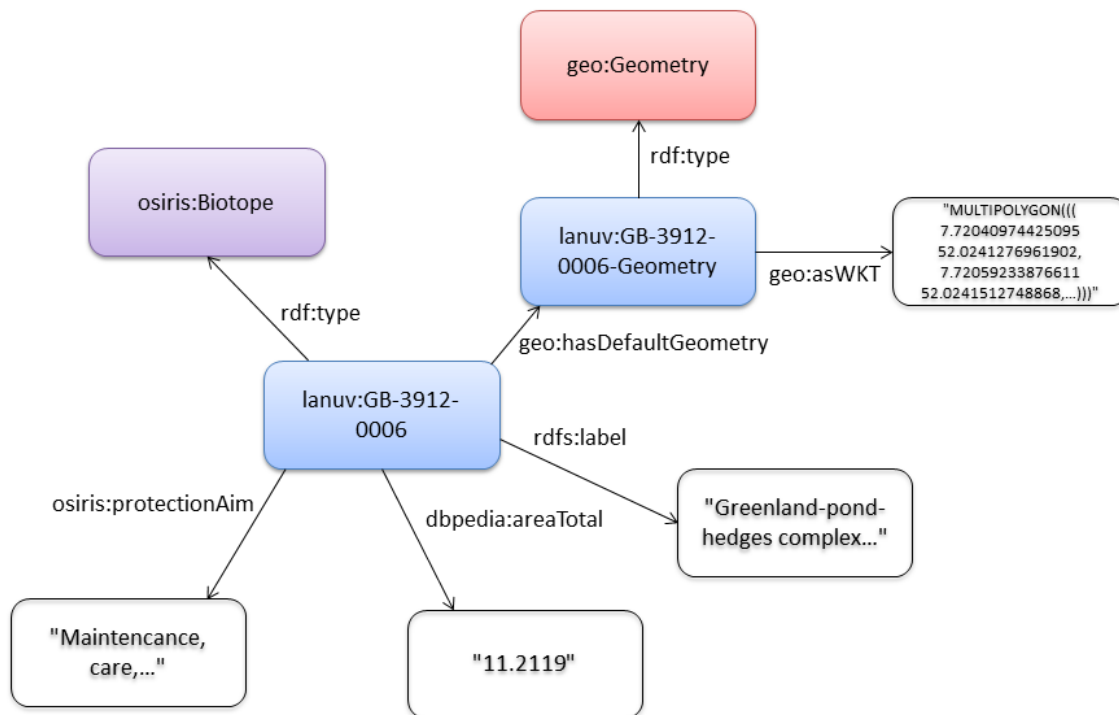[11] http://jena.apache.org/index.html

Figure 2: A concept map showing statements about a habitat

The structure of the example dataset shown in Figure 2 is inspired by the existing real-world data management system OSIRIS[12] used by the State Agency for Nature, Environment and Consumer Protection North Rhine Westphalia (Lanuv NRW)[13]. The dataset contains information about environmental habitats. The habitat is an instance of the class `osiris:Biotope` and includes many non-spatial attributes like a label or the protection aim. The GeoSPARQL vocabulary is used for representing the geometry that is a multipolygon with coordinates in the WGS84[14] datum. This assumes that the class `osiris:Biotope` is a subclass of `geo:Feature` in order to inherit the `geo:defaultGeometry` property. The coordinates are defined as literals in the well-known text (WKT) format. However, it is also possible to express them as a GML string via the `geo:asGML` property. Listing 1 shows the corresponding N3 serialization of the described dataset.

```
@prefix dbpedia: <http://dbpedia.org/ontology/>.
@prefix geo: <http://www.opengis.net/ont/geosparql#>.
@prefix osiris: <http://www.lanuv.nrw.de/osiris/>.
```

---

[12] http://www.osiris-projekt.de/

[13] http://www.lanuv.nrw.de/

[14] http://en.wikipedia.org/wiki/World_Geodetic_System

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

@prefix lanuv: <http://www.lanuv.nrw.de/osiris/biotope/>.


lanuv:GB-3912-0006 a osiris:biotope;

      rdfs:label "Greenland-pond-hedges complex...";

      dbpedia:areaTotal "11.2119";

      osiris:protectionAim "Maintenance, care, ...";

      geo:hasDefaultGeometry lanuv:GB-3912-0006-Geometry.


lanuv:GB-3912-0006-Geometry a geo:Geometry;

      geo:asWKT "MULTIPOLYGON(((7.72040974425095 52.0241276961902, 7.72059233876611

      52.0241512748868,…)))".
```

Listing 1: Dataset describing a habitat (serialized in N3)

The next example shows a sensor observation dataset (see Figure 3) which measures the concentration of carbon monoxide to provide information about the current air quality. Listing 2 shows the corresponding N3 serialization of this dataset. The origin sensor data is provided in OGC's Observations & Measurements (O&M) model which is used to store and publish sensor data. In this model, an observation consists of information about the sensor, the observed phenomenon, the observed feature, the time when the observation was taken, and the observation's result. The Semantic Sensor Network (SSN) ontology[15] developed by the W3C Semantic Sensor Network Incubator Group (W3C SSN-XG)[16] is similar to the O&M model and therefore facilitates the mapping from O&M into RDF. Hence, it is also used in this example to represent the sensor data in RDF. Since the sensor is installed at the habitat of the former example, the `ssn:featureOfInterest` property of the `obs:Observation` concept can be linked to the `lanuv:GB-3912-0006` habitat resource of the first example. To model the observed property 'carbon monoxid', the observation is connected to a resource (`sweet:CarbonMonoxide`) of the SWEET upper-level ontology[17] including thousands of concepts which are relevant to earth and environmental science. The graph can be linked to more non-spatial datasets. For example, the habitat resource can be connected to each dataset describing a species occuring in the biotope. Then it is possible to infer information about which species are affected if the amount of carbon monoxide in the air exceeds a threshold which is toxic to animals.

---

[15] http://purl.oclc.org/NET/ssnx/ssn

[16] http://www.w3.org/2005/Incubator/ssn/

[17] http://sweet.jpl.nasa.gov/

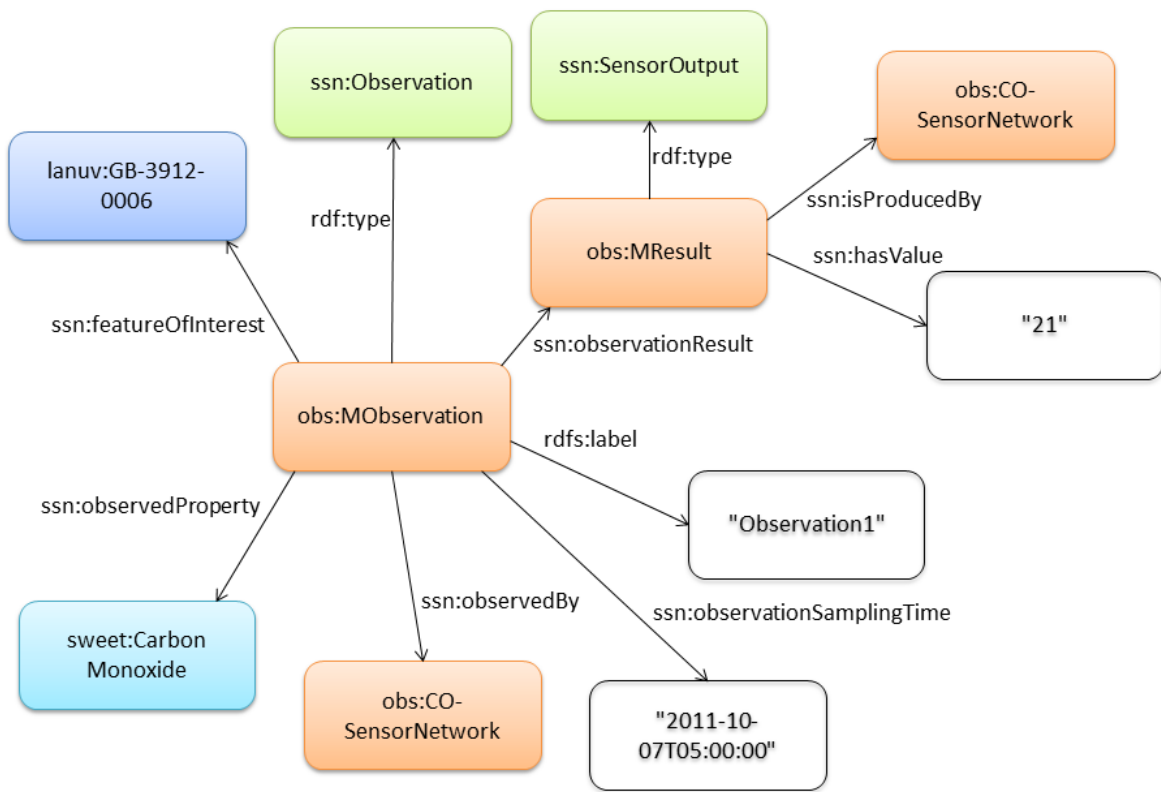**Figure 3: A concept map showing a sensor observation measuring the concentration of carbon monoxide**

```
@prefix ssn: <http://purl.oclc.org/NET/ssnx/ssn#>.

@prefix osiris: <http://www.lanuv.nrw.de/osiris/>.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

@prefix lanuv: <http://www.lanuv.nrw.de/osiris/biotope/>.

@prefix obs: <http://52north.org/observations/>.

@prefix sweet: <http://sweet.jpl.nasa.gov/ontology/substance.owl#>.


obs:MObservation a ssn:Observation;

     rdfs:label "Observation1";

     ssn:observedProperty sweet:CarbonMonoxide;

     ssn:featureOfInterest lanuv:GB-3912-0006;

     ssn:observedBy obs:CO-SensorNetwork;

     ssn:observationSamplingTime "2011-10-07

     T05:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

     ssn:observationResult obs:MResult.
```

```
obs:MResult a ssn:SensorOutput;

     ssn:isProducedBy obs:CO-SensorNetwork;

     ssn:hasValue 21.
```

Listing 2: Sensor data measuring the concentration of carbon monoxide (serialized in N3)

3.3. Querying Linked Geodata

In order to discover RDF triples in the Web of Data, users can query the data with SPARQL via a SPARQL endpoint. In this section, we will look at example SPARQL and GeoSPARQL queries using the datasets of the previous section to demonstrate the capabilities that each standard provides. One potential query over the sensor data shown in Listing 2 would be, "Which sensors observe a concentration of carbon monoxide that is higher than 20 ppm?" This query is shown in Listing 3. Table 1 shows that there is one feature matching this query.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX ssn: <http://purl.oclc.org/NET/ssnx/ssn#>

PREFIX sweet: <http://sweet.jpl.nasa.gov/ontology/substance.owl#>


SELECT ?observation ?result

WHERE  { ?observation a ssn:Observation;

          ssn:observedProperty sweet:CarbonMonoxide ;

          ssn:observationResult ?sensorOutput.

       ?sensorOutput ssn:hasValue ?result.

       FILTER (?result > xsd:integer('20') )

     }
```

Listing 3: Example SPARQL Query 1

observation	result
http://52north.org/observations/MObservation	21

Table 1: Example SPARQL Query 1 results

In order to perform spatial queries, we used the GeoSPARQL vocabulary to represent the geometry of the triples shown in Listing 1. Consider the query in Listing 4. This query is asking for all features within the biotope described by <http://www.lanuv.nrw.de/osiris/biotope/GB-3912-0006>.

```
PREFIX ex: <http://www.52nexample.org/Feature#>

PREFIX geo: <http://www.open<gis.net/ont/geosparql#>

PREFIX lanuv: <http://www.lanuv.nrw.de/osiris/biotope/>
```

```
SELECT ?feature

WHERE    { ?feature a ex:Feature;
                geo:hasGeometry ?mgeo.
           lanuv:GB-3912-0006 geo:hasDefaultGeometry ?bgeo.
           ?mgeo geo:within ?bgeo.
         }
```

Listing 4: Example GeoSPARQL Query 1

A more sophisticated query is shown in Listing 5. It asks for all animals within a habitat which might get a carbon monoxide poisoning since the concentration of carbon monoxide in the air is higher than 100 ppm.

```
PREFIX ex: <http://www.52nexample.org/Feature#>

PREFIX geo: <http://www.open<gis.net/ont/geosparql# >

PREFIX lanuv: <http://www.lanuv.nrw.de/osiris/biotope/>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX ssn: <http://purl.oclc.org/NET/ssnx/ssn#>

PREFIX sweet: <http://sweet.jpl.nasa.gov/ontology/substance.owl#>


SELECT ?animal

WHERE    { ?observation a ssn:Observation;
                ssn:observedProperty sweet:CarbonMonoxide ;
                ssn:featureOfInterest ?foi.
                ssn:observationResult ?sensorOutput.
            ?sensorOutput ssn:hasValue ?result.


            ?animal a ex:Animal;
                geo:hasGeometry ?mgeo.
            ?foi geo:hasDefaultGeometry ?foigeo.
            ?mgeo geo:within ?foigeo.
            FILTER (?result > xsd:integer('100') )
          }
```

Listing 5: Example GeoSPARQL Query 2

4. Converting SDI Services to Linked Geo Data

In the previous sections, we have only described conceptual approaches of how to provide and query geospatial data published as Linked Data. In this section, we present the semantic annotation pattern of the research project ENVISION[18]. It can be used to integrate descriptions of SDI Web services compliant to the standards of the OGC, such as WFS or SOS, into the Web of Data. These RDF descriptions are required by a second implementation (described in Section 4.2) which converts GML features and O&M observations into RDF.

4.1. Converting SDI Service Metadata to Linked Data

The ENVISION research project aims to provide a Web-based, environmental decision support platform assisting non ICT-skilled users to create and execute environmental models by integrating semantic technologies [18]. The semantic technologies help to minimize the potential of semantic conflicts ending in misunderstandings. Semantic conflicts emerge in communication processes, e.g. conversations between machines, due to semantic heterogeneities. Typical reasons are differences in language, culture, or knowledge in the field of interest. Non-semantic metadata are not sufficient to overcome these issues. To achieve semantic interoperability, i.e. the avoidance of semantic conflicts [19] within the environmental models, the geospatial Web services have to be semantically annotated. Semantic annotations establish links between Information Technology (IT) resources and explicit vocabularies explaining what the resource represents in the real world. Semantically annotated resources improve information retrieval, discovery, or the validation of requested resources to mention only some benefits [20].

The semantic annotations pattern used in the ENVISION project is described in [21] and illustrated in Figure 4. Geospatial data are served by an OGC Web service described by some XML-based metadata. The metadata comprises service-specific functional (i.e., where and how to access the service) and non-functional information (i.e., contact information). It additionally includes details about the data model. The metadata is linked to a Resource Vocabulary encoded in RDF. It re-models the service- and data-specific qualities of geospatial data via a service model and a data model. The former is based on the existing vocabularies for Procedure- or Resource-oriented Service Models (POSM/ROSM) [22]. Service-type specific extensions exist to support the various service interface standards from the OGC, i.e., the WFS, the SOS, and more. Each service model has a reference to the data model vocabulary. It describes the inner structure, i.e. the schema, of the provided data by instantiating concepts from vocabularies representing the OGC standards like GML or O&M. In addition, it can also include data quality information. As described in [23], the semantic annotation is an extension which links elements of resource vocabularies with globally shared concepts of domain vocabularies representing the terminology accepted within one information community. This connection is called domain reference and is provided through rules that are part of the data model.

---

[18] The acronym ENVISION stands for "Environmental services infrastructure with ontologies".
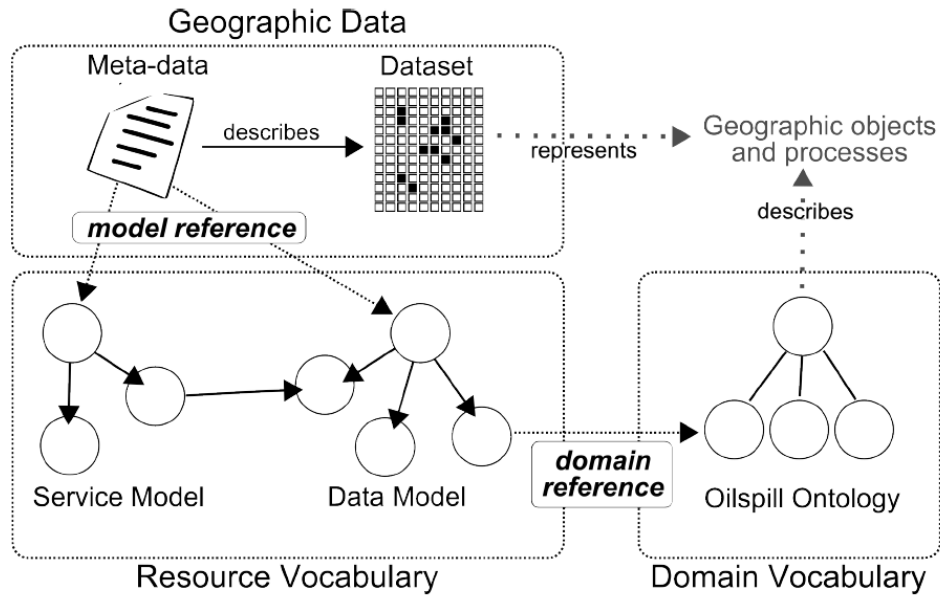
Figure 4: The Semantic Annotations Pattern of the ENVISION project described in [21].

A translation to the various RDF-based service descriptions is achieved by the Service Model Translator (SMT). It is a JAVA API[19] that produces the introduced service/data model vocabularies for a set of OGC Web services by fetching the original service description (e.g. the *GetCapabilities* document). Listing 6 depicts an example of a resource vocabulary of a WFS serving features of type "City".

```
@prefix posm: <http://purl.org/ifgi/posm#>.

@prefix ows: <http://purl.org/ifgi/ows/0.2#>.

@prefix wfs: <http://purl.org/ifgi/wfs#>.

@prefix gml: <http://purl.org/ifgi/gml/0.2#>.

@prefix sawsdl: <http://www.w3.org/ns/sawsdl#>.

@prefix sm: <http://www.example52n.org/Corine_c122.rdf#>.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix dc: <http://purl.org/dc/elements/1.1/>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.


// SERVICE MODEL VOCABULARY

sm:WFS11_SimpleUseCase a posm:Service;

      dc:title "SimpleUseCase";

      rdfs:seeAlso <http://link_to_web_service_capabilities_document>
```

```
        sawsdl:modelReference wfs:WebFeatureService;
        posm:hasOperation ows:DefaultGetCapabilities,
              wfs:DefaultDescribeFeatureType, sm:GetFeature.


sm:GetFeature a wfs:GetFeature;
        dc:title "Get Feature: City";
        posm:hasInput sm:GetFeatureRequest;
        posm:hasOutput sm:GetFeatureResponse.


sm:GetFeatureRequest a wfs:GetFeatureRequest;
        dc:title "Get Feature Request: City";
        sawsdl:modelReference sm:GetFeatureInput.


sm:GetFeatureInput a rdfs:Class;
        dc:title "Get Feature Input: City";
        rdfs:subClassOf ows:RequestParameters.


sm:GetFeatureResponse a wfs:GetFeatureResponse;
        dc:title "Get Feature Response: City";
        sawsdl:modelReference dm:FeatureType.


// DATA MODEL VOCABULARY
dm:FeatureType a rdfs:Class;
        dc:title "City";
        rdfs:isDefinedBy sm:GetFeatureOutputDefinition;
        rdfs:subClassOf geo:Feature.


dm:hasProperty a rdf:Property;
        rdfs:domain dm:FeatureType;
        rdfs:range dm:the_geom, dm:label, dm:state, dm:area_code, dm:population;
        rdfs:subPropertyOf gml:hasProperty.


dm:the_geom a rdfs:Class;
        dc:title "the_geom";
        rdfs:subClassOf geo:Geometry.
```

```
dm:label a rdfs:Class;

     dc:title "label";

     rdfs:subClassOf gml:FeatureProperty.

...
```

Listing 6: Extract of a RDF document describing a WFS that provides data of cities

4.2. Converting SDI Contents to Linked Data

Besides a few exceptions, the integration of geospatial data provided by OGC Web services is still missing yet. In the following, we present an API deployed as a Web service[20] (named as Geographic Feature Pipes [GFP]) working towards closing this gap. It translates O&M sensor data as well as GML features served by WFSs into a RDF representation. The resource vocabulary described in the previous section is required by the service to translate the data, because the data model, representing their schema, provides the vocabulary to express the geospatial data in RDF.

The translation procedure is illustrated in Figure 5. In step (1), the client registers a resource vocabulary to the service and retrieves a URI acting as identifier (ID) for each translated RDF dataset. The client uses the URI to either request the converted data or to address the data within SPARQL queries using the FROM clause after the translation. In the backend, the ID is mapped to a SPARQL query used by the service to retrieve the translated data sets later on. Each query is stored as RDF statement in one RDF document identified by an internal context. The service also stores the request parameters for each data set in order to retrieve the original geospatial data if they are out-of-date. The Web service *name*, coupled with the *feature type*, forms the identifier for an RDF encoded feature collection served by the WFS. The query string of the first URL in Listing 7 contains an example ID. An ID required for obtaining linked sensor data is represented by the query in the second URL. It consists of the Web service *name*, the *observedProperty* identifier, and the *featureOfInterest*.

In step (2), the service reads the service URL, the service name, and the data type (either *feature type* or *observed property*) from the service model and requests the corresponding data from the underlying service. If the data consists of sensor observations, an observation collection is requested for each pair of observed property and related feature of interest parameter.

In step (3), the obtained data is read, then translated using the data model vocabulary into RDF and finally added to an RDF repository associated with the ID. If the client requests the translation results via this ID, the service checks if the requested data, particularly the dynamic sensor data, is up-to-date. If that is the case, the stream is directly forwarded to the client. Otherwise, the original data will be requested with the previously queried request parameters, then translated into RDF, stored in the repository and is finally sent back to the user.

---

[20] The Web service is available at: http://dunedin.uni-muenster.de:8080/GFP/

```
(1)    http://dunedin.uni-muenster.de:8080/GFP/get?id=WFS11_SimpleUseCase/City/s

(2)    http://dunedin.uni-muenster.de:8080/GFP/get?id=SOS10_Pegelonline-
       sos/Pegelnullpunkt/Koeln(2730010)/s
```

Listing 7: URLs to access the translated data. The IDs end with the character 's' in order to identify them as initial datasets.
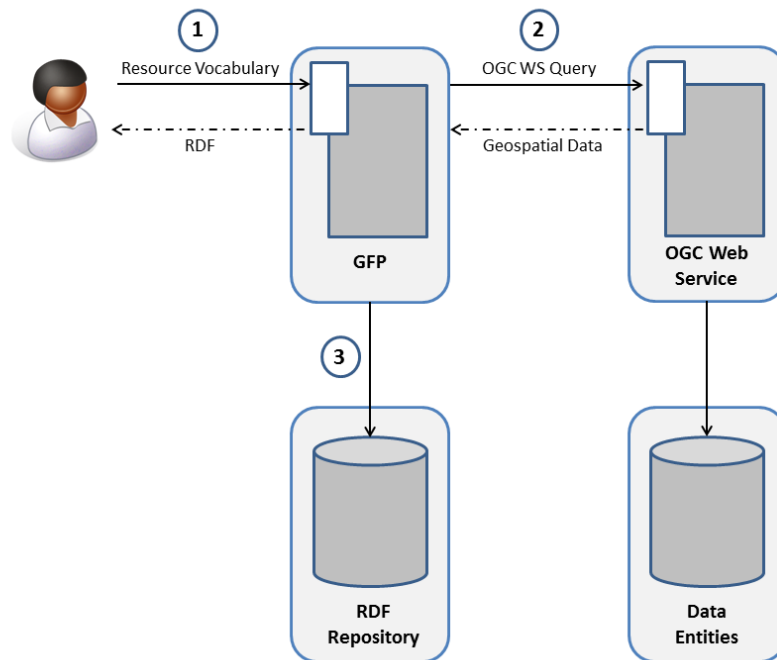
Figure 5: The translation procedure - Web Service converting OGC Data

Listing 8 shows the result of converting GML features served by the WFS, introduced in the previous section, into RDF. It shows a feature of type City of the translated feature collection. The feature and its properties are instances of the classes included in the registered resource vocabulary which namespace points to the generated ID. A dc:date property represents the creation time of this RDF encoded feature. This information is necessary to ensure that the user can get up-to-date information. The geometries are just represented as WKT strings. In the near future, they are compliant with GeoSPARQL.

```
@prefix gml: <http://purl.org/ifgi/gml/0.2#>.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

@prefix dc: <http://purl.org/dc/elements/1.1/>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

@prefix                          f:                          <http://dunedin.uni-
        muenster.de:8080/GFP/get?id=WFS11_SimpleUseCase/City/s/>.
```

```
@prefix rv: <http://dunedin.uni-
      muenster.de:8080/GFP/get?id=giso/WFS11_SimpleUseCase/City#>.


f:City_19 a rv:FeatureType;

      dc:date "2012-11-30T08:28:07+0200";

      gml:id "City.1";

      rdfs:label "Koeln";

      rv:hasGeometry _:node161pkco9fx91;

      rv:hasProperty  _:node161pkco9fx92,  _:node161pkco9fx93,  _:node161pkco9fx94,
      _:node161pkco9fx95.


_:node161pkco9fx91 a rv:the_geom;

      rdf:value "POINT (50.93341338662487 6.949928596019363)".


_:node161pkco9fx92 a rv:label;

      gml:hasValue "Koeln".


_:node161pkco9fx93 a rv:state;

      gml:hasValue "North Rhine-Westphalia".


_:node161pkco9fx94 a rv:area_code;

      gml:hasValue "0221".


_:node161pkco9fx95 a rv:population;

      gml:hasValue "1000660".
```

Listing 8: Extract of a GML feature collection encoded in RDF. This feature provides information about Cologne (Germany)

5. Conclusions & Outlook for the GLUES SDI

The vision of the Data Web assumes that information is published by following the Linked Data principles. This paper presents patterns to publish geospatial data as Linked Data and shows examples how to query the datasets. It shows a solution that contains on the one hand the Service Model Translator (SMT) which converts SDI Web service metadata to RDF, and on the other hand the Geographic Feature Pipes (GFP) which is able to convert geographic feature data to RDF.

Within the GLUES SDI, these principles and technologies can be applied to the already existing Web service infrastructure. The SMT can already be applied to the Web Processing Services (WPS), Web Feature Services (WFS), and Sensor Observation Services (SOS) to produce RDF representations of the service metadata, the so-called Capabilities. The GFP can already be applied to Web Feature Services (WFS), and Sensor Observation Services (SOS) to produce RDF representations of features or

observations, respectively. A current limitation of SMT and GFP is their inability to be applied to the Web Map Services (WMS) and Catalogue Services (CSW). Those services are mainly used in GLUES and hence, an extension of the available technology needs to be developed next.

Following the Linked Data principles within GLUES will have several benefits. First, publishing geospatial data as Linked Data breaks up data silos within the SDI itself but also towards the general Web. Utilizing Linked Data allows combining different data sources and datasets on-the-fly. Interlinked datasets leverage the discovery of information. Linked Geospatial Data also lowers the barrier to reuse spatial data by third parties. Clients can discover spatial datasets by browsing through interlinked data starting from non-spatial data. Metadata of spatiotemporal datasets can be embedded into Linked Data, which results in a seamless integration of geospatial content and its metadata. This applies for example to lineage and provenance information about datasets, which is already a topic within the MetaViz[21] application within the GLUES GeoPortal.

Besides the advantages of combining LOD and SDIs, there are still some challenges. First of all, opening up data has often licensing issues. Since Linked Data depends on HTTP URIs, data URIs have to be stable and the governance of URIs for common resources must be orderly. The infrastructures to resolve the URIs, need to be established and maintained. Responsibilities for link establishment and maintenance have to be identified as well. Increasing the performance when querying and accessing Linked Data over the Web is crucial. Furthermore, the usage of common vocabularies, such as GeoSPARQL, has to be ensured, otherwise client development is difficult.

6. References

[1]   P. Maué, S. Schade, *"*Data Integration in the Geospatial Semantic Web", *Cases on Semantic Interoperability for Information Systems Integration Practices and Applications,* vol. 11, no. 4, pp. 100–122, 2009.

[2]   C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far", *International Journal on Semantic Web and Information Systems (IJSWIS),* vol. 5, no. 3, pp. 1–22, 2009.

[3]   T. Berners-Lee, "Linked Data - Design Issues", 2006. Personal view available at http://www.w3.org/DesignIssues/LinkedData.html.

[4]   T. Heath, and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space", *Synthesis Lectures on the Semantic Web: Theory and Technology*, pp. 1-136, 2011.

[5]   M. Perry, and J. Herring, *OGC Implementation Specification 11-052-r4: OGC GeoSPARQL – A Geographic Query Language for RDF Data*, Open Geospatial Consortium, 2012. Available at: http://www.opengeospatial.org/standards/geosparql

[6]   J.R. Herring, *OpenGIS Implementation Standard 06-103r4: Simple feature access*

---

[21] http://geoportal.glues.geo.tu-dresden.de/geoportal/Applications/metaviz.html

- Part 1: Common architecture, Technical report, Open Geospatial Consortium, 2010.

[7]   S. Auer, J. Lehmann, and S. Hellmann, "LinkedGeoData: Adding a Spatial Dimension to the Web of Data", *The Semantic Web-ISWC 2009*, pp. 731–746, 2009.

[8]   S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A Nucleus for a Web of Open Data", *The Semantic Web*, pp. 722–735, 2007.

[9]   H. Patni, C. Henson, and A. Sheth, "Linked Sensor Data", *Collaborative Technologies and Systems (CTS)*, pages 362–370, IEEE, 2010.

[10]   K. Page, D. D. Roure, K. Martinez, J. Sadler, and O. Kit, "Linked sensor data: RESTfully serving RDF and GML", *Second Int'l Workshop on SSN2009*, 2009.

[11]   C. Keßler, and K. Janowicz, "Linking Sensor Data - Why, to What, and How?", *The 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010)*, 2010.

[12]   K. Janowicz, A. Bröring, C. Stasch, S. Schade, T. Everding & A. Llaves, "A RESTful Proxy and Data Model for Linked Sensor Data", *International Journal for Digital Earth*, 2011.

[13]   S. Tschirner, A. Scherp, S. Staab, "Semantic access to INSPIRE", *Terra CognitaWorkshop 2011*, 2011.

[14]   J. Brodeur, "Linked Data – Report to 32nd ISO/TC 211 plenary", Technical Report, ISO/TC 211 Adhoc Group on Linked Data, Delft, Netherlands, 2011.

[15]   S. Auer, S. Dietzold, J. Lehmann, S. Hellmann, and D. Aumüller, „Triplify: Light-Weight Linked Data Publication from Relational Databases", *Proceedings of the 18th international conference on World wide web*, pp. 621–630, ACM, 2009.

[16]   FJ. López-Pellicer, Vilches-Blázquez, FJ. Zarazaga-Soria, PR. Muro-Medrano, and O. Corcho, "The Delft Report: Linked Data and the challenges for geographic information standardization", *Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE 2011),* Barcelona, España, 2011.

[17]   T. Berners-Lee, and D. Connolly, "Notation3 (N3): A readable RDF syntax", *W3C Team Submission*, 2008. Retrieved from http://www.w3.org/TeamSubmission/n3/.

[18]   P. Maué, D. Roman, "The ENVISION Environmental Portal and Services Infrastructure", *Proceedings of International Symposium on Environmental Software Systems (ISESS),* June 27-29, Brno, Czech Republic, 2011.

[19]   W. Kuhn, "Semantic Interoperability", *Encyclopedia of Geographic Information Science*, pp. 383–384, SAGE Publications Inc., 2007.

[20]   P. Maué, S. Schade, and P. Duchesne, "Semantic Annotations in OGC Standards", Technical report, Open Geospatial Consortium (OGC), 2008.

[21]   P. Maué, and M. Roth, M., "Lost in Translation – Mediating between distributed environmental resources", In R. Seppelt, A. Voinov, S. Lange, & D. Bankamp (Eds.), *Proceedings of 2012 International Congress on Environmental Modelling and Software (iEMSs'12)*, IEMSS, 2012.

[22]   R. Krummenacher, B. Norton, A. Marte, A. Berre, A. Gómez-Pérez, K. Tutschku, and D. Fensel, „Towards Linked Open Services and Processes", *FIS'10 Proceedings of the Third future internet conference on Future internet*, pp. 68-77, Springer-Verlag Berlin, 2010.

[23]   P. Maué, H. Michels, and M. Roth, "Injecting semantic annotations into (geospatial) web service descriptions", *Semantic Web Journal (SWJ)*, 2011.