

# Relacionando la información geográfica con su componente semántico a través de las IDE

José Ángel Ramos<sup>1</sup>, Luis M. Vilches Blázquez<sup>1</sup>

<sup>1</sup>Ontology Engineering Group, Dpto. Inteligencia Artificial,  
Fac. Informática, Univ. Politécnica de Madrid,  
Av. Montepíncipe s/n, 29660 Boadilla del Monte, Madrid, España  
{jarg, lmvilches}@fi.upm.es

## Resumen

La cantidad de datos enlazados (*Linked Data*) publicados en la Web de los Datos ha experimentado un enorme crecimiento en los últimos años. La información publicada en formato RDF tiene asociada un modelo conceptual (ontología) por el que se rige. Las consultas a los conjuntos de datos publicados se realizan, generalmente, de manera individualizada. En este artículo, mediante la confluencia semántica, se desarrolla un enfoque y un sistema de consultas centralizadas que hace transparente para el usuario la diferente distribución de los datos en la nube que supone *Linked Data*. Los resultados obtenidos a través del sistema desarrollado se visualizan utilizando el WMS de la IDEE. En definitiva, el enfoque propuesto permite la integración de las Infraestructuras de Datos Espaciales y *Linked Data*.

**Palabras clave:** *Linked Data*, confluencia semántica, ontologías, WMS.

## 1 Introducción

La Web de los Datos enlazados (*Web of Linked Data*) supone un nuevo paradigma que pretende explotar la Web como un espacio global de información en el que la navegación se realiza a través de datos estructurados enlazados (*Linked Data*) en lugar de realizarse a través de documentos.

La cantidad de datos enlazados publicados en la Web de los Datos ha experimentado un enorme crecimiento en los últimos años. El fenómeno de *Linked Data* se está extendiendo a diversos sectores, entre los que destacan los medios de comunicación, infraestructuras y logística, el ámbito universitario y científico. En

el contexto de la información geográfica, la publicación de esta información y su integración con otras fuentes de datos puede ser un punto de referencia para el surgimiento de nuevos escenarios que permitan la explotación semántica de las fuentes de información.

Ante la diversidad de información disponible, cabe deducir que la explosión de información y del número de vocabularios (ontologías) asociados puede dar lugar a un problema de accesibilidad, ya que los datos presentes en la nube de *Linked Data* se encuentran en un formato uniforme (RDF) pero las ontologías, con frecuencia, obedecen a diferentes puntos de vista sobre un determinado dominio.

La integración de la información geoespacial proveniente de múltiples fuentes pone de manifiesto diversas problemáticas tanto en el contexto de la representación gráfica como en las clasificaciones de conocimiento asociadas. Estas problemáticas están vinculadas al término confluencia. Este concepto está adquiriendo una doble perspectiva: la visión tradicional, referida al proceso de unificación gráfica o representacional de múltiples fuentes de datos en una visión integrada; y una nueva visión que relaciona la perspectiva semántica con el concepto confluencia - vinculada con las clasificaciones de conocimiento-. Este hecho es resultado de la alta variabilidad de los catálogos (vocabularios) de los distintos sistemas que tratan la información geográfica. Este hecho supone una barrera cuando se trata de mezclar datos de diferentes fuentes de una forma diferente a la tradicional superposición de capas de información. Desde la perspectiva de la confluencia semántica, las terminologías deben integrarse en un modelo que sea capaz de armonizar los conceptos geográficos presentes en los múltiples recursos existentes.

En este trabajo se presenta un enfoque basado en el uso de un modelo (ontología) común. Este modelo se relaciona con los modelos particulares de cada fuente a través de *mappings*<sup>1</sup>. Usando esta infraestructura de modelos conectados se permite que las consultas se puedan realizar en base a un único modelo (el común) pero que los datos obtenidos provengan y mantengan sus modelos (los particulares). Este enfoque genérico se lleva a cabo usando aplicaciones software (descubridor de *mappings* entre ontologías e intérprete de *mappings* para la distribución de consultas) y se experimentará con *Linked Data* geoespacial. El resultado obtenido permite la combinación de dos áreas: Infraestructura de Datos Espaciales (WMS IDEE) y *Linked Data* (GeoLinked Data).

---

<sup>1</sup> Un *mapping* es una especificación formal de una relación entre elementos de diferentes fuentes (a partir de [11]).

Este artículo se organiza de la siguiente manera: en la sección 2 se presentan trabajos afines; en la sección 3 se describe el enfoque propuesto y su particularización al dominio geoespacial como caso de uso; en la sección 4 se detallan el proceso y los resultados de la experimentación llevada a cabo; en la sección 5 se comentan brevemente las conclusiones y los futuros trabajos

## 2 Trabajo relacionado

La información geográfica (IG) se captura, mantiene y actualiza por diversos productores de información geoespacial con distintos niveles de granularidad, calidad y estructura. Lo que se traduce en la aparición de múltiples conjuntos de fuentes de información geoespaciales con gran heterogeneidad en cuanto a sus modelos y datos asociados.

Una aproximación para resolver los problemas causados por la heterogeneidad en el contexto de la confluencia semántica se centra en la utilización de ontologías<sup>2</sup>. En este caso, las ontologías se utilizan con objeto de crear un esquema conceptual (común) que traduzca y contemple las diferentes fuentes de información. Esto ayudará a superar los problemas causados por la heterogeneidad semántica propia de la información geoespacial [6, 7]. En este contexto se ubican diferentes enfoques publicados [2, 4, 14, 15, 16].

Para la integración de información geográfica a través de ontologías, algunos autores proponen un sistema multiontología [3, 5]. En este enfoque cada organización integra sus fuentes de información usando una ontología local y otra ontología integra todas las ontologías locales de las organizaciones del sistema. Como se verá en este artículo, la diferencia de nuestra propuesta en comparación con este trabajo es considerar el tratamiento de *Linked Data*.

En [5] se presenta una propuesta para integrar esquemas de diferentes comunidades de IG, donde cada comunidad utiliza su propia ontología. Esta propuesta está basada en la fusión de ontologías siguiendo relaciones de similitud entre los conceptos de las diferentes ontologías integradas. Finalmente, la ontología fusionada se utiliza para derivar un esquema integrado, utilizado como esquema global en un sistema de base de datos federada [1,12].

---

<sup>2</sup> Una ontología es una especificación formal y explícita de una conceptualización compartida [13].

Sobre la relación entre *mappings* y consultas SPARQL<sup>3</sup>, cabe destacar el trabajo sobre SPARQL++ [9], donde se traducen los *mappings* operacionales a consultas en lenguaje SPARQL ampliado. Sin embargo, estos autores trabajan con *mappings* procedurales y en nuestro caso se trata de *mappings* declarativos<sup>4</sup>. Un trabajo similar, aunque con implicaciones en la modificación de las consultas debido a las relaciones de los *mappings* con los que trata es [8]. Igualmente, queda fuera del ámbito de este trabajo los enfoques que tratan sobre consultas distribuidas en *Linked Data*.

### 3 Consultas geográficas unificadas

La solución que se plantea para formular de manera unificada las consultas a varios repositorios de *Linked Data* consiste en tomar un modelo (ontología) de referencia, relacionarlo mediante *mappings* con los modelos locales de cada SPARQL *Endpoint*<sup>5</sup> y desarrollar un sistema que traduzca, usando los *mappings*, las consultas SPARQL con el vocabulario de la ontología de referencia a consultas SPARQL con los vocabularios adecuados de cada SPARQL *Endpoint*. Este sistema se ha denominado OMI (Intérprete de OEGMappings).

#### Caso de uso

La información del Instituto Geográfico Nacional publicada conforme a los requerimientos de *Linked Data* está disponible en la iniciativa **GeoLinked Data**<sup>6</sup>.

---

<sup>3</sup> <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>

<sup>4</sup> Los *mappings* pueden ser declarativos, cuando expresan la relación entre dos entidades como si de una sentencia se tratara, y procedurales, cuando la relación entre dos entidades está expresada como la transformación de una en otra como si de una fórmula se tratara. Ejemplo: La entidad “milla” y la entidad “metro” tienen un *mapping* entre ellas que es “*sibling*” (son ambas hermanas dentro de una jerarquía de unidades de medida de longitud); este *mapping* es un *mapping* declarativo. Las dos mismas entidades tienen un *mapping* de entre ellas que es “metro = 1609 x milla”; este *mapping* es un *mapping* procedural.

<sup>5</sup> A SPARQL endpoint is a conformant SPARQL protocol service as defined in the SPROT (stands for SPARQL Protocol for RDF) specification. A SPARQL endpoint enables users (human or other) to query a knowledge base via the SPARQL language. Results are typically returned in one or more machine-processable formats. Therefore, a SPARQL endpoint is mostly conceived as a machine-friendly interface towards a knowledge base. Both the formulation of the queries and the human-readable presentation of the results should typically be implemented by the calling software, and not be done manually by human users.

<sup>6</sup> <http://geo.linkeddata.es/>

Esta información tiene asociada una red de ontologías<sup>7</sup>, que es una colección de ontologías unidas a través de una variedad de diferentes relaciones, tales como la modularización, versionado y relaciones de dependencia. Las ontologías que componen esta red son: la ontología geopolítica de la FAO<sup>8</sup>, una ontología de fenómenos hidrográficos (laguna, surgencia, playa, etc.) denominada hydrOntology<sup>9</sup>, una ontología propia de transportes (pista, faro, aeropuerto, etc.), dos ontologías de modelado y representación de la información geográfica (GML<sup>10</sup> y WGS84Vocabulary<sup>11</sup>), una ontología sobre información estadística (SCOVO)<sup>12</sup> y la Ontología de Tiempo<sup>13</sup> del W3C.

Para llevar a cabo la conflación semántica de modelos se ha elegido **PhenomenOntology**<sup>14</sup>. Esta ontología ha sido tomada como estándar de facto por parte del IGN para la información geográfica española. Por ejemplo, son conceptos “Edificio religioso”, “Surgencia” y “Autopista”; el concepto “Autopista” tiene como atributo “acceso”, cuyos valores enumerados son “de peaje” y “libre”.

Para obtener las correspondencias entre PhenomenOntology y las ontologías de GeoLinked Data, se ha utilizado la herramienta **OEGMappingDiscoverer** [10] que posee comparadores genéricos y *ad hoc* para la información geográfica.

Las instancias con información geográfica generadas de estas ontologías se pueden consultar en los repositorios públicos, o SPARQL *Endpoints*, de GeoLinked Data<sup>15</sup>.

En la Figura 1 se muestra una visión de la solución propuesta particularizada con los componentes específicos de información geoespacial antes citados.

Con toda esta información, el intérprete **OMI** ofrece respuesta a las consultas de información geográfica en español que se realicen con el vocabulario de la ontología de referencia (PhenomenOntology). El proceso que sigue el intérprete es el siguiente:

- A partir de una consulta, OMI accederá a los alineamientos buscando los *mappings* relativos a los conceptos que aparecen en la consulta.

---

<sup>7</sup> <http://geo.linkeddata.es/web/guest/modelos>

<sup>8</sup> <http://www.fao.org/countryprofiles/geoinfo.asp?lang=es>

<sup>9</sup> <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/107-hydrontology>

<sup>10</sup> <http://loki.cae.drexel.edu/~wbs/ontology/2004/09/ogc-gml.owl>

<sup>11</sup> [http://www.w3.org/2003/01/geo/wgs84\\_pos](http://www.w3.org/2003/01/geo/wgs84_pos)

<sup>12</sup> <http://purl.org/NET/scovo>

<sup>13</sup> <http://www.w3.org/TR/owl-time/>

<sup>14</sup> <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/152-phenomenontology>

<sup>15</sup> <http://geo.linkeddata.es/sparql>

- A continuación, traducirá la consulta al vocabulario adecuado de cada SPARQL *Endpoint* sustituyendo los nombres de los conceptos por los que indican los *mappings*.
- Finalmente recopilará (atendiendo si procede a las relaciones de los *mappings* interpretados) las respuestas ofrecidas por los SPARQL *Endpoints* (URIs y coordenadas de los fenómenos) y las devolverá al usuario como una respuesta única a la consulta original.

El resultado obtenido se visualiza mediante marcadores superpuesto sobre el WMS IDEE-Base.

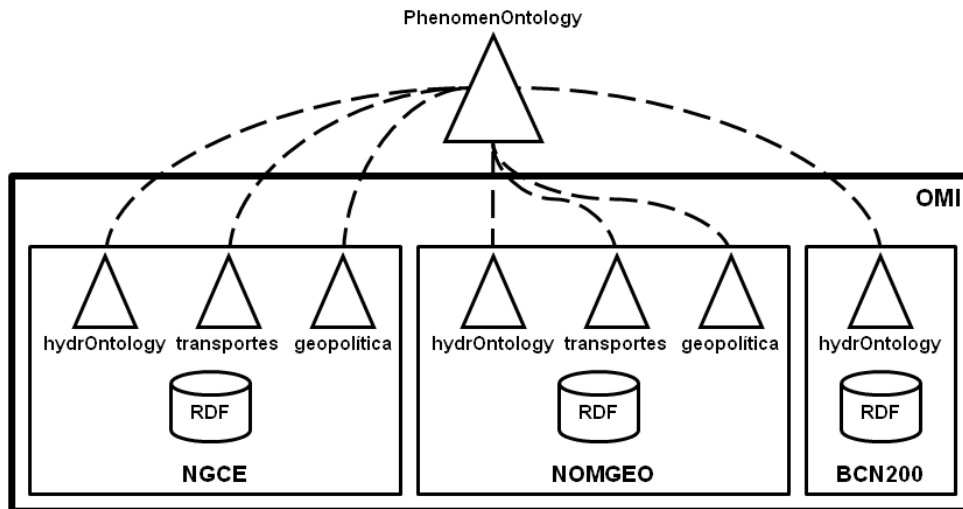


Figura 1. Esquema de la solución propuesta

#### 4 Caso de uso real

Para la experimentación se cuenta con las mencionadas ontologías (PhenomenOntology y la red de ontologías GeoLinked Data) y 3 SPARQL *Endpoints* en los que se encuentran instancias de las siguientes ontologías:

- NGCE: hydrOntology, transportes, geopolítica
- NOMGEO: hydrOntology, transportes, geopolítica
- BCN200: hydrOntology

Se han generado con OEGMappingDiscoverer los alineamientos entre PhenomenOntology (335 conceptos) e hydrOntology (331 conceptos), transportes (18 conceptos) y geopolítica (6 conceptos). Estos alineamientos tienen 90, 40 y 7 *mappings*, respectivamente.

Para la visualización se ha usado el servicio WMS IDEE-Base (relieve e hidrografía) de la IDEE y una capa de marcadores donde se representan los resultados (Linked Data) obtenidos.

Sobre esta base de información, se han realizado experimentos de consultas usando el intérprete OMI. A modo de ejemplo, se van a mostrar a continuación las ejecuciones detalladas (considerando la limitación de espacio) de tres consultas.

### **Ejemplo 1: Laguna**

El concepto seleccionado de PhenomenOntology es “Laguna”.

OMI a partir de la URI del concepto “Laguna” busca en los ficheros de *mappings* y obtiene 1 *mapping* del alineamiento con hydrOntology.

<http://www.semanticweb.org/ontologies/2010/04/Phenom4.0.owl#Laguna##LABEL##Laguna> → <id, Laguna, Laguna, =, 1>.

Como ha localizado *mappings* en el alineamiento con hydrOntology, OMI genera consultas para cada SPARQL *Endpoint* donde tiene instancias esta ontología (NGCE, NOMGEO y BCN200), con cada uno de los *mappings*. Se genera una consulta para cada SPARQL *Endpoint* (sólo hay un *mapping*).

En todos los SPARQL *Endpoints* encuentra instancias. Se incluye un extracto representativo de las URIs de la respuesta a continuación:

URI: <a href="http://geo.linkeddata.es/NGCE/resource/Laguna/Tae%C3%B1a%20Laguna%20de">http://geo.linkeddata.es/NGCE/resource/Laguna/Tae%C3%B1a%20Laguna%20de</a>
URI: <a href="http://geo.linkeddata.es/NGCE/resource/Laguna/Tollos%20Laguna%20de%20los">http://geo.linkeddata.es/NGCE/resource/Laguna/Tollos%20Laguna%20de%20los</a>
[...]
URI: <a href="http://geo.linkeddata.es/NOMGEO/resource/Laguna/Hilejas%20Laguna%20de%20las">http://geo.linkeddata.es/NOMGEO/resource/Laguna/Hilejas%20Laguna%20de%20las</a>
URI: <a href="http://geo.linkeddata.es/NOMGEO/resource/Laguna/Tejo%20Laguna%20del">http://geo.linkeddata.es/NOMGEO/resource/Laguna/Tejo%20Laguna%20del</a>
[...]
URI: <a href="http://geo.linkeddata.es/BCN200/resource/Laguna/Laguna%20de%20Paniagua">http://geo.linkeddata.es/BCN200/resource/Laguna/Laguna%20de%20Paniagua</a>
URI: <a href="http://geo.linkeddata.es/BCN200/resource/Laguna/Laguna%20Laguna%20">http://geo.linkeddata.es/BCN200/resource/Laguna/Laguna%20Laguna%20</a>
[...]

La respuesta se compone de 1.806 URIs (38 provenientes de NGCE, 1.416 provenientes de NOMGEO y 406 provenientes de BCN200) y sus correspondientes coordenadas. En la Figura 2 se muestran los resultados obtenidos sobre el WMS IDEE base.

### **Ejemplo 2: Pista**

El concepto seleccionado de PhenomenOntology es “Pista”.

OMI a partir de la URI del concepto “Pista” busca en los ficheros de *mappings* y obtiene 1 *mapping* del alineamiento con transportes.

<http://www.semanticweb.org/ontologies/2010/04/Phenom4.0.owl#Pista##LABEL##Pista> → <id, Pista, Pista, ≡, 1>.

Como ha localizado *mappings* en el alineamiento con transportes, OMI genera consultas para cada *SPARQL Endpoint* donde tiene instancias transportes (NGCE y NOMGEO). Se genera una consulta para cada *SPARQL Endpoint* (sólo hay un *mapping*).

Sólo hay instancias de este concepto de la ontología transportes en NOMGEO. Se incluye un extracto representativo de las URIs de la respuesta a continuación:

```
URI: http://geo.linkeddata.es/NOMGEO/resource/Pista/Arenas%20Negras%2C%20Pista%20de
URI: http://geo.linkeddata.es/NOMGEO/resource/Pista/Oso%2C%20Pista%20del
URI: http://geo.linkeddata.es/NOMGEO/resource/Pista/Derrabado%2C%20Pista%20del
[...]
```

La respuesta final está compuesta por 24 URIs, todas provenientes NOMGEO y sus correspondientes coordenadas. En la Figura 3 se muestran los datos obtenidos.

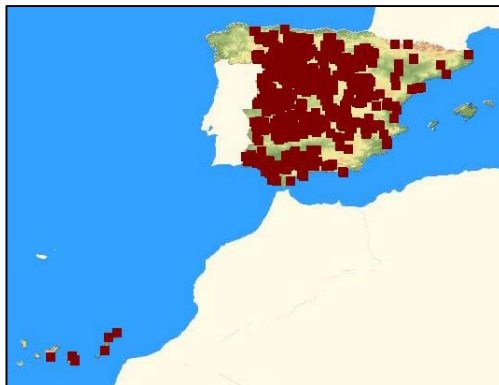


Figura 2. Mapa de “Laguna”



Figura 3. Mapa de “Pista”

### Ejemplo 3: Surgencia

El concepto seleccionado de PhenomenOntology es “Surgencia”.

OMI a partir de la URI del concepto “Surgencia” busca en los ficheros de *mappings* y obtiene 3 *mappings* del alineamiento con hydrOntology.

<http://www.semanticweb.org/ontologies/2010/04/Phenom4.0.owl#Surgencia##LABEL##Surgencia>  
→ <id, Surgencia, Surgencia, ≡, 1>, <id, Surgencia natural, Surgencia, <, 0,7> y <id, Surgencia artificial, Surgencia, <, 0,7>.

Como ha localizado *mappings* en el alineamiento con hydrOntology, OMI genera consultas para cada *SPARQL Endpoint* donde tiene instancias hydrOntology



(NGCE, NOMGEO y BCN200), con cada uno de los *mappings*. Se generan tres consultas para cada *SPARQL Endpoint*.

Sólo encuentra instancias de “Surgencia” en el *SPARQL Endpoint* NOMGEO, no teniendo instancias “Surgencia natural” ni “Surgencia artificial” en ningún *SPARQL Endpoint*. Se incluye un extracto representativo de las URIs de la respuesta a continuación:

```
URI:  
http://geo.linkeddata.es/NOMGEO/resource/Surgencia/Masc%C3%BAn%2C%20Surgencias%20del  
URI: http://geo.linkeddata.es/NOMGEO/resource/Surgencia/R%C3%ADo%20Leza%20o%20de%20  
la%20Fuente%20del%20Restauro%2C%20Surgencia%20del%20%7C  
[...]
```

La respuesta final está compuesta por 9 URIs y sus correspondientes coordenadas. En la Figura 4 se muestran los datos obtenidos y en la Figura 5 se puede ver ampliada una zona con marcadores donde se ve la capa de hidrografía.



Figura 4. Mapa de “Surgencia”

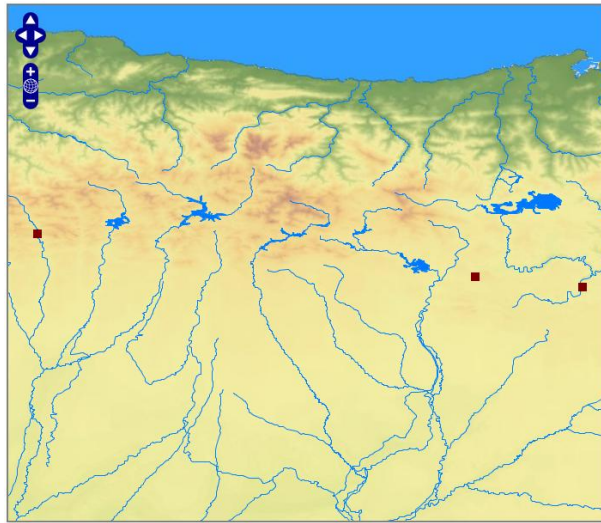


Figura 5. Zoom sobre el mapa de “Surgencia”

## 5 Conclusiones y trabajos futuros

La solución planteada para unificar las consultas haciendo uso del enfoque de conflación semántica se ha llevado a cabo mediante una aplicación, el intérprete de *mappings* OMI. Este sistema realiza la interpretación y consultas de forma dinámica, atendiendo a los ficheros de alineamientos fácilmente actualizables (con

la herramienta de descubrimiento automático) y a las asociaciones de los *SPARQL Endpoints* y sus modelos (ontologías). La experimentación, realizada con un caso de uso real en el dominio geoespacial, ha demostrado la utilidad del planteamiento. Igualmente, debido a que la experimentación se ha limitado a una consulta SPARQL determinada, condicionada por el caso de uso real, es trabajo futuro la ampliación del sistema para admitir todo tipo de consultas del estándar y lograr que sean tratadas adecuadamente por el intérprete OMI.

El resultado de las consultas que proporciona OMI ha sido integrado con éxito utilizando el servicio WMS IDEE Base. Queda como trabajo futuro una mayor usabilidad de este resultado, dotando a los marcadores de enlaces directos a los datos y la integración del sistema desarrollado en el sitio web de la iniciativa GeoLinked Data.

**Agradecimientos.** Este trabajo ha sido financiado por el proyecto España Virtual (CENIT-2008-1030), auspiciado por el Centro Nacional de Información Geográfica y CDTI, enmarcado en el contexto del programa Ingenio 2010.

## Referencias

- [1] Blanco JM, Illarramendi A, Goñi A (1994) Building a federated database system: An approach using a knowledge based system. *Int. Journal on Intelligent and Cooperative Information Systems*, vol. 3, no. 4, pp. 415-455.
- [2] Fonseca F, Davis C, Câmara G (2003) Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. *GeoInformatica*, Volume 7(4): 355-378.
- [3] Fonseca FT, Egenhofer MJ, Davis CA, Câmara G (2002) Semantic Granularity in Ontology-Driven Geographic Information Systems. *Annals of Mathematics y Artificial Intelligence*. Vol. 36, 2002. Issue: 1-2. Pp: 121-151.
- [4] Gómez-Pérez A, Ramos JA, Rodríguez-Pascual A, Vilches-Blázquez LM (2008) The IGN-E Case: Integrating through a hidden ontology. In *Headway in Spatial Data Handling. SDH'08. LNGC*. Pp. 417-435. Montpellier, France.
- [5] Hakimpour F (2003) "Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata". Ph.D. thesis, Zurich University. Switzerland.
- [6] Harvey F, Kuhn W, Pundt H, Bishr Y, Riedemann C (1999) Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science*, vol. 33 (2): 213-232.

- [7] Kuhn W (1995) *Semantics of Geographic Information: GeoInfo 7*. Technical University of Vienna, Austria, Department of Geoinformation.
- [8] Makris K, Gioldasis N, Bikakis N, Christodoulakis S (2010) *Ontology Mapping and SPARQL Rewriting for Querying Federated RDF Data Sources*. International conference on On the move to meaningful internet systems 2010. Vol. 2, pp: 1108-1117.
- [9] Polleres A, Scharffe F, Schindlauer R (2007) *SPARQL++ for mapping between RDF vocabularies*. In 6th Int. Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2007), Vilamoura, 2007.
- [10] Ramos JA, Fernández-López M, Gómez-Pérez A (2009) *Algoritmo de agregación de mappings basado en reglas de selección*. Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA'2009). Sevilla, 10-12 noviembre de 2009. Pp: 137-146.
- [11] Ramos JA, Gómez-Pérez A (2008) *Mappings for the Semantic Web*. IEEE Intelligent Systems - Trends & Controversies. Noviembre-diciembre 2008. Pp: 79-81.
- [12] Sheth AP, Larson JA (1990) *Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases*. ACM Computing Surveys, 22(3).
- [13] Studer R, Benjamins VR, Fensel D (1998) *Knowledge Engineering: Principles and Methods*. IEEE Transactions on Data and Knowledge Engineering 25 (1-2). Pp: 161–197.
- [14] Uitermark HT (2001) *Ontology-Based Geographic Data Set Integration*. Ph.D. thesis. University of Twente, The Netherlands.
- [15] Uitermark HT, Van Oosterom PJM, Mars NJI, Molenaar M (2005) *Ontology - based integration of topographic data sets*. International Journal of Applied Earth Observation and Geoinformation;7(2): pp. 97-106.
- [16] Visser U, Stuckenschmidt H, Schuster G, Vögele T (2002) *Ontologies for geographic information processing*. Computers & Geosciences, Vol. 28, pp. 103-117.