

Semántica Espacial
Modelo conceptual de Información Geográfica para la
IDE – Uruguay

Diana Comesaña¹

Resumen

La mayoría de la información geográfica de actualidad se halla dispersa en la web. Abarca disciplinas que van desde la Geociencias hasta las Ciencias Sociales. El problema surge al considerar la variedad y dispersión de datos existente, formatos y denominación de objetos y acontecimientos, lo que dificulta la interoperabilidad y recuperación de los datos. Se propone una solución partiendo desde las Ciencias de la Información y tomando herramientas y conceptos de la Informática para desarrollar un modelo conceptual que aporte semántica a este sector de la información. Se siguen los pasos del método de desarrollo de ontologías Methontology, para construir un modelo en el lenguaje OWL-DL. El resultado es un conjunto de once ontologías que puedan actuar en una red, para conformar un meta-modelo, que permita una mejor interoperabilidad de los datos y recuperación de los mismos, desarrollando un producto aplicable en otros dominios de datos reales.

Palabras clave: IDE-Uruguay, Ontologías, Objetos geográficos, Información geográfica, Semántica espacial.

Space semantics
Conceptual model of Geographic Information for the SDI-Uruguay

Abstract

Most current geographic information is scattered across the web. This information includes disciplines such Geosciences to Social Sciences. The problem arises when considering the variety and dispersion of data, the diversity of formats and how to call these objects and events, making it difficult to interoperability and data recovery. The solution proposed is based on the Information and Communication Sciences and taking tools and concepts from Computing Science

¹ Departamento de Tratamiento y Transferencia de la Información, Instituto de Información y Comunicación. Facultad de Información y Comunicación, Universidad de la República. Emilio Frugoni 1427. Montevideo-Uruguay. Correo electrónico: diana.comesana@fic.edu.uy

to develop a conceptual model that provides semantic information to this sector. The Metontology method of development of ontologies, to build a model in OWL-DL. The result is a set of eleven ontologies that can operate on a network, to form a meta-model which enables better interoperability of data and their recovery, developing a product applicable in other domains of real data.

Keywords: SDI-Uruguay; Ontologies; Geographic Objects; Geographic Information; Space Semantic.

1. Introducción

Cuando hablamos de datos geográficos nos referimos a hechos y objetos naturales y de origen antrópico que suceden o existen en la superficie de la Tierra. Como define el Centro de Desarrollo e Investigación en Información Geográfica del Instituto Geográfico Agustín Codazzi de Colombia: "Los **datos geográficos** son entidades espacio-temporales que cuantifican la distribución, el estado y los vínculos de los distintos fenómenos u objetos naturales y sociales. Un **dato** se caracteriza por tener: Posición absoluta: sobre un sistema de coordenadas (x, y, z)." (IGAC, s. d.)

Con el desarrollo de la Geomática, es decir de la Informática aplicada a la Geografía, la información geo-espacial de por sí estratégica, se ha convertido en un insumo imprescindible para la toma de decisiones. En su mayoría se halla dispersa en la web y es una información heterogénea, que abarca disciplinas tan diversas, como la Geociencia, la Geografía, las Ciencias Naturales y las Ciencias Sociales.

La multiplicidad de actores en la obtención de los datos geográficos, no solamente lleva a la dispersión de los mismos. Los datos se construyen bajo diferentes estándares y reciben diferentes denominaciones y alcances. Se produce duplicidad de esfuerzos y dificultades en cuanto a la interoperabilidad de los datos y al acceso y evaluación de la calidad de los mismos.

En respuesta, y como evolución de los Sistemas de Información Geográfica, surgen los proyectos de Infraestructura de Datos Espaciales (IDE), tanto a nivel nacional, como regional. Estos proyectos tienden a solucionar parte de los problemas, pero no han resultado suficientes. Deben complementarse.

Las IDE surgen en la década del '90 y su primer modelo se basó en el producto, en lo que se denomina IDE de 1ª Generación. Alrededor del año 2000, el desarrollo se orienta hacia los procesos y se comienza a hablar de IDEs orientadas a usuarios o a demanda (Rajabifard, 2006).

Nuestro país ha comenzado el desarrollo de su Infraestructura de Datos Espaciales y la normalización de la producción y descripción de la información geográfica. Diferentes documentos nacionales, presentados en el I Congreso Uruguayo de Infraestructura de Datos Espaciales Uruguayo, y posteriores, mencionan la necesidad de conformar una IDE Semántica mediante el uso de

Metadatos Semánticos y desarrollo de Ontologías.

Enrique Latorres (2010) expresa:

El Sistema utilizará la información espacial existente como vínculo para el uso de conocimiento por los diferentes participantes del proyecto. El sistema tendrá tres niveles de servicios y múltiples áreas de conocimiento. Dispondrá de sistemas de consultas avanzadas dentro de las bases de conocimiento basadas en ontologías, disponibles para aquellos expertos de las organizaciones que tanto agregan contenido como lo consumen.

En el mismo documento, se establece, para las Áreas del Conocimiento:

El conocimiento se registrará por parte de Organismos y Entidades en forma Vertical. Se elaboraran Ontologías separadas que cada organización, se asegurará la coherencia y consistencia vertical, pero sus componentes de conocimiento podrán ser reutilizados por los usuarios de otros.

Se hablaba ya de la necesidad de proporcionarle semántica, pero no se presentaron posturas acerca de cuál es el modelo conceptual que más se ajuste a nuestra realidad.

2. La hipótesis

Nos propusimos demostrar la existencia de un modelo conceptual para este tipo de información, mediante la construcción de un prototipo funcional, con la mirada puesta en las Ciencias de la Información, pero tomando herramientas y conceptos pertenecientes a la Informática, de suerte que partiendo de un Análisis de Dominio, en el concepto de Birger Hjørland, y un Análisis Terminológico, estudiamos diferentes modelos de Organización y Representación del conocimiento, hasta alcanzar el desarrollo del prototipo que nos resultó más adecuado a nuestra realidad, tal cual se aprecia en la Figura 1.

3. Metodología empleada

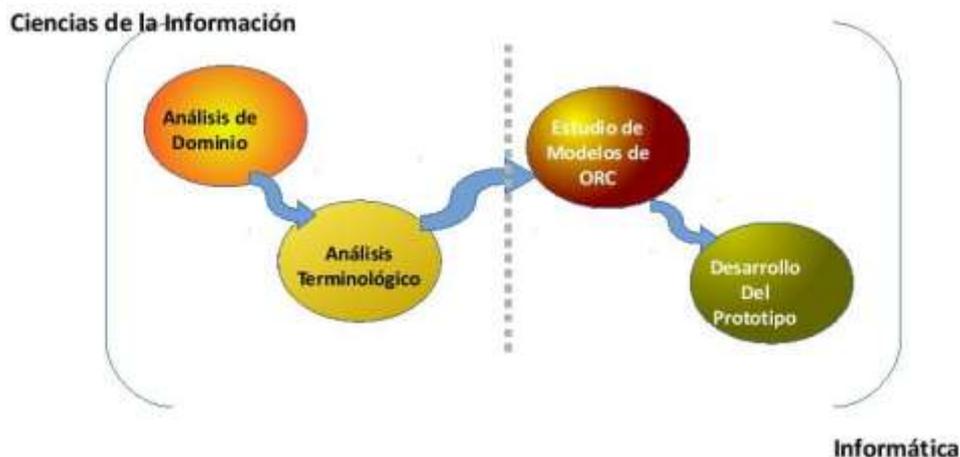


Figura 1: Diagrama de flujo del trabajo realizado

En este diagrama vemos los pasos realizados en el trabajo: partiendo del paradigma de Análisis del Dominio de Hjørland, estructurado como una combinación de los enfoques por él señalados, nos permitió alcanzar el conjunto de términos que representan los objetos y hechos geográficos para nuestra Infraestructura de Datos Espaciales. Para este análisis se realizaron entrevistas pautadas a los actores involucrados, se realizó un análisis documental de los documentos textuales y cartográficos por ellos producidos, de los geo-portales en funcionamiento y los catálogos de metadatos disponibles.

El Análisis Terminológico posibilitó clarificar definiciones que en algunos casos eran ambiguas o inexistentes y establecer relaciones entre los términos, para, por comparación con los modelos existentes, hallar el más apropiado.

Se presentaron problemas de índole cultural al enfrentar los términos correspondientes al dominio: por un lado la influencia del portuñol en la denominación del objeto geográfico tajamar, azude en nuestra frontera con Brasil, y por otro la definición ambigua de términos como arroyo-cañada-río, ciudad-pueblo-villa, definiciones que debieron construirse por el Método de la Grilla del Prof. Mario Barité, a partir de las leyes de nuestro país que se aplican a ellos (Ley de Aguas y Ley de Centros Poblados) y de los manuales y glosarios que manejan los integrantes de la IDE-Uruguay.

El desarrollo del prototipo constituyó un abordaje marcadamente informático. Al enfrentar los problemas de interoperabilidad, el primer y gran desafío fue comprender que nuestra semántica, con un significado lingüístico en las Ciencias de la Información, y nuestros Términos, equivalentes a una palabra o conjunto de palabras que expresan un concepto, se transformaban, el término en una "entidad", definida por una serie de atributos, y la semántica en el hecho de que esta entidad fuera comprensible para las máquinas, permitiendo llegar a la automatización de los procesos, como muestra la Figura 2.

Conceptos Término y Semántica



Figura 2: Transformación de los conceptos de Término y Semántica

Según Vizcaya (1997), la organización de la información consiste en “la estructura formal que, como resultado de las variables usuario, temática, tipología de fuentes, tecnología y presupuesto, dispone, define y coordina las diferentes etapas del ciclo de vida de la información, con el propósito de cumplir con los objetivos de un sistema.” Las formas más tradicionales de Organización y Representación del Conocimiento (ORC), aplicadas a la información espacial (catálogos y tesauros), si bien se muestran útiles a la hora de recuperar información, no son eficaces para solucionar los problemas derivados de la interoperabilidad de los datos geo-espaciales.

Analizamos tesauros y su normativa SKOS (Simple Knowledge Organization System), modelos RDF (Resource Description Framework) y RDFS (RDF Schema o Esquema RDF que es una extensión semántica de RDF) y Ontologías. Los Tesauros están destinados a ser empleados por los seres humanos. Presentan relaciones jerárquicas, asociativas y equivalentes. Pueden decir que dos términos están relacionados, pero no cuál es esta relación, como vemos para el término “rio” en el microtesauro específico incluido en el Macrotesauro en línea de la UNESCO.

Término Río, Macrotesauro de la UNESCO

MT Hidrología (Término perteneciente al Microtesauro específico)
UP Arroyo
TG Aguas superficiales
 TG2 Recursos Hídricos
TE Cuenca fluvial
TE Desagüe de Ríos
TR Canal
TR Contaminación del agua
TR Delta

Figura 3: Término Río, en el Macrotesauro en línea de la UNESCO

Cuando un tesauro está construido según la normativa SKOS, puede combinarse con diferentes editores de geo-portales para actuar desde sus catálogos de metadatos. Si se buscara simplemente la mejor recuperación de la información, ésta sería una buena opción.

La idea de este trabajo es llegar a la automatización, por lo que se estudiaron modelos que provienen de la informática: RDF y Ontologías.

Los modelos RDF se basan en la teoría de grafos, corresponden a una estructura simple: sujeto-predicado-objeto.

Grafo RDF



Figura 4: Grafo RDF para las instancias “Río Uruguay”, “Río Arapey”, “Arroyo Arerungua”

En la Figura 4 vemos un grafo que representa a los ríos Uruguay, Arapey y Arroyo Arerungua. Vemos que podemos establecer la relación que los une en cuanto a que uno desemboca en el otro y su relación inversa, uno es curso principal del otro. Nos permiten determinar las relaciones que necesitamos entre dos objetos, pero no si dos clases de objetos son disjuntas, por lo que si se indica una propiedad para una, se transmitirá a las demás. En el caso de la figura, el Arroyo Arerungua desemboca en el Río Arapey, que desemboca en el Río Uruguay. ¿Es lícito entonces decir que el Arroyo Arerungua desemboca en el Río Uruguay?

Tampoco es posible establecer la cardinalidad, no es posible expresar que un curso de agua desemboca en un único lugar, pero que puede recibir “n” afluentes.

El siguiente modelo estudiado, fueron las ontologías que se basan en la lógica descriptiva. Por estar basadas en lógica constituyen un modelo formal y

están destinadas al procesamiento automático de los datos. Una ontología, al estar basada en la lógica permite alcanzar definiciones de términos que son comprensibles para las máquinas. Nos proporciona la estructura jerárquica de los términos como los tesauros, nos permite establecer relaciones como en RDF, pero también indicar la cardinalidad y realizar inferencias con motores de razonamiento.

Tomaremos, para ontología, la definición de Gruber, que nos dice que una ontología es “una especificación explícita y formal sobre una conceptualización compartida” (Gruber, 1993), es decir está conformada por términos, definidos por consenso por los integrantes de cierto colectivo, sus definiciones y axiomas que los relacionan con otros términos, organizados en una taxonomía y que permiten realizar inferencias. Constituyen masas de conocimientos que se organizan en dos conjuntos los T-box, que contienen sentencias que describen conceptos jerárquicos y sus relaciones, y los A-box, donde hallamos relaciones entre individuos y conceptos.

Una ventaja de los modelos constituidos por ontologías, es que incorporan tanto el lenguaje natural como el controlado.

Por ejemplo, consideremos el término “tajamar”, que emplean las regiones centro, sur y litoral de nuestro país. En nuestra frontera con Brasil se emplea la denominación “azude”. Ambos corresponden al concepto “embalse”. Si el modelo que se adopta corresponde a la estructura de un tesoro, deberíamos indicar un término preferencial al que remitirnos. En las ontologías es suficiente indicar que estos términos expresan el mismo concepto, permitiendo que el equipo realice las inferencias correspondientes.

4. Modelos de ORC

Modelo ORC	Destino	Relaciones	Base	Carencias
Tesoro	Uso humano	Jerárquica Asociativa Equivalente	Unidades léxicas	No permite automatización de los procesos
Normativa SKOS	Normalización de tesauros, tiende a la automatización	Jerárquica Asociativa Equivalente	Unidades léxicas	Algunos editores de geo- portales, como el Geonetwork, pueden incorporarlo
RDF y RDFS	Automatización	Pueden establecerse el tipo de relaciones que se desee entre	Teoría de grafos	No puede establecerse que dos clases sean disjuntas ni cardinalidad

Modelo ORC	Destino	Relaciones	Base	Carencias
Ontología		términos		
	Automatización	Jerárquicas Las que sean necesarias entre términos, descripciones lógicas entre los mismos, restricciones y cardinalidad	Lógica descriptiva	Según el lenguaje OWL empleado puede existir límites al establecer restricciones.

Tabla 1: Comparación de los Diferentes modelos de ORC

Información geodésica, información catastral, información meteorológica, información turística, etc. la variedad de disciplinas, formatos y soportes en que se encuentra dispersa la información geográfica, dificulta la construcción y mantenimiento de una macro-ontología que los englobe a todos.

La tabla 1 muestra las características básicas de cada uno de los modelos estudiados, con sus ventajas y desventajas. El estudio de estos modelos lleva a que la selección recaiga en un conjunto de ontologías, que cubran todo el dominio y conformen una red que aporte semántica a la información geográfica contenida en la Infraestructura de Datos Espaciales uruguaya.

El lenguaje seleccionado, OWL-DL (Ontology Web Language), permite un limitado grado de metamodelado, por lo que al formar esta red deben desarrollarse “puentes” que las conecten entre sí, apoyándonos en las características especiales del meta-modelo indicado por Edelweis Rohrer en su trabajo con la descripción lógica ALCQM (Rohrer et al, 2014).

Modelo de ORC para la IDE-Uruguay



Figura 5: Conjunto de ontologías que conforman la red de objetos geográficos

Este gráfico nos muestra un primer nivel del meta-modelado, donde las ontologías **Objetos_Culturales**, **Hidrografía**, **Vegetación**, **Información_Aeronáutica**, **Información_Catastral**, **Fisiografía**, **Hipsografía**, **Información_General**, **Demarcación y Uso_especiales**, constituyen instancias de la Meta-clase **Objetos_Geográficos**.

5. Desarrollo del prototipo

Para el desarrollo de cada una de las ontologías que constituyen el meta-modelo, de este prototipo se siguen los pasos del método de desarrollo de ontologías Methontology. Se escoge el lenguaje OWL-DL (Web Ontology Language) por su grado de madurez al momento de la implementación y el editor de ontologías Protégé 4.3 por su practicidad.

Método Methontology



Figura 6: Los siete pasos del método de desarrollo de ontologías Methontology

Methontology es un método que se desarrolla en siete pasos, y estos siete pasos, representados en la Figura 6, nos muestran cuán cercanos son al estudio de la Terminología y la Informática, al apreciar que incluye construcción de glosarios, taxonomías de conceptos y diccionarios. La cercanía del método a ambas disciplinas determinó su uso.

Ontografo de la Ontología Objetos_Geográficos

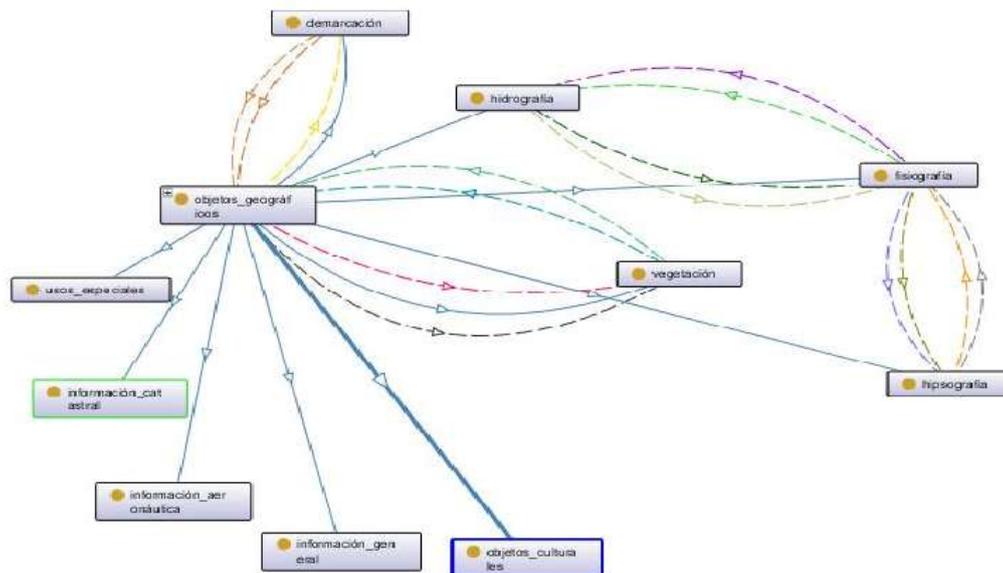


Figura 7: Gráfico del primer nivel del meta-modelo y algunas de sus relaciones

En esta figura vemos como la hidrografía, está determinada por la forma del terreno (fisiografía), que a su vez es modelado por ella y se asocia a la hipsografía. La demarcación de límites administrativos o de otra índole está asociada a los objetos geográficos, o como la vegetación se asocia a los mismos.

6. La sintaxis empleada

El análisis terminológico nos llevó a alcanzar definiciones para aquellos términos que no la tenían clara, por ejemplo los términos “cañada” y “río”:

Río: *Corriente de agua continua, más o menos caudalosa, que va desembocar en otro río, lago o mar y tiene la condición de ser Agua Fiscal, por lo que es navegable o flutable, artificial o naturalmente, y que no desaparece en el terreno en ninguna estación.*

Cañada: *Corriente de agua de poco caudal, que puede desaparecer en el terreno en tiempos de sequía, por lo que está sujeta al régimen de lluvias, y que no forma parte del sistema de Aguas Fiscales, al no ser flutable ni navegable.*

La figura 8, nos muestra el tipo de sintáxis que deberemos emplear para construir esta definición en forma lógica.

OWL constructor	DL	Protégé (Manchester Syntax)	Explicación
SubClassOf	$C \subseteq D$	SubClassOf	C es sub-clase de D
intersectionOf	$C \sqcap D$	C and D	Ambas clases C y D
unionOf	$C \sqcup D$	C or D	Clase C o D
complementOf	$\neg C$	Not C	No es C
someValuesFrom	$\exists R.C$	R some C	R son algunos valores de C
allValuesFrom	$\forall R.C$	R only C	R son todos los valores de C
cardinality	$= n.R$	R exactly n	R es exactamente n
null	\perp		Valor nulo

Figura 8: Sintaxis a emplear en el desarrollo de la ontología

Con ésta sintaxis, la definición anterior se transforma en:

curso_de_agua \subseteq Hidrografía [“curso_de_agua” es sub-clase de “Hidrografía”]

Cañada \subseteq curso_de_agua [“cañada” es sub-clase de “curso_de_agua”]

Río \subseteq curso_de_agua [“río” es sub-clase de “curso_de_agua”]

río \sqcap cañada = \perp [la intersección de las clases río y cañada es nula]

\exists (Desagua Curso_de_Agua. Curso_de_Agua U Desagua Curso_de_Agua.
Masa_de_agua [un curso siempre desagua en otro curso o en una masa de agua]

\exists (Desagua Río. Río U río Desagua. río U río Desagua. Masa_de_agua) [un río siempre es afluente de (desagua) en otro río o en una masa de agua]

\exists (cañada Desagua. Río U cañada Desagua. arroyo U cañada Desagua.
Masa_de_agua U cañada Desagua.cañada) [una cañada, por otra parte desemboca en otra cañada, un arroyo, un río o un espejo de agua]

\forall río. Agua_fiscal [un río siempre es un agua fiscal]

\neg cañada. Agua_fiscal [una cañada nunca es agua_fiscal]

Y como están sujetas al régimen de lluvias:

\forall cañada. Tiene_regimen_de_lluvias \forall U cañada. desaparece en el terreno
[una cañada está asociada al régimen de lluvias]

7. Consultas al modelo (Consultas SPARQL)

Nuestra intención era demostrar la existencia del modelo mediante la construcción de un prototipo funcional. Para ello veamos como responde ante preguntas que surgieron en el análisis terminológico.

Las definiciones de “río” y “cañada”, lagos y lagunas, están íntimamente relacionadas con la Ley de Aguas, que determina cuales son las “aguas fiscales” y esta característica está dada por la navegabilidad de las aguas.

Veamos su respuesta al preguntarle al modelo ¿Que Aguas Superficiales tienen la condición de ser Aguas Fiscales?

```
SELECT ?Aguas_superficiales ?Agua_fiscal WHERE { ?Aguas_superficiales  
<http://www.semanticweb.org/diana/ontologies/2013/8/untitled-ontology-9#es> ?  
Agua_fiscal
```

La consulta se realiza a la ontología “Hidrografía”, y como respuesta nos lista cuales son las aguas superficiales que son navegables o flotables, tal como lo indica la Ley:

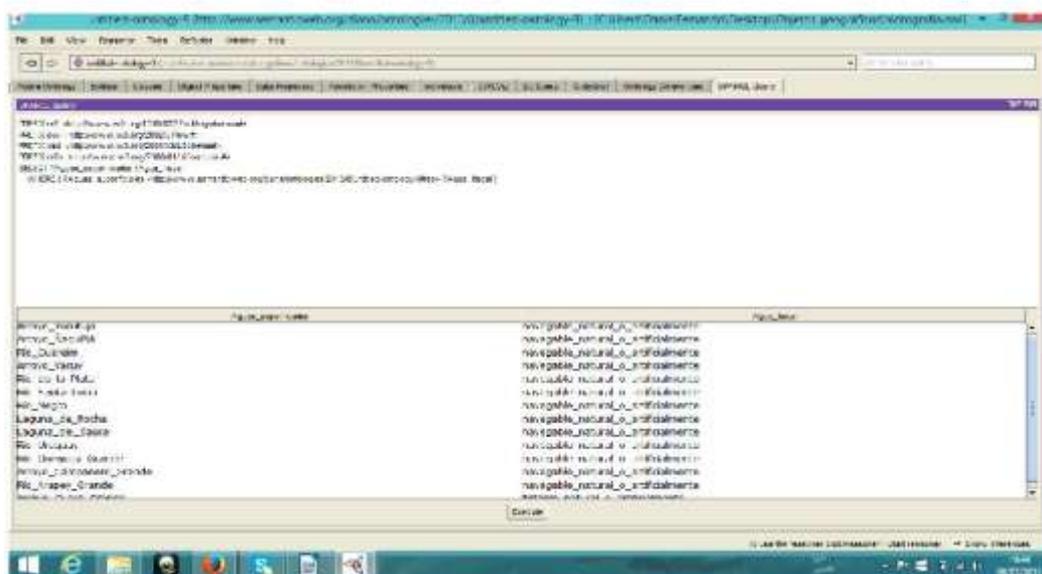


Figura 9: consulta SPARQL al modelo

8. Conclusiones

El modelo propuesto logra una representación del conocimiento que mejora la interoperabilidad y recuperación de los datos, contribuyendo a situarnos mejor en el panorama mundial del desarrollo de la información geográfica, al aportar semántica a la información contenida en nuestra Infraestructura de Datos Espaciales.

Por estar conformado por ontologías, constituye una estructura adaptable que permite integrar distintos tipos de información y posibilita la integración de datos, metadatos y conocimiento espacial. Recoge la conceptualización de los datos y formaliza el conocimiento en torno a ellos y sus relaciones. También es un enlace físico para la recuperación óptima de información, principalmente en aquellos datos que no están asociados a sus metadatos.

La facultad integradora que tienen las ontologías permite obtener de ellas productos de información, como el gazetteer que aún están ausentes en nuestro país.

Referencias

GRUBER T. (1993). "Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Technical Report KSL 93-04". *Knowledge Systems Laboratory, Stanford University*. Recuperado de:

<http://citeseerx.ist.psu.edu/viewdoc/downloaddoi=10.1.1.89.5775&rep=rep1&type=pdf>

IGAC (s. d.). "Datos Geográficos y almacenamiento". Recuperado de: http://geoservice.igac.gov.co/contenidos_telecentro/fundamentos_sig/cursos/sem_2/uni2/index.php?id=1

LATORRES, E. (2010). "IDEVS.UY: Un propuesta de Integración Semántica de Conocimiento e Infraestructura de Datos Espaciales." *I Congreso Uruguayo de Infraestructura de Datos Espaciales. Contribuyendo al Desarrollo de una Red Regional*. Montevideo: ACESIC, 188 -200.

RAJABIFARD, A. (2006). "The role of sub-national government and the private sector in future spatial data infrastructures. *International Journal of Geographical Information Science*, 20 (7), Agosto, 727-741.

ROHRER, E.; MOTZ, R.; SEVERI, P. (2014). "Reasoning for *ALCQM* Extended with a Flexible Meta-Modelling Hierarchy". *JIST2014*, 47-62.

Recuperado de: http://link.springer.com/chapter/10.1007%2F978-3-319-15615-6_4

UNESCO. Tesauro de la UNESCO [en línea]
Recuperado de: <http://databases.unesco.org/thessp/>

VIZCAYA, D. (1997). *Información: Procesamiento de contenido*. Rosario: Parhadigma.