(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

Applying Semantic Linkage in the Geospatial Web

Aneta J. Florczyk, Francisco J. Lopez-Pellicer, Rubén Béjar, Javier Nogueras-Iso and F.Javier Zarazaga-Soria

Computer Science and Systems Engineering Department
University of Zaragoza, Spain
{florczyk, fjlopez, rbejar, jnog, javy}@unizar.es

Abstract

The Semantic Web is an attempt to add meaningful annotations to Web resources, services and content that requires developing reference ontologies, which help to understand these annotations. The venue of the Web of Data makes the geographic information, which has become an important part of the current Web, widely usable.
This paper demonstrates how the Geospatial Web might take advantage from the Semantic Web. The show case is a services catalog dedicated to support the visualization applications based on on-the-fly data integration. The presented infrastructure for improving the catalog functionality applies an *administrative geography*, i.e. an ontology of political organization of the territory, published as Linked Data. The principal advantage of this approach is reflected by enhancing the functionality of the user application.

1 Introduction

From the very beginning, the growth of the Web has been uncontrolled and rapid, implying as a result the creation of disorganized content loosely

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

connected. Soon, it was obvious that information searching and organizing techniques based on lexical analysis of the content were not satisfactory. A new approach has been needed to transform the Web from a document repository into an information resource which would improve reflection of human reasoning in sharing and processing information of Web resources by automated tools, such as search engines.

The machine-understandable Web resulted in the Semantic Web is based on the ideas of (1) semantic description of every resource available on Web and (2) knowledge representation for reasoning on the relation among concepts. Thus, the Semantic Web involves the idea of an ontology – "a formal, explicit specification of a shared conceptualisation" (Gruber, 1993). The semantics that capture the cognitive content of Web resources might be presented in different ways. The easiest way is to add simple *metadata*, e.g., specially designed tags in XML-based format. The semantics might be also represented as data models via other Web resources that provide conceptual structures, for example RDF (Klyne & Carroll, 2004). The most complex but also the most rich in meaning are ontology-based semantics expressed in the form of RDF+RDFS (Brickley and Guha, 2004) or OWL (McGuinness & Harmelen, 2004).

The combination of RDF documents and the HTTP protocol has gained recently considerable interest in the Semantic Web community, as it allows publishing structured data on the Web as *Linked Data* (Heath, 2009). This best practice from the Semantic Web offers logic references to any related resources. The potential of the created *Web of Data* consists of the identification of a concept via dereferenceable URI that permits retrieving the description of the concept from Web as the RDF document. This document may contain references to other documents about the same concept (i.e. identifies its instances) or states the logic relation with other concepts referenced via their URIs. In this simple manner it is possible to create a web of interlaced concepts.

Nowadays, geographic information has become an important part of the Web. The services, such as geocoding or map services, offered out-of-charge by commercial providers (e.g. Google, Yahoo) or open communities, such as OpenStreetMap (Haklay &Weber, 2008), have become essential elements of many Web applications. Historically, it has been difficult to integrate digital spatial data from different geoprocessing sources and integrate it to non-spatial information systems (McKee, 2004). The most important standardization bodies that deal with digital spatial data are (1) ISO/TC 211[1], which works on standardization in the field of digital geo-

[1] http://www.isotc211.org/

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

graphic information; and (2) Open Geospatial Consortium (OGC)[2], a voluntary consensus standards organization for interoperability issues of geospatial and location based services. Also, there are many important initiatives from official institutions which aim to improve geographic information organization and accessibility via Spatial Data Infrastructures (SDI), e.g. Infrastructure for Spatial Information in the European Community (INSPIRE) established by the European Union directive (Inspire, 2007). The common Implementation Rules (IRs), adopted as Commission Decisions or Regulations, ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community. The IRs are adopted in a number of specific areas (Metadata, Data Specifications, Netowrk Services, Data and Service Sharing, Monitoring and Reporting). The OGC services has been indicated as possible implementations of the INSPIRE compliant services, i.e., view services (OGC Web Map Service, WMS), download services (OGC Web Feature Service, WFS), invoke spatial service services (OGC Web Processing Service, WPS), transformation services (Application Profile of OGC WPS) and discovery services (OGC Web Catalogue Service, CSW). Following the Service Oriented Architecture approach the discovery service is the linking point responsible for the reusability of resources offered in SDI.

Geographic information has to be accompanied with a knowledge backbone to be appropriately handled due to its peculiarities (Egenhofer, 2002). Therefore, the proper semantic description of geographic information seems to be the first step in the improvement of its usage. The core of standards communities' efforts (e.g. ISO, OGC) focuses on *metadata* to describe services and their content, data models and spatial data itself. Recently, an OGC discussion paper has been published which proposes a methodology for referencing plain-text annotations to a backbone ontology (Maue, et.al., 2009). These additional annotations might be added on three levels, (1) resource meta-data (e.g., an OWS Capabilities Document), (2) data model (e.g. a GML Application Schema), and (3) data entities (e.g., a GML file). The formal specifications of concepts from the reference ontologies can be used then for tasks such as semantics-based information retrieval during workflow definition process.

As the contents of Spatial Data Infrastructures are concealed for common Web users, the Web community has created proper solutions applying successfully linking approaches, such as geographic Web platforms (e.g., semantic geocoding service of GeoNames[3]) and for publishing geo-data

[2] http://www.opengeospatial.org/
[3] http://sws.geonames.org/

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

(e.g., LinkedGeoData[4]). On the other hand, the Linked Data has been applied successfully in spatial solutions (e.g., DBpedia Mobile (Becker & Bizer, 2008)). Both technologies, OGC services and Linked Data, might complement each other. There are many advantages of applying the best practices from the Semantic Web, such as Linked Data, in SDI. For example, a created ontology of geographic features might link instances of the same geographic feature across different sources. This framework would provide an integrated view of geographic features rich in logic and spatial information. The richness of geographic feature description, direct or provided by linked sources, might be helpful in dealing with conflation for database integration, the well known problem from the database field (Dolbear & Hart, 2008). Additionally, such unified ontology might be used for defining and publishing complementary logic relations among geographic features (e.g., to represent different territorial organization) instead of creating new instances of OGC services. As for the Semantic Web community, the publication of geographic ontologies by official providers (i.e., a public administration organ) according to Linked Data principles might be a valuable source of references, for example the Administrative Geography of Great Britain (Goodwin, 2009).

One of the main advantages of applying the linking-based approaches to reference geographic features is the maintenance of the abstraction from their spatial representations. Usually, geographic features are characterized by blurriness of their footprints. Topological elements of physical world usually lack well defined conceptual boundaries and, consequently, this influences theirs spatial definition (e.g., rivers, chains of mountains). What is more, the computational representation of a geographic feature footprint is limited due to the resolution limits. Any geographic feature may be represented via different spatial objects which depend on the system application, the data model, and the applied technological solution. For example, a point is the best choice as a location reference while a polygon is for landscape visualization. Therefore the spatial object should be interpreted as one of the possible representations of an individual. Figure 1 presents the global view of spatial reference definition.

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

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

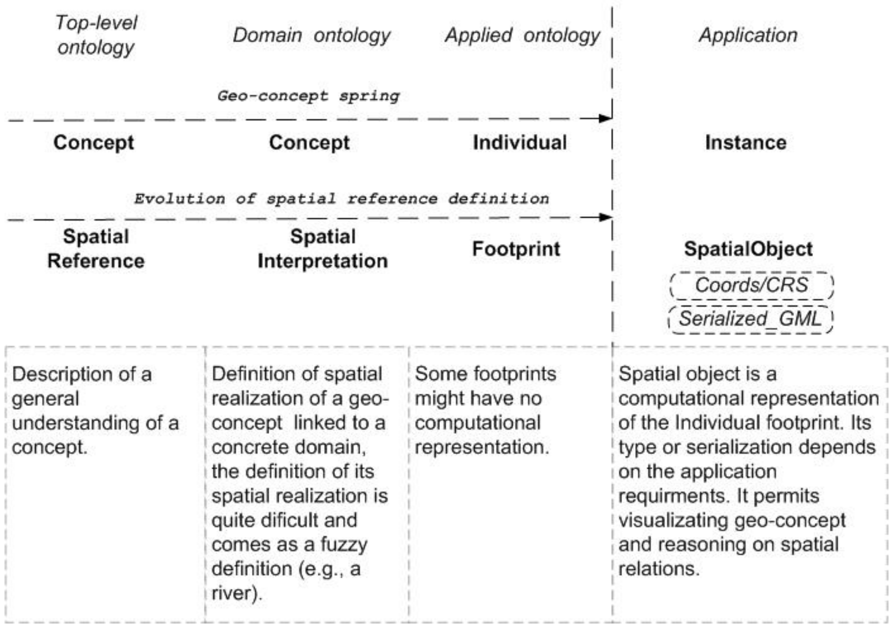

Fig. 1. Modelling the spatial representation of a geo-concept

The objective of this paper is to present how a SDI might take advantage of best practices from the Semantic Web. An *administrative geography* (AG), i.e. an ontology of political organization of the territory, published as Linked Data will be introduced as a relevant element of a SDI, since it permits (1) reasoning on logic relations among administrative units, and (2) accessing their footprints. In the use case, presented as an example later on, we will show how such ontology can improve functionality of the services catalog deployed within the Spatial Data Infrastructure of Spain (Nogueras-Iso, et.al., 2009) by applying geographic reasoning. One of the principal characteristics of a service deployed in a SDI is its geographic extent provided by the publisher as part of metadata description (i.e., *getCapabilities* response). According to the INSPIRE Metadata Implementating Rules (Craglia, 2009), the geographic location is defined as *minimum bounding box* (MMB). The descriptive metadata are used by the discovery service to answer the user requests and among the requestable parameters, there is the geographic location of the offered spatial data. Since services from SDI are provided by public administrations, frequently their geographic extent corresponds with the administrative area of the provider. The usage of MMB introduces false positive in the catalog response as the administrative areas are not rectangular. A catalog provided with knowledge about the hierarchy of administrative units and accurate

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

spatial objects of each administrative area might increase considerably the precision and recall of the requests, which is especially important for applications based on on-the-fly data integration.

The rest of this paper is organized as follows. After this introduction, second section presents the state of art in linking geographic features in the Semantic Web community (starting from Linked Data approaches to deal with geographic information, and ending with examples of semantic geographic platforms existing on the current Web) and the Geospatial Web. The next section describes the infrastructure for the improvement of services catalog and the prototype application. Finally, some conclusions are drawn and future work is outlined.

2 State of art

The geographic feature is managed differently in the Semantic Web and the Geospatial Web. The first community treats geographic features as the additional contextual information, which might link to some other concepts. For the Geospatial Web, the concept of geographic feature is the core element of a geographic platform. However, the Geospatial Web has focused on the interoperability issues maintaining the boundaries among the concepts from different geographic sources.

2.1 Linking geographic features in the Semantic Web

Nowadays, deploying the data gathered in relational data bases as Linked Data on the Web is possible using integrated technological solutions such as OpenLink Virtuoso (Erling & Mikhailov, 2007) or D2R Server (Bizer & Seaborne, 2004). These approaches are based on mapping data base models onto a reference ontology and may provide a Semantic Web Browser and a SPARQL client. It is also possible to join RDF data from different endpoints providing a transparent on-the-fly view to the end user (Langegger, et al., 2008). The most important initiative related to creating and publishing interlinked contents on the Web is the Linked Open Data project. In May 2009, the amount of Linked Data datasets consist of over 4.7 billion RDF triples interlinked by around 142 million RDF links (Bizer, et al., 2009). The RDF links are navigable using Semantic Web browsers, and Semantic Web Search Engines can apply sophisticated queries over crawled data. The expressive semantic queries might be executed via SPARQL access points as well.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

Since on-line contents involve geographic information, geographic features have become the part of Linked Data datasets, such as DBPedia, where geographic information has been extracted from Wikipedia (Auer, et al., 2008). Another example is the LinkedGeoData (Auer, et al., 2009b), which aims at adding geo-semantic meaning to the Web. It offers a Linked Geo Data Knowledge Base with RDF descriptions of more than 350 million spatial features from the OpenStreetMap database linked to DBPedia.

An example of applying Linked Data principles in location based solution is the DBpedia Mobile (Becker & Bizer, 2008), a location-aware client for the Semantic Web for mobile devices. The current user location is used to extract corresponding datasets from the underneath DBpedia database which are interlinked with various other location-related datasets.

An interesting proposal is Triplify (Auer, et al., 2009) which supports a kind of circular spatial requests. This system uses directly DB Views model as base for creating the RDF documents and URLs of published datasets, which facilitate the development. The underneath data base is responsible for processing spatial and semantic queries which are encoded explicitly in the request URL. The spatial query permits to retrieve the geographic features located in a circular region defined via a point and radius added to the request URL. The proposal enables limited spatial query (just a circular region) notwithstanding the Semantic Web techniques can not take advantage of this facility.

The idea of linking geographic features to create interlinked web influenced the appearance of geographic Web platforms, such as Yahoo! GeoPlanet[5] and GeoNames[6]. Both of them belong to a new branch of geocoders, the *semantic geocoders*, which return URIs to identify uniquely the named places instead of standardized textual description and location reference (e.g., a point). Although, they use the idea of linking, they are not following the pure Linked Data approach. GeoPlanet uses unique identifiers (URIs) to identify the named place which permits to retrieve its semantic description, however, the platform uses simple xml file instead of RDF. Additionally, the important spatial relations (e.g., “child”, “neighbour”, “siblings”) are encapsulated into the URI definitions. GeoNames is *almost* Linked Data based. It also uses unique identifiers of concepts to identify the named places. The RDF description of features contains spatial relations defined in the published OWL reference ontology. However, GeoNames distinguishes the *concept* from the *descriptive document*. The feature (i.e. concept) is identified via an URI but the geonames server uses

[5] http://developer.yahoo.com/geo/
[6] http://www.geonames.org/

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

303 redirection to display its location on map. The RDF description is available by adding the "/about.rdf" at the end of the feature URI.

The spatial requests supported by the presented solutions of current Semantic Geo Web are based on a branch of predefined spatio-logic relations (e.g., *near-by*, *belongs-to*, *child*, *siblings*). Since it is impossible to express all spatial relations among geographic features via definition of logic relations, the Semantic Web needs to use spatial representation of features. Currently, the spatial objects usually used in the Semantic Web are limited to points or MMB. The complex spatial requests that mix spatial objects and spatial relations defined in a reach ontology still remain the open issue.

2.2 Linking in the Geospatial Web

The requirement of unique identifiers for geographic features, *geoidentifiers*, in the geospatial community has been present from its beginning. Any Geospatial Web framework which publish information about geographic features uses unique identifiers (within this framework at least) and might be seen as a source of geoidentifiers. Therefore, any gazetteer (Hill, 2006) or OGC Web Feature Service from a SDI might be such a source of geoidentifiers. Currently, there are several instances of WFS services in the SDI of Spain, which frequently contain instances of the same geographic feature. The services model the geographic feature in different manner and use different identifier usually derived keys from relational databases, therefore, there exist problem of individual identification among different contents and consequently common problems of data integration.

One of the earlier proposals from the OGC community to apply common geographic identifiers for linking purposes is a framework based on *Geolinked Data Access Service* and *Geolinking Service.* This approach is dedicated to publish geographically linked information (e.g., statistical data) separately from spatial representation (spatial objects). Since this proposal is based on merging datasets from different sources by using a linkage field found in the sources, it fixes geographic data to only one spatial representation source and geoidentifiers are used only as syntactic link. In practice, it can be seen as technological facilitation for data publishers.

A proposal for providing the integrated view across distributed services is the EuroGeoNames project (Jakobsson, & Zaccheddu, 2009), a prototype of an integrated gazetteer for Europe from the INSPIRE directive. Apart from defining the data model to be followed by all community members, it provides rules for identifier definition. The named place identifier has to be composed of (1) its name, (2) two-letter ISO 3166 code (e.g., ES, NL) and (3) a code generated according to the BASE36-system

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

(e.g. “2YC67000B”). However, these identifiers still remain unique only in this distributed gazetteer.

Interlinking the corresponding instances of the same feature across different providers might be interesting for the Geospatial Web. It might be the way of adding logic relations among features and avoid the necessity of providing a new separate platform. The next section shows the application of an administrative geography of Spain published as Linked Data for the improvement of a SDI services catalog.

3. OGC services catalog application

The services catalog is one of the elements of a SDI. Since it is responsible for service discovery, its functionality determines the reusability of the offered services in the SDI and might be improved by applying best practices from the Semantic Web. This paper proposes a framework, where the administrative geography published as Linked Data is one of the core elements. A services catalog dedicated to support the visualization applications based on on-the-fly data integration is presented as a use case. The principal advantage of this approach is reflected by the improvement of the functionality of the end application.

3.1 Coverage issue

In INSPIRE, service discovery requests allow restricting the geographic extent of searched datasets and services to a required MMB. Frequently, the geographic extent of published data and service corresponds with the coverage of an individual from a geographic ontology (e.g., Europe from a geographic region ontology, Europe Union from a political organization ontology). Therefore, using MMB to describe available resources usually introduces false positives in the collection of results. For example, in the SDI of Spain the coverage of published service frequently corresponds with an administrative unit area of provider (e.g., council of Zaragoza). The administrative boundaries are far away from being rectangular and their MMBs overlaps significantly. The figure 2 shows the issue of overlapping MMBs of administrative areas. The shadowed rectangle represents the required MMB from the service discovery request. In this scenario the application would return services which geographic extend MMB corresponds with BBOX1, BBOX2 or BBOX3 when in reality it should provide only those services whose MMB corresponds with BBOX3.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

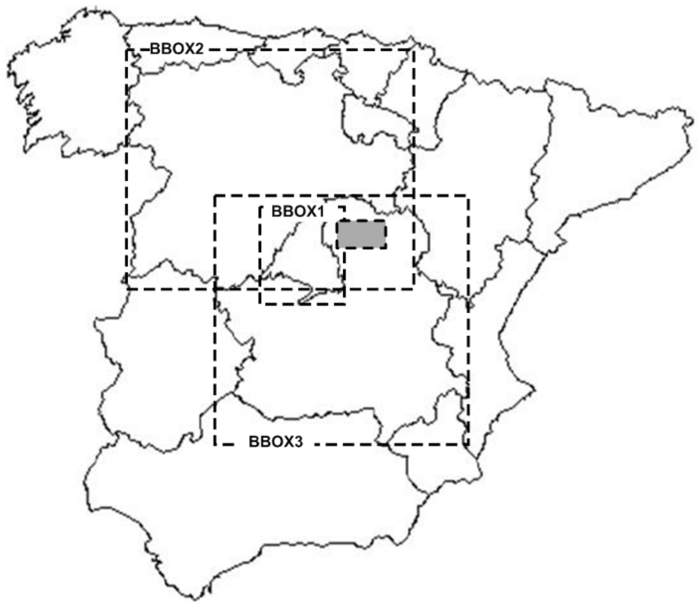

Fig. 2. Overlapping MMBs of administrative areas issue (Spain case).

The functionality of the catalog might be improved by operating on more precise spatial objects. A more promising approach could be application of the identifiers from an administrative geography which not only provides footprints but also permits reasoning on logic relations among concepts.

3.2 Administrative geography of Spain

Currently, there is no administrative geography published as Linked Data for Spain. Within the Spanish SDI there are various OGC WFS services which publish information about administrative unit entities. These services are provided by central and local authorities. Some of them offer administrative boundaries with different resolution (e.g., the *Infraestructura de datos Espaciales de España-WFS* service, IDEE-WFS[7]) separately for each level of administrative division, and others such as the gazetteers focused on gathering named places (e.g., IDEE-WFS-Nomenclator-NGC[8]), contain administrative units among published features. However, neither of these models permits to express the full administrative model that exists

[7]http://www.idee.es/IDEE-WFS/ogcwebservice?
[8]http://www.idee.es/IDEE-WFS-Nomenclator-NGC/services?

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

in Spain. The existing domain ontologies for modelling political organizations of the territory are usually based on the 'part-of' relation (parental relation). Such model is not enough flexible to scale the complexity of the territorial organization of countries, which apart from main division units (e.g., municipality, province and autonomous community in Spain) has to involve the units of different status (e.g., autonomous cities of Ceuta and Melilla, or associations of administrative units in Spain). Therefore, it is required usually a dedicated administrative ontology (e.g., Ordnance Survey) to model the political organization of territory of a country. In the case of Spain the Administrative Unit Ontology has been proposed as the domain ontology (Lopez-Pellicer, et.al., 2008). Apart from “part-of” and “has-part” relations, the “is-member-of” and “has-member” relations were defined to distinguish the association of the administrative units whose spatial representation might overlay the boundaries of direct parental units. An example of such association might be “comarca” of Aragon Autonomous Community which groups municipalities. Each municipality might lay in boundaries of only one province, however, one comarca might aggregate municipalities from different provinces.

The D2R Server has been used to publish the administrative geography of Spain, and to generate a data dump. The national Gazetteer, *IDEE-WFS-Nomenclator-NGC*, has been used as the reference source to extract the part of administrative geography. Although this model does consider logical relations among features (e.g., “parent-child”), the location of features in the administrative structure is defined indirectly by their names offered via *LocationEntity* element, whose structure contains concepts from the territorial organisations of Spain (e.g., *autonomous community*, *province, municipality* or *island*). Since the published data are not complete (e.g., there is no assignation of comarca names to municipalities), the INE online catalog (*Instituto Nacional de Estadística*, the National Statistics Institute of Spain) was used to complement the data. Then, the result data has been linked to their corresponding instances from different WFS services via the *skos:relatedMatch* relation. This approach produced the *administrative geography of Spain* published as Linked Data (see figure 3), and one of its advantages might be the maintenance of the references to different instances across the SDI of Spain.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.


```
@prefix au: <http://purl.org/iaaa/sw/gsw/ont/au-spain.owl#>.
@prefix agont: <http://purl.org/iaaa/sw/gsw/ont/app/agont.owl#>.
@prefix ag: <http://purl.org/iaaa/sw/gsw/georef/ag/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix skos: <http://www.w3.org/2009/08/skos-reference/skos.rdf#>.

ag:ZaragozaMun rdf:type au:Municipality;
      rdfs:label "Zaragoza"@es,"Zaragoza"@en;
      au:part-of ag:Spain, ag:AragonComunidad;
      au:member-of ag:ZragozaComarca;
      skos:relatedMatch <http://www.idee.es/IDEE-WFS/
ogcwebservice?SERVICE=WFS&VERSION=1.1.0&REQUEST=GetFeature&MAXFEATUR
ES=1&NAMESPACE=xmlns(ideewfs=http://www.idee.es/
wfs)&TYPENAME=ideewfs:BDLL1000Municipio&FILTER=%3CFilter%20xmlns:ide
ewfs=%22http://www.idee.es/
wfs%22%3E%3CPropertyIsEqualTo%3E%3CPropertyName%3Eideewfs:nombre%3C/
PropertyName%3E%3CLiteral%3EZARAGOZA%3C/Literal%3E%3C/
PropertyIsEqualTo%3E%3C/Filter%3E>;
(. . .)
      agont:bond-100 <http://www.idee.es/IDEE-WFS/
ogcwebservice?SERVICE=WFS&VERSION=1.1.0&REQUEST=GetFeature&MAXFEATUR
ES=1&NAMESPACE=xmlns(ideewfs=http://www.idee.es/
wfs)&TYPENAME=ideewfs:BDLL1000Municipio&FILTER=%3CFilter%20xmlns:ide
ewfs=%22http://www.idee.es/
wfs%22%3E%3CPropertyIsEqualTo%3E%3CPropertyName%3Eideewfs:nombre%3C/
PropertyName%3E%3CLiteral%3EZARAGOZA%3C/Literal%3E%3C/
PropertyIsEqualTo%3E%3C/Filter%3E>.
```

Fig. 3. An example of the RDF description that represents the Zaragoza municipality.

The section 3.3 describes an architecture and implementation of a services catalog component which applies the administrative geography of Spain during the service selection process.

3.3 Architecture and Implementation

The usage of geographic feature identifiers requires the annotation of registered resources in a catalog with corresponding geoidentifiers. Creation of metadata of registered services is one of the characteristics of the services catalog deployed in the SDI of Spain (Nogueras-Iso, et.al., 2009). Its architecture has been extended with the *Knowledge Content (KC)* that is responsible for the service search process (see Nogueras-Iso, et.al., 2009 for the description of the services catalog architecture). Figure 4 presents the main elements of the KC component. The KC has access to two RDF dataset sources: the *Administrative Geography* (AG), i.e. the administrative geography of Spain, and *Service Description Register* (SDR) which

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

contains RDF serialization of the registered service description. The reference ontologies, i.e. *Administrative Geography Ontology* (*agont*) and the *Service Description Ontology* (*svont*), are applied during the reasoning process. The concepts from the administrative geography (1) are linked to the entities from the reference WFS service, the source of boundary spatial definitions, and (2) the URI of the administrative units are used in the service description as indication of the geographic extent (*dc:coverage* property). Figure 5 represents an example of the link between the administrative geography and the service description.

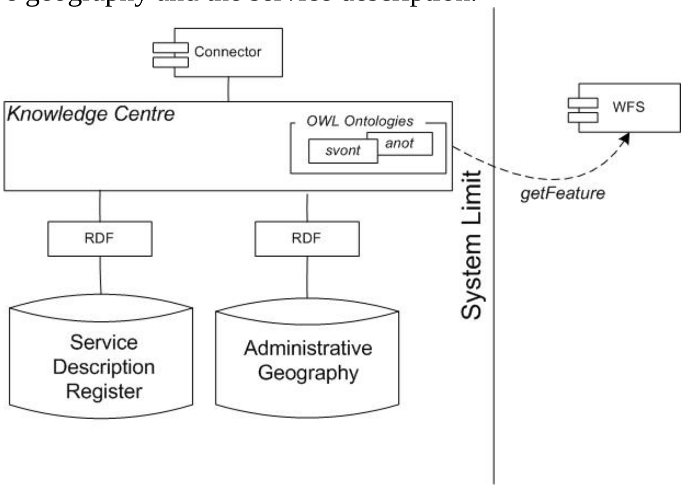

Fig. 4. Administrative geography as support for services catalog.

During the registration of a new service, the catalog uses the *getCapabilities* response to create proper description of the service. One of the elements is the service geographic extent expressed via MMB. The metadata model of the description has been extended with the *geoidentifier* metadata to contain an URI from the administrative geography of Spain. This element is not an ISO19119 element; therefore, it is neither visible to users nor published by the OGC CSW. The geoidentifier value is obtained from the analysis of the service MMB offered by provider. This MMB is used to request the KC that identifies the most extended administrative unit within the MMB (i.e., the *prime administrative unit*). Validation of the result consists in checking if the service provides any data from the disjoint area of the MMB and the prime administrative unit coverage (a set of retrieval tests). If it is impossible to identify the prime administrative unit (e.g. in the case of the hydrography service of the Ebro river basin, some data lies on France as well) or the validation fails, the URI of the *Non* concept (i.e., the disjoint concept with *administrativeUnit*) is returned. The

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

service registration ends by deploying the registered service description in the RDF container.

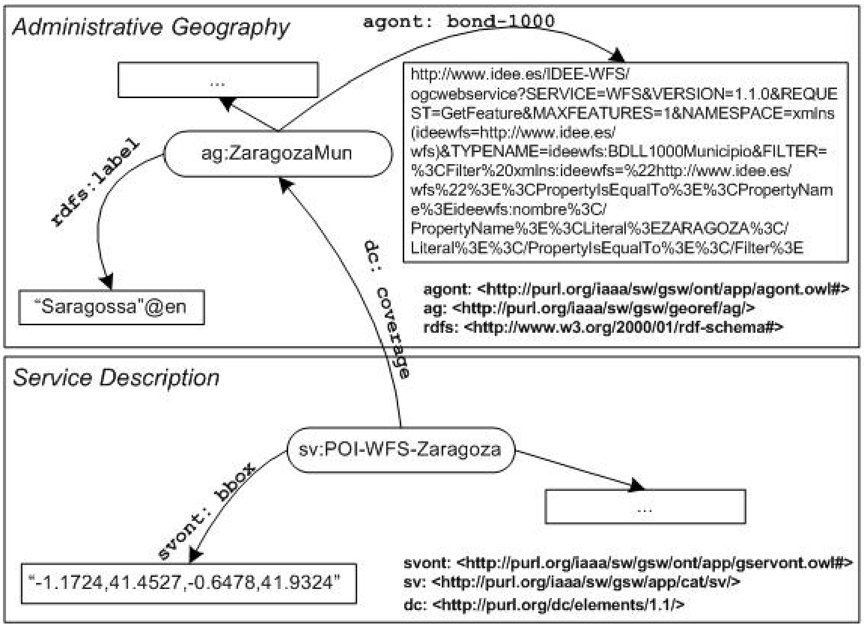

Fig. 5. Relations between the service description and the administrative geography.

The searching process for services with a MMB restriction exploits the semantic links between the semantic description of services and the features from the AG. The simplified example of a request (only the spatial part) consumed by KC is shown in figure 6. The request pattern uses as the input the searched MMB (*$BBOX*) expressed as literal (e.g., "1.16311, 41.0937, 1.7132, 41.6686"). The first part extracts those services which spatial reference of coverage (*dc:coverage*) refers to *administrativeUnits* which boundaries are in an interaction with the searched MMB (within it or intersects it). The *administrativeUnit* boundary is defined via the *bond-1000* property containing reference to the correspondent WFS entity, which is converted automatically into the spatial object by applying the profile instructions. The second part of the request extracts those services which geoidentifier is defined as *Non* and the collection is filtered similarly as in the previous version of services catalog, i.e. comparing the service MMB and the requested one.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.


```
PREFIX svont: <http://purl.org/iaaa/sw/gsw/ont/app/gservont.owl#>
PREFIX agont: <http://purl.org/iaaa/sw/gsw/ont/app/agont.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX sf: <http://purl.org/iaaa/sw/gsw/app/gfun>
SELECT ?s
WHERE {
  ?s rdf:type svont:service;
     dc:coverage ?x.
?x rdf:type agont:au;
     agont:bond-1000 ?g.
FILTER(sf:INTERACT(?g, $BBOX))
}UNION
SELECT ?s
WHERE {
?s rdf:type svont:service;
     dc:coverage agont:Non;
     svont:bbox ?g.
FILTER(sf:INTERACT(?g, $BBOX))
}
```

Fig. 6. SPARQL request pattern for service selection via a MMB ($BBOX is the MMB literal, eg., “-1.1724,41.4527,-0.6478,41.9324”) .

The KC component has required implementation of the spatial functions, such as *intersect*, or *within*, known form spatial data bases. Therefore, the Jena framework[9] has been chosen to deploy RDF datasets as its proprietary extension to SPARQL RDF query language (Jena ARQ[10]) permits to implement such additional functionality.

3.4 The use case

The catalogs that use precise spatial representation instead of MMB approximation might offer better functionality for the applications based on on-the-fly data integration. An example might be an application which allows displaying geographic information from different OGC services found in the services catalog (see figure 7). For improving the user experience, the list of selectable layers should depend on current displayed area.

[9] http://openjena.org/
[10] http://jena.sourceforge.net/ARQ/

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

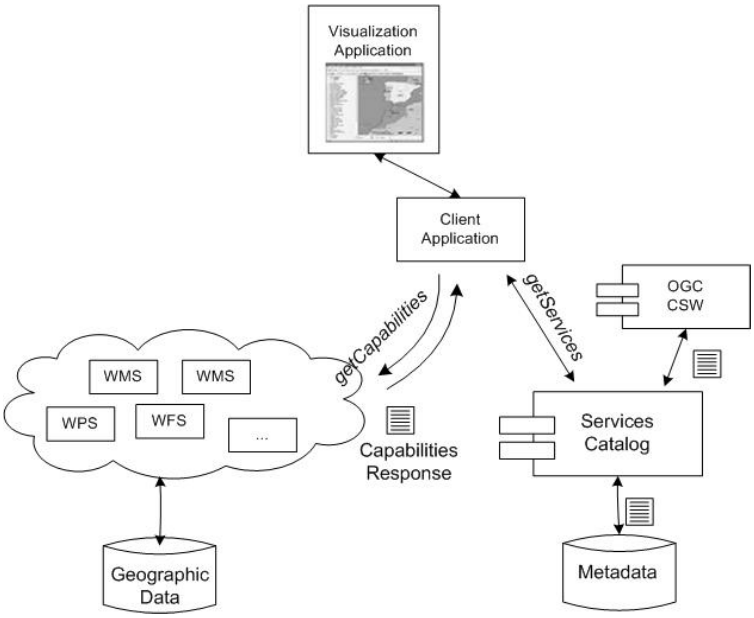

Fig. 7. The services catalog as the support component for application based on on-the-fly data integration.

The prototype of the improved services catalog has been used as the core component of Web application[11] which displays spatial data provided from different OGC services. The main disadvantage of this proposal is the response time of the WFS service. To solve this problem we have created a local repository of spatial objects retrieved previously from reference service. The cache techniques have also improved the catalog response time and the behaviour of the end application.

4 Results and future works

This paper shows the current approaches in referencing and identification of geographic features in the Semantic Web and the Geospatial Web. Applying best practices from the Semantic Web might be useful for the Geospatial Web. An administrative geography published in accordance with Linked Data principles might be useful for data integration as it permits referencing the geographic concept to the corresponding instances from other sources. Such ontology might be used as source of geoidentifiers in

[11] http://www.idee.es/IDEE-ServicesSearch/ServicesSearch.html

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

geospatial solutions and its main advantage lies in using more precise spatial representation and spatial reasoning on semantic level.

Additionally, the usage of geoidentifiers along with minimum bounding boxes to represent the service geographic extents might improve the recall of OGC services catalog. For instance, we have demonstrated that this improvement has enabled the development of web-based applications that facilitate on-the-fly data integration.

The principal advantage of using Linked Data technology in geospatial solutions is the possibility of explicit identification of features and abstraction of their spatial definition from footprint and computational representation. The different spatial representation might be accessible via linked instances and chosen according to the application requirements.

One of the future tasks will be applying the administrative geography of Spain as guideline to map instances of the same geo-concept individuals between two different gazetteers for merging purpose. Next, there will be investigated an adaptable framework to support complex spatial requests applying different spatial representations of features.

Acknowledgments

This work has been partially supported by Spanish Government (projects "España Virtual" ref. CENIT 2008-1030, TIN2007-65341 and PET2008_0026), the Aragón Government (project PI075/08), the National Geographic Institute (IGN) of Spain, GeoSpatiumLab S.L. The work of Aneta Jadwiga Florczyk has been partially supported by a grant (ref. AP2007-03275) from the Spanish government.

References

Auer, S., Dietzold, S., Lehmann, J., Hellmann, S. and Aumueller, D. (2009) Triplify – Light-Weight Linked Data Publication from Relational Databases, Proceedings of the 18th International World Wide Web Conference, April 20—24, 2009, Madrid, Spain, pp. 621—630.

Auer S., Lehmann J. and Hellmann S. (2009b) LinkedGeoData - Adding a Spatial Dimension to the Web of Data, In: 8th International Semantic Web Conference (ISWC2009), October 25—26, 2009, Washington, DC, USA.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. (2008) DBpedia: A Nucleus for a Web of Open Data, The Semantic Web, Proceedings of the 6th International Semantic Web Conference, the 2nd Asian Seman-

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

tic Web Conference (ISWC+ASWC 2007), November 23—30, 2008, Karlsruhe, Germany, pp. 722—735.

Becker, C. and Bizer, C. (2008) DBpedia Mobile: A Location-Enabled Linked Data Browser, Proceedings of the Linked Data on the Web Workshop, April 22, 2008, Beijing, China.

Bizer, C. and Seaborne A. (2004) D2RQ -Treating Non-RDF Databases as Virtual RDF Graphs, Poster at the 3rd International Semantic Web Conference, November 7—11, 2004, Hiroshima, Japan, http://www4.wiwiss.fu-berlin.de/bizer/pub/Bizer-D2RQ-ISWC2004.pdf, Last date accessed 10.2009.

Bizer, C., Heat, T. and Berners-Lee, T. (2009) Linked Data – The Story So Far, International Journal on Semantic Web and Information Systems.

Brickley, D. and Guha, R.V. (2004) RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, 2004, http://www.w3.org/TR/rdf-schema/. Last date accessed 10.2009.

Craglia M. (Eds.) (2009) INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Revised edition), Drafting Team Metadata and European Commission Joint Research Centre, 2009.

Dolbear, C. and Hart, G. (2008) Ontological Bridge Building - using ontologies to merge spatial datasets, Proceedings of the AAAI Spring Symposium on Semantic Scientific Knowledge Integration, AAAI/SSS Workshop, March 26—28, 2008, Stanford U, CA.

Egenhofer, M.J. (2002) Toward the Semantic Geospatial Web, Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems, pp.1—4.

Erling, O. and Mikhailov, I. (2007) RDF Support in the Virtuoso DBMS, Conference on Social Semantic Web, Vol. 113, 2007, pp. 59—68.

Goodwin, J.; Dolbear, C. and Hart, G. (2009) Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web Transactions in GIS, 2009

Gruber, T.R. (1993) A translation approach to portable ontology specifications, Knowledge Acquisition, 5(2), pp. 199—220.

Haklay, M. and Weber, P. (2008) OpenStreetMap: User-Generated Street Maps, IEEE Pervasive Computing, 7(4), pp. 12—18.

Heath T. (2009) The Linked Data community homepage, http://linkeddata.org/, Last data accessed 10.2009.

Hill, J. (2006) Georeferencing, The Geographic Associations of Information, Digital Libraries and Electronic Publishing, MIT Press.

Inspire (2007) Directive 2007/2/EC of the European Parliament and of the Council. Official Journal of the European Union, 2007.

Jakobsson, A. and Zaccheddu P.G. (2009) EuroGeoNames (EGN) - A Prototype Implementation for an INSPIRE Service, GSDI 11 World Conference and the 3rd INSPIRE Conference, Rotterdam 15—19 June 2009.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Geospatial Thinking. 2010, vol. , p. 201-220. ISSN 1863-2246.

Klyne, G. and Carroll, J.J. (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax W3C Recommendation, 2004, http://www.w3.org/TR/rdf-concepts/, Last date accessed 10.2009.

Langegger, A., Wöß, W. and Blöchl, M. (2008) A Semantic Web Middleware for Virtual Data Integration on the Web, The Semantic Web: Research and Applications, pp. 493—507.

López-Pellicer, F.J.; Florczyk, A.J.; Lacasta, J.; Zarazaga-Soria, F.J. and Muro-Medrano, P.R. (2008) Administrative Units, an Ontological Perspective. ER Workshops, 5232, Springer, pp. 354—363.

Maue, P., Schade, S. and Duchesne, P. (2009) OGC Discussion Paper 08-167r1: Semantic annotations in OGC standards, Technical report, Open Geospatial Consortium.

McGuinness, D.L. and van Harmelen F. (2004) OWL Web Ontology Language Overview, W3C Recommendation, 2004, http://www.w3.org/TR/owl-features/. Last date accessed 10.2009.

McKee, L. (2004) The Spatial Web, An Open GIS Consortium (OGC) White Paper, http://portal.opengeospatial.org/files/?artifact_id=3859. Last date accessed 10.2009.

Nebert, D., Whiteside, A. and Vretanos, P. (Eds.) (2007) OpenGIS - Catalogue Services Specification (version: 2.0.2). OGC 07-006r1, Open Geospatial Consortium Inc., http://portal.opengeospatial.org/files/?artifact_id=20555, Last date accessed 10.2009.

Nogueras-Iso, J., Barrera, J., Rodrígez, A.F., Recio, R., Laborda, C. and Zaragaza-Soria, F.J. (2009) Development and deployment of a services catalog in compliance with the INSPIRE metadata implementing rules, SDI Convergence: Research, Emerging Trends, and Critical Assessment. The Netherlands Geodetic Commission (NGC), 2009, pp. 21—34.