

CONFLACIÓN SEMÁNTICA: UN ESTUDIO SOBRE LA INTEGRACIÓN DE INFORMACIÓN GEOESPACIAL BASADA EN ONTOLOGÍAS

LUIS M. VILCHES-BLÁZQUEZ¹, JOSÉ ÁNGEL RAMOS GARGANTILLA²
Ontology Engineering Group, Facultad de Informática, Univ. Politécnica de Madrid
Avda Montepríncipe s/n, 28660 Boadilla del Monte (Madrid), España
{¹lmvilches, ²jarg}@fi.upm.es

RESUMEN

La conflación semántica es la integración de información geoespacial de múltiples fuentes en un conjunto compuesto de fenómenos (*features*) que poseen mayor precisión en posición y atributos que las fuentes originales. En la conflación semántica el tratamiento de la información tiene una componente más abstracta y afecta a todos los actores que intervienen en el ciclo de vida de la información geográfica. La existencia de gran cantidad de productores de información geográfica provoca la aparición de multitud de fuentes con información similar pero con modelizaciones (representaciones de esa información) y realizaciones (tecnologías que soportan las modelizaciones) muy dispares entre sí.

En este artículo se proporciona una revisión de los diferentes enfoques existentes para llevar a cabo la conflación semántica entre diversos conjuntos de datos geoespaciales. Para ello, se realiza un análisis de las principales fuentes de información geográfica existentes (de nivel mundial a regional). Esto da una idea de la gran heterogeneidad que existe entre ellas. Asimismo, se estudian todos los tipos de heterogeneidad que se presentan y a qué niveles se dan en estas fuentes de información. A continuación, se realiza una breve introducción a las ontologías de dominio geoespacial. Dichas ontologías resultan un elemento principal de los diversos enfoques descritos sobre conflación semántica. Finalmente, se presentan unas conclusiones.

Palabras clave: Información geográfica, heterogeneidad, conflación semántica, ontología.

SEMANTIC CONFLATION

ABSTRACT

Semantic conflation merges geospatial information from multiple sources into a combined set of features that is superior in positional accuracy and attribution than the original sources. Semantic conflation involves an abstract dealing with information and relates all members associated with life cycle of geographical information. There is multiple geographical information producers, which entails that different sources appear. Often, these sources have similar information, but associated models (shapes of information) and implementations (technologies related to models) are very different between them.

In this paper we provide a survey of existing approaches related to semantic conflation in the geospatial realm. So, we perform an analysis of main geographical information sources (from worldwide to regional level). This analysis shows existing heterogeneity in the examined sources. Besides, we describe different heterogeneity types that appear and identify level where heterogeneity is presented in the sources. Next, we provide a brief overview of geospatial domain ontologies, which are main component of the diverse semantic conflation approaches described. Finally, we present some conclusions.

Keywords: Geographic information, heterogeneity, semantic conflation, ontology.

1. Introducción

La Información Geográfica (IG) está constituida por dos componentes: por un lado, el cognitivo o conceptual (fenómenos con una localización en el espacio, cuya instanciación serían los topónimos o nombres geográficos, tradicionalmente conocida como información alfanumérica) y, por otro lado, el componente representacional o cartográfico (plasmación gráfica de los elementos conceptuales, también conocida como información vectorial o ráster). A modo de ejemplo, el plano cognitivo de IG sería el siguiente:

Fenómeno: Edificio religioso cristiano. Catedral

Topónimo: Catedral de Santiago de Compostela

Coordenadas (Grados decimales):

Latitud: 42.88048;

Longitud: -8.545721

Línea de contorno: gruesa

Color de fondo: 235 180 87 (RGB)

Tipo de figura: polígono

En el plano representacional, un ejemplo de IG (información vectorial) se muestra en la [figura 1](#).

Existen diversas comunidades encargadas de producir y explotar este tipo de información geográfica. Con frecuencia, estas comunidades no comparten plenamente intereses y necesidades. Unas proporcionan información de referencia o topográfica, otras usos del suelo,

aspectos geológicos, etc. También existen diferentes escalas de representación y otros tipos de información. Estas diferencias son reflejo de la distinta abstracción de los fenómenos geográficos presentes en el mundo real y originan que haya diversidad, tanto en la representación gráfica (cartografía) como cognitiva (clasificación del conocimiento). Además, estas variaciones están condicionadas por las necesidades de las aplicaciones que explotan este tipo de información.

La demanda actual de IG y la utilización de ésta como instrumento para facilitar la toma de decisiones y la gestión de recursos en diversas áreas (catastro, recursos naturales, etc.) hace que los geodatos se estén convirtiendo en una necesidad de primer orden, consecuencia de su potencial. 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 conflación. Este concepto está adquiriendo una doble perspectiva; por un lado, la visión tradicional del término conflación (OGC, 2008), vinculada a la representación gráfica de la IG mediante información vectorial o ráster, está referida al proceso de unificación gráfica o representacional de múltiples fuentes de datos en una visión integrada que abarque toda la información. Por tanto, el núcleo del proceso de conflación se centra en identificar y asociar los elementos geométricos comunes entre las múltiples fuentes, reconciliar las diferencias entre ellos y construir un resultado integrado. Un ejemplo de esta perspectiva se muestra en la [figura 2](#).

Por otro lado, aparece un nuevo punto de vista que relaciona la perspectiva semántica con el concepto conflación (vinculada con las clasificaciones de conocimiento). Este hecho es resultado de la alta variabilidad de los catálogos de fenómenos (u objetos) de los distintos sistemas que soportan IG, como son los Sistemas de Información Geográfica e Infraestructuras de Datos Espaciales. Esta variabilidad en las clasificaciones repercute en diferencias y variaciones notables a nivel de tipos de fenómenos geográficos, como manifiesta el uso de diferentes terminologías para representar una misma información. Este hecho es una barrera al superponer y mezclar datos de diferentes fuentes en un entorno interactivo. Desde la perspectiva de la conflación semántica, estas terminologías deben integrarse en un modelo que sea capaz de armonizar los conceptos geográficos presentes en los múltiples catálogos.

Desde una visión general, las soluciones aportadas por la Ingeniería Ontológica (Gómez-Pérez *et al.*, 2003), aunque poco aplicadas al campo de la IG, son en general muy satisfactorias. En la actualidad parece reconocerse ampliamente que las ontologías tienen significativos beneficios para capturar y compartir la semántica de la información geoespacial (Klien y Probst, 2005) y servir para el desarrollo de sofisticados SIG (por ejemplo, Fonseca *et al.*, 2002) y/o IDE (por ejemplo, Lacasta *et al.*, 2005 o Lutz *et al.*, 2009). Este hecho queda respaldado con iniciativas como la del *W3C Geospatial Incubator Group* o con el inicio del desarrollo de estándares internacionales en el contexto de *International Organization for Standardization* (lo que será el estándar ISO 19150).

Una posible solución para superar estos problemas de heterogeneidad semántica es la explicitación del conocimiento mediante ontologías que identifiquen asociaciones entre conceptos de recursos ya existentes en el ámbito geográfico (Wache *et al.*, 2001). Actualmente

no existe ninguna implementación de fusión de IG con diferentes catálogos de fenómenos que resuelva el problema en tiempo real. De hecho, un requisito previo sería disponer de tales catálogos de fenómenos publicados en la *Web* y de forma estándar (conforme a ISO, 2005) o de una ontología formalizada en RDF u OWL, cosa que hoy en día apenas ha empezado a desarrollarse.

Este artículo se organiza de la siguiente manera. En la sección 2 se realiza una breve descripción de las principales formas de estructuración de la información geográfica y se detallan, de forma no exhaustiva, a diferentes niveles, algunas fuentes de información actuales. La sección 3 recoge una visión general del estado de la cuestión sobre heterogeneidad y, además, se circunscribe al contexto de la información geoespacial. La sección 4 trata sobre ontologías de dominio geoespacial y en la sección 5 se presentan diferentes enfoques vinculados a la conflación semántica. Finalmente, se exponen unas breves conclusiones y líneas futuras.

2. Tipos de fuentes de información

En la actualidad, el catálogo de fenómenos (*feature catalogue*) es la forma comúnmente empleada para estructurar nombres, códigos, atributos y otras características (inclusive las referidas a la representación gráfica) asociadas a los fenómenos geográficos. La Norma 19110 (ISO, 2005), define el elemento más básico de estos catálogos geográficos (**fenómeno**) como una abstracción de un ente del mundo real asociado a una localización relativa de la Tierra, sobre el que se recolectan, mantienen y diseminan datos. Los fenómenos pueden incluir representaciones de un amplio rango de elementos que pueden localizarse en el tiempo y el espacio (como edificios, ciudades o pueblos), una red geométrica, o una imagen, punto o capa temática georreferenciada.

Los **catálogos de fenómenos** (ISO, 2005) definen los tipos de fenómenos, sus operaciones, atributos y asociaciones, además, son indispensables para convertir los datos en información utilizable. Estos catálogos promueven la diseminación y el uso compartido de los datos geográficos ya que proporcionan un mejor entendimiento del contenido y significado de los mismos. De esta manera, los proveedores y usuarios de datos geoespaciales pueden tener un entendimiento común de los tipos de elementos del mundo real representados por los datos y los usuarios pueden juzgar si se ajustan a su propósito. En la práctica, estos catálogos (como Nomenclátor Geográfico Nacional o *EuroGlobalMaps*) consisten en una lista de fenómenos geográficos agrupados en diferentes clases y donde dichos fenómenos tienen instancias con características comunes. En ocasiones, este tipo de catálogos pueden contener, asociados a los diferentes tipos de fenómenos, definiciones y diferentes tipos de atributos.

Otra forma común para organizar la información geoespacial es el uso de **tesauros**. Sin embargo, éstos no están extendidos entre los productores de IG. Según ISO (1985) e ISO (1986), un tesauro es un vocabulario de un lenguaje de indización controlado (conjunto controlado de términos extraídos del lenguaje natural y utilizados para representar, de forma breve, los temas de los documentos), organizado formalmente con objeto de hacer explícitas las relaciones, a

priori, entre conceptos (por ejemplo "más genérico que" o "más específico que"). Una recopilación de tesauros geográficos e históricos aparece en Vilches-Blázquez *et al.* (2007b).

La construcción de tesauros supone una considerable mejora en la estructuración respecto a un simple listado de fenómenos geográficos, consecuencia de la desaparición de la imprecisión y ambigüedad en el uso del lenguaje (motivada por la existencia de sinónimos y polisemias) y del aporte de semántica adicional (aunque ligera) en sus relaciones (por ejemplo: "Término Genérico", "Término Específico", "Usa", etc.) entre los términos, haciendo que sean interpretadas de modo inequívoco. Sin embargo, según Lassila y MacGuinness (2001), los tesauros no proporcionan una jerarquía explícita (aunque el *brother than* puede usarse como "parte-de" o "subclase-de", tiene problemas a la hora de aplicar herencia de propiedades).

Desde un punto de vista cognitivo, el concepto más importante para la IG es el **fenómeno**. Esta afirmación se deriva de la declaración, por parte del *Open Geospatial Consortium* (OGC, 2003), de que un fenómeno geográfico es el punto de partida para el modelado de la IG. Por esta razón, la unidad básica de la IG en la mayoría de los modelos es el **fenómeno**, que se define como una abstracción de un ente del mundo real asociado a una localización relativa de la Tierra, sobre el que se recolectan, mantienen y diseminan datos (ISO, 2005). De esta manera, tradicionalmente, un fenómeno encapsula en una entidad todo aquello que un dominio dado consideraba sobre un fenómeno geográfico único (Greenwood y Hart, 2003). Desde este punto de vista, se puede ver que la heterogeneidad asociada al término fenómeno varía mucho por los intereses y necesidades de los diferentes productores de IG.

En 2001, Lassila y colaboradores (Lassila y MacGuinness, 2001) presentan una clasificación de diferentes tipos de ontologías en las que incorpora recursos tradicionales que consiste en un orden creciente de complejidad semántica, tal como se muestra en la [figura 3](#). Una recopilación detallada sobre los diferentes tipos y clasificaciones de ontologías existentes aparece en Lacasta (2009).

En definitiva, al igual que ocurre con otros dominios, es posible incrementar la semántica formal existente en catálogos, tesauros y lexicones y transformar en ontología. Las ontologías solucionan los siguientes inconvenientes de las fuentes: (1) la ausencia de una estructuración óptima, ya que, con frecuencia, estas formas se limitan a simples representaciones jerárquicas, más o menos explícitas, determinadas, en ocasiones, por identificadores (ID) o especificaciones del tipo más genérico y/o más específico asociadas a cada fenómeno; (2) la simplicidad y, en ocasiones, la ausencia de relaciones entre los términos y (3) la ausencia de una semántica formal claramente definida. Estos factores provocan un impacto negativo sobre la eficiencia de los procesos de búsqueda, acceso y recuperación de información, así como importantes dificultades en procesos de integración semántica de fuentes heterogéneas de IG. A continuación se van a describir de forma no exhaustiva algunas de las principales fuentes de IG.

- **Fuentes a nivel mundial**

Dentro de estos recursos, una de las fuentes más extendidas es el *Getty Thesaurus of Geographic Names* (TGN) (Harpring, 1997). En TGN cada registro muestra las coordenadas espaciales asociadas, la posición jerárquica, una breve explicación de la región y los nombres de lugares asociados. Además, los nombres para un lugar pueden incluir nombres en la lengua vernácula, inglés, otros idiomas, nombres históricos. Para más detalles, sobre éste y otros recursos, puede consultarse la recopilación de tesauros geográficos e históricos realizada en Vilches-Blázquez *et al.* (2007b).

- **Fuentes a nivel europeo**

Entre las fuentes de información europeas, se van a presentar tres fuentes de referencia a este nivel: *EuroGlobalMap*, *EuroRegionalMap* y *CORINE Land Cover*.

*EuroGeographics*¹ representa a casi todos los Organismos Cartográficos y Catastrales Nacionales de Europa (*National Mapping and Cadastral Agencies-NMCA*). Esta asociación supervisa, entre otros, los proyectos *EuroGlobalMap* y *EuroRegionalMap*. El proyecto *EuroRegionalMap*² consiste en crear una base de datos topográfica vectorial paneuropea a escala 1:250.000, que se usará como base geográfica de referencia para la presentación, visualización y análisis espacial. Está basada en la armonización de las bases de datos geográficas nacionales de las NMCA ya existentes. *EuroGlobalMap*³ es un conjunto de datos topográficos que cubre toda Europa a escala 1:1.000.000, producido en cooperación con los Organismos Cartográficos Nacionales (*National Mapping Agencies – NMA*) de Europa, usando bases de datos oficiales.

Finalmente, *CORINE (Coordination of Information on the Environment) Land Cover* se llevó a cabo por la Comisión Europea en una primera versión entre 1985 y 1990. De este programa se obtuvieron:

- Un sistema de información del estado del medio ambiente en la Unión Europea (sistema *CORINE*), compuesto de una serie de bases de datos que también incluyen información topográfica de referencia.
- Nomenclátor (44 clases) y metodologías para llevar a cabo el programa, usados ambos como referencia a nivel de la Unión.

En una fase posterior se llevó a cabo un proceso de actualización de este proyecto. *CORINE Land Cover 2000* cubre las actividades de fotointerpretación y cartografía de los cambios en la ocupación del suelo. Los productos resultantes de esta fase son: la base de datos de ocupación del suelo para el año 2000 y la base de datos de cambios en la ocupación del suelo durante la década 1990-2000. *CORINE Land Cover* finalizó en el año 2000 y en la actualidad tiene su continuación en el proyecto denominado *Image & CORINE Land Cover 2000 (I&CLC2000)*, cuyo objetivo es actualizar la base de datos *CORINE Land Cover*.

- **Fuentes a nivel nacional**

Por encima de las múltiples fuentes de información existentes a nivel nacional destacan las procedentes del Instituto Geográfico Nacional (IGN), consecuencia de su papel como productor oficial de IG en el ámbito nacional. El IGN dispone, entre otras, de las siguientes fuentes de información: el Nomenclátor Conciso (NC) (1:1.000.000), el Nomenclátor Geográfico Nacional (NGN) (1:50.000), la Base Cartográfica Numérica a escala 1:200.000 (BCN200) y la Base Cartográfica Numérica a escala 1:25.000 (BCN25). Estas fuentes se mantienen separadamente y presentan gran heterogeneidad en diferentes aspectos.

Cada una de las dos Bases Cartográficas Numéricas dispone de su propio catálogo de fenómenos que proporcionan una abstracción de la realidad y una clasificación definida de los fenómenos. A continuación se va a proceder a describir cada una de estas fuentes de manera más detallada.

La Base Cartográfica Numérica BCN25 se diseñó como un producto derivado del Mapa Topográfico Nacional creado a escala 1:25.000 (MTN25). Fue construida para obtener información geográfica que cumpliera con los requisitos de datos requeridos para ser explotados dentro del entorno de los Sistemas de Información Geográfica (SIG). Por lo tanto, BCN25 contiene esencialmente la misma información que MTN25, con propiedades topológicas y geométricas adicionales, siguiendo un modelo específico orientado a base de datos y un catálogo de fenómenos (Rodríguez Pascual y García Asensio, 2005). La [figura 4](#) muestra una pequeña parte del catálogo de fenómenos de BCN25. Por otro lado, la primera versión de la Base Cartográfica Numérica a escala 1:200.000 (BCN200) se inició en 1985. Este trabajo se realizó mediante la digitalización de mapas analógicos de esta escala (Sevilla *et al.*, 2006). A continuación, se muestra un extracto del catálogo de fenómenos de BCN200:

La información contenida en ambas Bases Cartográficas Numéricas está estructurada en ocho temas diferentes (límites administrativos, relieve, hidrografía, vegetación, edificaciones, comunicaciones, líneas de conducción y toponimia). Estos temas, a su vez, recogen diferentes tipos de fenómenos geográficos. Cada fenómeno está codificado con tres pares de dígitos: dos para el tema, dos para el grupo (parte de información homogénea estructurada en tópicos) y dos para el subgrupo (una extensión del fenómeno geográfico asociado al grupo). Estos números describen y clasifican diferentes fenómenos de acuerdo con su temática y atributos. El siguiente ejemplo muestra esta estructura:

030101 Río permanente representado con una línea simple

Tema: **03** Hidrografía

Grupo: **01** Curso de agua permanente

Subgrupo: **01** Curso de agua simbolizado con una línea simple

Asimismo, en la [figura 5](#) se pueden ver otros códigos en diferentes columnas que simbolizan características gráficas de los fenómenos para indicar su representación en los Sistemas Asistidos por Ordenador (*Computer Aided Design-CAD*). Estos dígitos corresponden con los atributos: nivel (LV), color (COL), peso (PS) y estilo (LC).

Respecto a los **nomenclátors**, la definición extendida de este concepto proviene de la norma ISO 19112 (ISO, 2003). Este estándar define un nomenclátor como un directorio de instancias de una clase o clases de fenómenos que contienen alguna información de posición.

El Nomenclátor Geográfico Nacional (NGN), también llamado Base de Datos Georreferenciada de Nombres Geográficos o NOMGEO, alberga información desde 1:50.000 a escalas superiores. Tiene aproximadamente 460.000 registros con coordenadas geográficas, que pertenecen a 44 fenómenos diferentes en español, gallego, catalán, vasco y aranés (lenguas oficiales de España). El nomenclátor es, además, fuente de información del servicio *web* de nombres geográficos⁴ de la Infraestructura de Datos Espaciales de España (IDEE)⁵. Puede verse la clasificación de los fenómenos de este nomenclátor en la [figura 6](#).

El Nomenclátor Conciso (NC) recoge información a escala 1:1.000.000. Esta fuente constituye un *corpus* básico de topónimos estandarizados creado por la Comisión Española de Nombres Geográficos. La primera versión de esta fuente tiene 2.667 topónimos que pertenecen a 22 fenómenos ([figura 7](#)). Este nomenclátor sigue las recomendaciones de la Conferencia de Naciones Unidas sobre normalización de nombres geográficos. Además, el nomenclátor tiene 17 campos, de los cuales algunos son obligatorios: nombre, lengua del nombre, grupo, tipo de fenómeno, provincia, comunidad autónoma, latitud, longitud, mapa y fuente del nombre; y otros opcionales: variante, lengua de la variante, anterior, lengua de anterior, municipio, fuente de la variante y observaciones. Estos *ítems* son conformes al Modelo de Nomenclátors Español (Rodríguez Pascual *et al.*, 2006). Para más detalles, se puede consultar (NGC, 2006).

• Fuentes a nivel regional

Algunas Comunidades Autónomas (CCAA) tienen organismos o institutos propios destinados a la producción de IG y cada uno tiene catálogos de información propios. A continuación se describen las características principales de algunos de éstos:

- Castilla y León (JCYL, 2005): Los fenómenos están codificados jerárquicamente con 6 dígitos, agrupados en temas (11), grupos (60) y subgrupos.
- Galicia (Xunta de Galicia, 2004): Los fenómenos están codificados jerárquicamente con 6 dígitos, agrupados en 17 bloques que se dividen, a su vez, en sub-bloques. Los nombres de los fenómenos aparecen en gallego.
- Cataluña (ICC, 2001): Los fenómenos están codificados jerárquicamente con 5 dígitos, el primero de ellos los divide en 7 bloques y los dos siguientes en 61 sub-bloques.
- País Vasco (GV, 2005): Los fenómenos se llaman capas y se codifican jerárquicamente con 6 dígitos, de los cuales la primera división indica el tema al que pertenecen, habiendo 28 temas posibles.
- Extremadura (Junta de Extremadura, 2005): Los fenómenos están codificados con 5 dígitos, los dos primeros fijos y el tercero indica el tema (10).
- La Rioja (GLR, 2005): Los fenómenos están codificados jerárquicamente con 6 dígitos y organizados jerárquicamente en temas (13), grupos y subgrupos.

- Navarra (GN, 2000)0: Los fenómenos están codificados con 4 dígitos y divididos en 16 temas.

De todo este repaso al estado de la cuestión se puede concluir que existe mucha heterogeneidad entre las fuentes, que se refleja fundamentalmente en una granularidad diferente y en un conjunto diverso de esquemas de representación de la información.

3. Heterogeneidad: Contexto de la información geoespacial

La heterogeneidad se define como la "mezcla de partes de diversa naturaleza en un todo". En el dominio de la IG, cualquier diferencia en las fuentes de datos, entre disciplinas, herramientas y/o repositorios de datos puede provocar heterogeneidad (Alonso y Abbadí, 1994). Fielding y colaboradores (Fielding *et al.*, 2004) identifican que la causa de este problema se encuentra en las visiones específicas del mundo o a las diferentes comunidades de información geoespacial. Estas razones para la heterogeneidad conducen a la identificación de diversos tipos. A continuación, se presenta un conjunto representativo de estos tipos de heterogeneidad comúnmente distinguidos en el entorno de la IG.

En la **heterogeneidad esquemática** (Bishr, 1998) o heterogeneidad en el modelo conceptual (Hakimpour, 2003) puede variar la clasificación, la estructura jerárquica y la representación de los tipos de fenómenos geográficos, dentro o a través de las disciplinas. Se puede ver un ejemplo de esta heterogeneidad en el extracto de dos bases de datos reflejado en la [figura 8](#).

Con respecto a la **heterogeneidad sintáctica** (Bishr, 1998). Ésta se divide en dos tipos; por un lado, está relacionada con el modelo de datos lógico y su subyacente Sistema Gestor de Base de Datos (SGBD), por ejemplo, relacional u orientado a objetos. Mientras, por otro lado, está relacionado con la representación de los objetos espaciales en la base de datos.

La **heterogeneidad en el modelado espacial** (Hakimpour, 2003) de un tipo de fenómeno podría estar representado por un polígono (o un sector de píxeles) en un sistema, mientras que este tipo de fenómeno, en un segundo sistema, podría estar representado por líneas. Un ejemplo de disparidad representacional aparece en la [tabla 1](#).

Por otro lado, se encuentran las heterogeneidades relacionadas con la semántica. Desde el dominio de la lingüística, la semántica se define como las relaciones entre las palabras y las cosas a las que esas palabras se refieren (ter Bekke, 1991). En un plano más práctico, la **heterogeneidad semántica** aparece, habitualmente, como la fuente de la mayoría de los problemas al compartir la información. De forma generalizada, este tipo de heterogeneidad es causada por la representación, significado, interpretación o uso desigual de los datos (es decir, al relacionar los datos con lo que representan) causada por un entendimiento particular del mundo asociado a un determinado contexto (Sheth y Larson, 1990; Naiman y Ouksel, 1995; Wache *et al.*, 2001; Hakimpour, 2003). Este tipo de heterogeneidad recoge diferentes aspectos dependiendo de las consideraciones de los autores.

Según (Bishr, 1998) este tipo de heterogeneidad ocurre como consecuencia de la variación de los modelos entre las diferentes disciplinas y necesidades, incluso aunque los fenómenos geográficos, probablemente, compartan un interés común. Esta heterogeneidad se subdivide en heterogeneidad cognitiva y heterogeneidad de nombrado:

- La heterogeneidad cognitiva es frecuente cuando no existe una base común de definiciones para los fenómenos comunes de diferentes catálogos o bases de datos.
- La heterogeneidad de nombrado es debida a que fenómenos similares semánticamente, pueden nombrarse de forma diferente. Por ejemplo, "curso de agua" y "río" son dos nombres utilizados para describir el mismo fenómeno geográfico. Este tipo de heterogeneidad se identifica y concreta respecto a sinónimos y homónimos en Chatterjee y Segev (1995); Batini *et al.* (1986); Naiman y Ouksel (1995).

En Hakimpour (2003), un sistema puede adoptar un punto de vista sobre un tipo de fenómeno o instancia, mientras que otro sistema distinto puede adoptar un punto de vista totalmente diferente. Además, es habitual encontrar diferentes definiciones de un mismo fenómeno. Desde la comunidad de bases de datos y en el contexto de la federación de bases de datos, Kim y Seo (1991) también proponen en su clasificación disparidades a nivel semántico. Estas se refieren al contenido y significado de cada elemento de información.

Mientras, Naiman y colaboradores (Naiman y Ouksel, 1995) mencionan dos conflictos semánticos: abstracción y niveles de heterogeneidad:

- Abstracción: este tipo de conflicto establece una relación de abstracción entre dos elementos esquemáticos. Este conflicto se subdivide en generalización y agregación.
 - Generalización, ocurre consecuencia de la existencia de conceptos abstraídos a diferentes niveles en bases de datos distintas.
 - Agregación, ocurre cuando un objeto en una base de datos se relaciona (*mapping*) con un grupo de objetos en otra base de datos.
- Niveles de heterogeneidad: los conceptos correspondientes pueden representarse a diferentes niveles en cada base de datos y esta relación estructural se expresa por los siguientes pares: "objeto, objeto", "objeto, atributo", "objeto, instancia" y otros.

Goh (1997) identifica tres áreas principales como parte de la heterogeneidad semántica:

- Conflictos de confusión. Éstos suceden cuando los campos de información parecen tener el mismo significado, pero en realidad son diferentes, por ejemplo, conflictos propios de diferentes contextos temporales.
- Conflictos de escala. Ocurren cuando se utilizan diferentes sistemas de referencia para medir un valor. Un ejemplo típico en este tipo de conflictos son las diferentes medidas que se utilizan en lo que a unidades de longitud se refiere; así, mientras los anglosajones utilizan como unidad básica la yarda, el resto de la comunidad emplea como unidad el metro.
- Conflictos de nombrado. Dichos conflictos acaecen cuando los esquemas de nombrado de

la información difieren significativamente. Un fenómeno frecuente es la presencia de homónimos y sinónimos.

A modo de ilustrar las diferentes clasificaciones de disparidades presentes en la literatura, cabe citar la recopilación y puesta en conjunto de Barrasa (2007), resumida en la [figura 9](#).

4. Ontologías de dominio geoespacial

En el contexto de la informática se define ontología como una especificación explícita y formal de una conceptualización compartida (Gruber, 1993). Detrás de cada término clave de la definición hay una interpretación (Studer *et al.*, 1998): "explícita" se refiere a que los tipos de los conceptos utilizados y las restricciones de su uso se definen explícitamente; "formal" se refiere a la particularidad de que la ontología ha de poder ser leída por una máquina; "conceptualización" es un modelo abstracto de algún fenómeno que tiene identificados los conceptos relevantes de dicho fenómeno; y "compartida" se refiere al consenso de los conocimientos que contiene, la aceptación de los conocimientos por un grupo de personas.

Desde el punto de vista ontológico, el uso de vocabularios estandarizados, como por ejemplo *CORINE Land Cover*⁶, *EuroGlobalMap*³ o *EuroRegionalMap*², implica una simplificación de la compleja realidad existente en cada país, ya que cada productor de IG (tanto nacional como local) tiene diferentes catálogos de fenómenos (recogiendo intereses particulares), lo que provoca que los solapamientos entre fenómenos no sean evidentes totalmente. El uso de la Ingeniería Ontológica (Gómez-Pérez *et al.*, 2008) es un factor clave que puede contribuir a solucionar los problemas de heterogeneidad vinculada a los fenómenos. Las ontologías geográficas son una nueva rama de investigación relacionada con la IG. Para Fonseca y colaboradores (Fonseca *et al.*, 2006) el concepto **geo-ontología** tiene dos tipos básicos de conceptos. Por un lado, los conceptos correspondientes a los fenómenos físicos en el mundo real. Por otro, los conceptos que corresponden a fenómenos del mundo que han sido creados para representar construcciones institucionales y sociales. De esta manera, los conceptos en una geo-ontología se dirigen hacia los objetos espaciales en el mundo.

Este tipo de ontologías (geo-ontologías), según Fonseca *et al.* (2002), son divididas en: ontología de dominio fenomenológico (*Phenomenological Domain Ontology*-PDO) y ontología de dominio de aplicación (*Application Domain Ontology*-ADO). El primer tipo de ontologías (PDO) captura las diferentes dimensiones y propiedades internas de los fenómenos geográficos. Mientras, el segundo tipo (ADO) afectan a la descripción de tareas y temas específicos que los científicos de la IG utilizan como una fuente de información. En definitiva, PDO está relacionado con cómo los fenómenos geográficos pueden ser capturados y representados por los sistemas informáticos. Mientras, ADO trata con la descripción del fenómeno en sí, dónde encaja en el mundo y cómo puede ser mejor descrito.

A esta tipología se le unen los diferentes puntos de vista sobre el mundo procedentes de las variadas comunidades de información geoespacial. Estas vistas pueden ser formalizadas en

diversas ontologías. Por tanto, es necesario acomodar múltiples ontologías para la representación de la realidad geoespacial. Así, en Fonseca *et al.* (2002) aparecen dos formas de integrar ontologías. La integración vertical está relacionada con la integración dentro de un mismo tema. El otro tipo de integración es denominado integración horizontal e involucra la integración de ontologías de diferentes temas.

Estas definiciones hacen entender que las ontologías no están formadas únicamente por meros conceptos con una cierta organización, lo que no distaría de las formas más habituales de organización de la IG (catálogos de fenómenos, diccionarios de datos y/o tesauros), sino que también se van a definir relaciones, atributos, reglas y axiomas entre conceptos que enriquecen y contribuyen, entre otros aspectos, a formalizar el vocabulario y conocimiento del área, dejando a un lado el simple almacenamiento realizado sobre la parte más específica de la información, es decir, los datos.

Por tanto, las ontologías pueden ayudar al mundo geográfico a definir los significados de los fenómenos contenidos en los geo-datos, pudiendo proporcionar la base del entendimiento en el dominio de la IG. Aunque, a menudo, desarrollar una ontología de un dominio no es la meta en sí (sólo es un proceso de definición de un conjunto de datos y sus estructuras para que otros programas las usen), algunas de las razones principales que pueden conducir a la construcción de una ontología son (Noy y MacGuinness, 2001):

- Compartir el entendimiento común sobre un área de conocimiento (dominio geográfico) entre personas y máquinas. Por ejemplo, hidrografía, vegetación o vías de comunicación.
- Permitir la reutilización de conocimientos de un dominio (por ejemplo, hidrográfico) o área de conocimiento (como geografía o geología).
- Permitir cambiar las especificaciones de conocimiento de un dominio si se producen cambios en el mismo. Además, las especificaciones explícitas del dominio de conocimiento son útiles para nuevos usuarios que deben aprender el significado de los términos del área.
- Analizar el conocimiento de un dominio, una vez realizada una especificación completa de los términos que componen el mismo. El análisis formal de los términos es extremadamente valioso al intentar reutilizar ontologías existentes y pretender extenderlas.

De esta sección se pueden extraer varias conclusiones: por un lado, en las ontologías el conocimiento se especifica a través de conceptos, es decir, están formadas por conjuntos de conceptos que dan como resultado una forma de ver la realidad geográfica, común y compartida, acorde con cierta perspectiva ante un dominio de conocimiento. Por otro lado, la posibilidad de formalizar relaciones semánticas, reglas de inferencia, axiomas, etc. entre los conceptos contribuye a enriquecer el conocimiento del dominio en cuestión. Por último, hay que destacar, que esas conceptualizaciones formalizadas permiten la comunicación entre expertos y sistemas informáticos, consecuencia directa de que la información no sólo es entendible por todos los usuarios, como sucedía hasta este momento, sino que también es "comprendida" por las máquinas, lo que repercutirá en una constante reutilización del conocimiento. Por tanto, desde la óptica geoespacial, la construcción de ontologías dará origen a una importante mejora en la representación de la IG, repercutiendo de forma directa en los sistemas de recuperación, consulta y análisis de la misma.

5. Enfoques propuestos en conflación semántica

La IG, como quedó reflejado con anterioridad, 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 su catálogo de fenómenos y su modelo de datos (Vilches-Blázquez *et al.*, 2007a).

Una aproximación para resolver los problemas causados por la heterogeneidad en el contexto de la conflación semántica se centra en la utilización de ontologías. En este caso, las ontologías se utilizan con objeto de crear un esquema conceptual común que traduzca y contemple las diferentes fuentes de información. Por tanto, la descripción de las fuentes de datos mediante ontologías hace que los procesos de identificación, acceso y uso compartido de información puedan manejarse de una manera más eficaz. Esto ayudará a superar los problemas causados por la heterogeneidad semántica propia de la información geoespacial (Kuhn, 1995; Harvey *et al.*, 1999; Mark *et al.*, 1999; Pundt y Bishr, 2002). En este contexto se ubican los diferentes enfoques, reflejados a continuación, que se pueden encontrar en la literatura.

A un nivel genérico existen diferentes estudios sobre el proceso de integración de diversas fuentes. Según Chrisman (2000) se considera que es necesario, para las diferentes opiniones del mundo real, desarrollar un método que pueda integrar y transformar todos los tipos de sistemas de categorías geográficas y resolver su heterogeneidad semántica. Kuhn (2002) investigó el modelado semántico de las categorías geográficas desde la visión de la teoría de la integración conceptual en la ciencia cognitiva. Su estudio revela las ventajas y desventajas de las ontologías que se organizan mediante una estructura en árbol o nivel de conceptos, por ejemplo la ontología representativa de *WordNet*. Smith y colegas (Smith y Mark, 2001) realizaron una serie de experimentos para establecer cómo sujetos no expertos conceptualizaban los fenómenos geográficos y desvelaban emparejamientos considerables entre los significados asignados a los términos "geografía" y "geográfico" por geógrafos y no expertos. En Mark y Turk (2003) se muestra cómo las diferencias entre regiones e idiomas influyen en los individuos en lo que respecta el proceso de conceptualización geográfica y su modelado a través de ontologías.

A un nivel menos teórico y, por tanto, más cercano a la información, se encuentran los siguientes enfoques: en Fonseca *et al.* (2002) el objetivo es establecer una arquitectura SIG que permita la integración de IG, teniendo en cuenta su valor semántico e independencia de representación, de una forma homogénea y flexible. La solución propuesta en este trabajo es un sistema de información geográfica guiado por ontologías (ODGIS) (Fonseca y Egenhofer, 1999) que actúa como un integrador del sistema independiente del modelo. En ODGIS, una ontología es un componente más, como puede serlo una base de datos, que coopera para cumplir con los objetivos del sistema. Asimismo, la utilización de términos de *WordNet* para resolver las cuestiones de conexión entre fuentes de información y el establecimiento de correspondencias entre múltiples ontologías, logrado a través de técnicas de orientación a objetos, proporciona un elevado nivel de integración.

En Uitermark *et al.* (1999a) y Uitermark (2001) se presenta una propuesta para integración de datos espaciales entre dos conjuntos de datos. Este proceso se lleva a cabo por actualización del conjunto de datos más antiguo según el conjunto de datos más actualizado. Los autores definen dos conjuntos de relaciones; un conjunto entre ontología de aplicación y de dominio (relaciones de agregación e igualdad) y el segundo conjunto sólo entre ontologías de aplicación (relaciones de relevancia, relación e igualdad). Las relaciones de similitud son conocidas y utilizadas para relacionar los objetos en dos conjuntos de datos. La localización espacial de los objetos se utiliza para refinar el resultado. El producto final se utiliza para actualizar los conjuntos de datos más antiguos (Uitermark *et al.*, 1999b).

En Stuckenschmidt *et al.* (1999) y Visser *et al.* (2002) se propone una visión general de las ontologías formales y cómo éstas pueden utilizarse en el proceso de integración de información geoespacial de diferentes catálogos. El uso de ontologías en esta propuesta se enfoca hacia el establecimiento de traductores semánticos entre las fuentes de datos. De esta manera, la utilización de ontologías, formalizadas con Ontolingua⁷, proporciona vistas integradas de datos de los diferentes catálogos y procesos de verificación de los datos de ATKIS-OK-250 con los datos de *CORINE Land Cover* o viceversa.

El proceso de integración de conjuntos de datos topográficos realizado en Uitermark *et al.* (2005) establece relaciones entre las instancias de objetos similares, procedentes de diferentes y autónomos conjuntos de datos topográficos del mismo espacio geográfico, considerando las diferencias en contenido y abstracción. Para ello, desarrollan un marco conceptual cuyos componentes principales son las ontologías y los conjuntos de reglas contempladas. Asimismo, introducen un modelo de referencia al que pertenece, únicamente, la combinación de conjuntos de datos topográficos que van a ser integrados.

Fonseca y colegas (Fonseca *et al.*, 2003) proponen una forma de establecer conexión de la representación formal de la semántica (ontologías) a esquemas conceptuales que describen la información almacenada en las bases de datos geográficas. De esta manera, estos autores establecen un marco formal que establece una correspondencia entre una ontología espacial y un esquema conceptual geográfico. La correspondencia de ontologías al esquema conceptual se lleva a cabo mediante el establecimiento de tres niveles de abstracción (formal, dominio y aplicación). A nivel formal, los conceptos abstraídos se utilizan para expresar el esquema y las ontologías. Respecto al nivel de dominio, el esquema es conforme con una instancia de un modelo de datos genérico. Finalmente, el nivel de aplicación se centra en aplicaciones geográficas.

Gómez-Pérez y colaboradores (Gómez-Pérez *et al.*, 2008) integran diferentes fuentes de información con el objetivo de construir un vocabulario común de referencia y, de esta manera, gestionar la enorme cantidad de información manejada por el Instituto Geográfico Nacional. Para ello se creó automáticamente una primera versión de una ontología, denominada *PhenomenOntology*, utilizando la semántica de los catálogos seleccionados. *PhenomenOntology* conceptualiza fenómenos y sus conceptos se relacionan con las filas de las tablas de fenómenos correspondientes en cada base de datos. Tras ser depurada por expertos del dominio, como línea de continuidad de este trabajo se va a proceder a descubrir *mappings* de manera automática con

el objetivo de relacionar los conceptos de la ontología con las instancias de diferentes bases de datos.

Para la integración de información geográfica a través de ontologías, algunos autores proponen un sistema multiontología (Stuckenschmidt *et al.*, 1999; Hakimpour y Timpf, 2001; Fonseca *et al.*, 2002). 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.

En los trabajos de Hakimpour y colegas (Hakimpour y Geppert, 2001) Ose 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 (Sheth y Larson, 1990; Blanco *et al.*, 1994) (este tipo de bases de datos corresponden a una propuesta de integración de información a nivel de esquema).

Green y colaboradores (Green *et al.*, 2008) presentan un prototipo que lleva a cabo la integración de datos de diferentes organizaciones de una manera explícita, produciendo una ontología fusionada que resuelve las cuestiones de integración. Este trabajo ilustra, en el contexto de la contaminación difusa de aguas, diversos aspectos del mencionado prototipo, tales como:

- La descripción del recubrimiento de los dominios por parte de los datos a través de las ontologías de dominio.
- El establecimiento de correspondencias entre las ontologías de dominio y las bases de datos físicas se lleva a cabo mediante una ontología de datos.
- La producción de una ontología fusionada (ontología de aplicación) describe la aplicación por documentación de tareas y combinación de ontologías de dominio que, de forma indirecta, conecta los datos.
- La ejecución de las consultas semánticas incluye elementos espaciales sobre fuentes de datos múltiples utilizando el vocabulario de la ontología de aplicación.

6. Conclusiones

Hoy en día existen varios enfoques y desarrollos de prototipos para realizar la tarea de la conflación semántica en el dominio de la IG. Del repaso aquí descrito a los trabajos de investigación más importantes en el área, se puede extraer que el uso de ontologías es la solución que mejores características aporta tanto para el modelado de las fuentes en sí, como para su compartición dentro de la comunidad o su explotación por aplicaciones y diferentes sistemas que soportan la IG. Las ontologías son el elemento armonizador, que contribuye a solventar las grandes diferencias de heterogeneidad que se dan entre los múltiples recursos y fuentes de información geoespacial.

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, n° 12, p.147-171. ISSN: 1578-5157

Las principales fuentes de información (tesauros y catálogos de fenómenos) y sus correspondientes datos almacenados en bases de datos son integrables a través de la modelización de ontologías. Éstas ofrecen la oportunidad de que las fuentes sigan manteniéndose por separado, a un nivel de integración superior a la organización local (ya sea una integración nacional entre organizaciones regionales o una internacional entre organizaciones nacionales).

Por tanto, la implantación de esta solución mediante el desarrollo de ontologías permite desarrollar los enfoques descritos por estas arquitecturas, ontologías y sistemas contribuyendo a soportar los procesos de integración de múltiples fuentes de información, característicos de la conflación semántica. Asimismo, la generalización de la aplicación de las ontologías conseguirá que la IG goce de una disponibilidad y facilidad de acceso de la que hasta el momento carece o que se encuentra muy limitada.

Referencias bibliográficas

- AdV (1998): *Amtliches Topographisch-Kartographisches Informationssystem ATKIS*. Bonn.
- Alonso, G. y Abbadi, A.E. (1994): "Cooperative modelling in applied geographic research", *Intr. Journal of Intelligent and Cooperative Information Systems*, vol. 3, no. 1, pp 83-102.
- Barrasa, J. (2007): *Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales*. Tesis doctoral, Facultad de Informática, Universidad Politécnica de Madrid. Enero 2007.
- Batini, C.; Lenzerini, M. y Navathe, S. (1986): "A comparative analysis of methodologies for database schema integration", *ACM Computer Survey*, vol. 18, no. 4, pp. 323-364.
- Bishr, Y. (1998): "Overcoming the semantic and other barriers to GIS interoperability", *International Journal of Geographical Information Science*, vol. 12, no. 4, pp. 299-314.
- Blanco, J.M.; Illarramendi, A. y Goñi, A. (1994): "Building a federated database system: An approach using a knowledge based system", *International Journal on Intelligent and Cooperative Information Systems*, vol. 3, no. 4, pp. 415-455.
- Chatterjee, A. y Segev, A. (1995): "Rule Based Joins in Heterogeneous Databases", *Decision Support Systems*, vol. 13, no. 1, pp. 313-333.
- Chrisman, N. (2000): "Building GIS without foundations: ontology from a social practice perspective", *Proceedings of GIScience*, Savannah, Georgia, USA.
- Fielding, J.M.; Simon, J.; Ceusters, W. y Smith, B. (2004): "Ontological Theory for Ontological Engineering: Biomedical Systems Information Integration", *Proceedings of the Ninth Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR2004)*, Whistler, BC.
- Fonseca, F. y Egenhofer, M. (1999): "Ontology-Driven Geographic Information Systems", In Medeiros, C. B. (ed.) *7th ACM Symposium on Advances in GIS*, Kansas City, MO, pp.14-19.

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

Fonseca, F.; Egenhofer, M.; Davis, C.A. y Câmara, G. (2002): "Semantic Granularity in Ontology-Driven Geographic Information Systems", *Annals of Mathematics and Artificial Intelligence*, vol. 36, no. 1-2, p. 121-151.

Fonseca, F.; Davis, C. y Câmara, G. (2003): "Bridging Ontologies and Conceptual Schemas in Geographic Information Integration", *GeoInformatica*, vol. 7, no. 4, pp. 355-378.

Fonseca, F.; Câmara, G. y Monteiro, A.M. (2006) "A Framework for Measuring the Interoperability of Geo-Ontologies", *Spatial Cognition and Computation*, vol. 6 (4), pp. 307-329.

GLR (2005): *Catálogo de topónimos 1:5.000*, Gobierno de La Rioja.

GN (2000): *Revisión de la metodología de producción y actualización de la cartografía a escala 1:5.00'*, Gobierno de Navarra. Departamento de Obras Públicas, Transportes y Comunicaciones.

Goh, C.H. (1997): "Representing and reasoning about semantic conflicts in heterogeneous information sources", Ph.D. dissertation, Massachusetts Institute of Technology (MIT).

Gómez-Pérez, A.; Fernández-López, M. y Corcho, O. (2003): *Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*, Springer-Verlag. London.

Gómez-Pérez, A.; Ramos, J.A.; Rodríguez-Pascual, A.F. y Vilches-Blázquez, L.M. (2008): 'The IGN-E case: Integrating through a hidden ontology', *The 13th Int. Symposium on Spatial Data Handling (SDH 2008)*. Montpellier (France), pp. 417-135. ISBN: 978-3-540-68565-4.

Green, J.; Dolbear, C.; Hart, G.; Engelbrecht, P. y Goodwin, J. (2008): "Creating a semantic integration system using spatial data", *International Semantic Web Conference 2008*, Karlsruhe (Germany).

Greenwood, J. y Hart, G. (2003): "Sharing Feature Based Geographic Information - A Data Model Perspective". *7th Int. Conference on GeoComputation*. United Kingdom.

Gruber, T. (1993): "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220.

GV (2005): *Relación de formas gráficas a escala 1:10.000 y 1:25.000*, Gobierno vasco. Departamento de Ordenación del Territorio y Medio Ambiente.

Hakimpour, F. y Timpf, S. (2001): "Using Ontologies for Resolution of Semantic Heterogeneity in GIS", *4th. AGILE Conference on Geographic Information Science*, Brno (Czech Republic).

Hakimpour, F. y Geppert, A. (2001): "Resolving semantic heterogeneity in schema integration: an Ontology based Approach", *Proc. of Int. Conf. on Formal Ontologies in Information Systems FOIS'01* (Eds. Ch. Welty and B. Smith), New York, ACM Press, pp. 297-308.

Hakimpour, F. (2003): *Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata*, Ph.D. thesis, Zurich University. Switzerland.

Harpring, P. (1997): "Proper words in proper places: The Thesaurus of Geographic Names", *Museum Document Association Information*, vol. 2, no. 3, pp. 5-12.

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): “Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías”, *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

Harvey, F.; Kuhn, W.; Pundt, H.; Bishr, Y. y Riedemann, C. (1999): “Semantic interoperability: A central issue for sharing geographic information”, *The Annals of Regional Science*, vol. 33, no. 2, pp. 213-232.

ICC (2001): *Diccionario de datos de la Base topográfica 1:5.000 v2.0 (BT-5M)*, Institut Cartogràfic de Catalunya.

ISO (1985): *Guidelines for the establishment and development of multilingual thesauri*, International Organization for Standardization 5964:1985.

ISO (1986): *Guidelines for the establishment and development of monolingual thesauri*, International Organization for Standardization 2788:1986.

ISO (2003): *Geographic information -- Spatial referencing by geographic identifiers*, International Organization for Standardization 19112.

ISO (2005): *Geographic Information – Methodology for feature cataloguing*, International Organization for Standardization 19110.

JCYL (2005): *Diccionario de datos cartografía 1:5.000*, versión 1.1. Junta de Castilla y León.

Junta de Extremadura (2005): *Codificación de la cartografía Digital*, Junta de Extremadura. Dirección General de Urbanismo, Arquitectura y Ordenación del Territorio.

Kim, W. y Seo, J. (1991): “Classifying schematic and data heterogeneity in multidatabases systems”, *IEEE Computer*, vol. 24, no. 2, pp. 12-18.

Klien, E. y Probst, F. (2005): “Requirements for geospatial ontology engineering”. *Conference on Geographic Information Science (AGILE 2005)*. Estoril (Portugal).

Kuhn, W. (1995): *Semantics of Geographic Information: GeoInfo 7*, Technical University of Vienna, Austria, Department of Geoinformation, 108 p.

Kuhn, W. (2002): “Modeling the semantics of geographic categories through conceptual integration”, *Proceedings of GIScience*, Boulder, Colorado, USA.

Lacasta, J. (2009): *Contributions to the problem of knowledge management in Spatial Data Infrastructures*. Ph.D. thesis. Universidad de Zaragoza.

Lacasta, J.; Muro-Medrano, P.; Noguerras-Iso, J. y Zarazaga, F.J. (2005): “Web Ontology Service, a key component of a Spatial Data Infrastructure”. *Proceedings of the 11th EC GI&GIS Workshop: ESDI: ESDI Setting the Framework*.

Lassila, O. y McGuinness, D. (2001): *The Role of Frame-Based Representation on the Semantic Web, Technical Report*. Knowledge Systems Laboratory, Stanford University. KSL-01-02.

Lutz, M.; Sprado, J.; Klien, E.; Schubert, C. y Christ, I. (2009): “Overcoming semantic heterogeneity in spatial data infrastructures”. *Computers & Geosciences*, Vol. 35(4), pp.739-752.

Mark, D.M.; Smith, B. y Tversky, B. (1999): “Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization”, *Spatial Information Theory*, Freksa, Chr., Mark, D.M., (eds.). Lecture Notes in Computer Science 1661, Springer, Berlin, New York, pp. 283 – 298.

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

Mark, D.M. y Turk, A.G. (2003): "Landscape Categories in Yindjibarndi: Ontology, Environment, and Language", *Spatial information theory: foundation of geographic information science*, Kuhn, W., Worboys, M. F., Timpf, s. (eds.), LNCS Springer-Verlag, no.2825, p.28-45.

Naiman, C.F. y Ouksel, A.M. (1995): "A classification of semantic conflicts in heterogeneous database systems", *Journal of Organizational Computing*, vol. 5, no. 2, pp. 167-193.

NGC (2006): *Nomenclátor Geográfico Conciso de España versión 1.0. Presentación y Especificaciones*, Instituto Geográfico Nacional. Octubre 2006. (Disponible en: <http://www.ideo.es/ApliVisio/Nomenclator/NGCE.pdf> - Último acceso 2012/02/07)

Noy, N.F. y McGuinness, D.L. (2001), "Ontology Development 101: A Guide to Creating Your First Ontology", *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05* and *Stanford Medical Informatics Technical Report SMI-2001-0880*.

OGC (2003): *OpenGIS Reference Model Version 0.1.2*, Open Geospatial Consortium, OGC Inc. Wayland, MA, USA.

OGC (2008): *OWS-5 Conflation Engineering Report. Discussion Paper*, Pete Brennan (ed.). Version: 0.1.0. Open Geospatial Consortium

Pundt, H. y Bishr, Y. (2002): "Domain Ontologies for Data Sharing-An Example from Environmental Monitoring Using Field GIS", *Computer & Geosciences*, vol. 28(1) pp. 95-102.

Rodríguez Pascual, A.F. y García Asensio, L. (2005): "A fully integrated information system to manage cartographic and geographic data at a 1:25,000 scale", *XXII International Cartographic Conference (ICC2005)*. A Coruña, Spain. ISBN: 0-958-46093-0

Rodríguez Pascual, A.F.; López Romero, E.; Abad Power, P. y Sánchez Maganto, A. (2006): *Modelo de Nomenclátor de España v1.2*, (Disponible en: http://www.ideo.es/resources/recomendacionesCSG/MNEv1_2.pdf - Último acceso 2012/01/07)

Sevilla Sánchez, C.; Rodríguez Pascual, A.F.; González Matesanz, F.J.; Blanco Ortega L.M. y Vilches-Blázquez, L.M. (2006): "Un SIG corporativo en el IGN para la gestión integrada, publicación y análisis de datos geográficos", *XII Congreso Nacional de Tecnologías de la Información Geográfica*. Camacho Olmedo, M.T., Cañete Pérez, J.A. y Lara Valle, J.J. (Eds.)

Sheth, A. y Larson, J. (1990): "Federated database systems for managing distributed, heterogeneous and autonomous databases", *ACM Computing Surveys*, vol. 22(3), pp. 183-236.

Smith, B. y Mark, D.M. (2001): "Geographical Categories: an Ontological Investigation", *International Journal of Geographical Information Science*, vol. 15, no. 7, pp. 591-612.

Stuckenschmidt, H.; Visser, U.; Schuster, G. y Vögele, T. (1999): "Ontologies for geographic information integration", In *Proc. of Intelligent Methods in Environmental Protection: Special Aspects of Processing in Space and Time*, Ubbo Visser and Hardy Pundt (eds.), 13th International Symposium of Computer Science for Environmental Protection (CSEP '99), pp 81-107.

Studer, R.; Benjamins, V.R. y Fensel, D. (1998), "Knowledge Engineering: Principles and Methods", *IEEE Transactions on Data and Knowledge Engineering*, vol. 25, no. 1-2, pp. 161-197.

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): “Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías”, *GeoFocus (Artículos)*, n° 12, p.147-171. ISSN: 1578-5157

ter Bekke, J.H. (1991): *Semantic data modeling in relational environments*, PhD. Thesis, University of Delft.

Uitermark, H.T.; van Oosterom, P.; Mars, N. y Molenaar, M. (1999a), “Ontology-Based Geographic Data Set Integration”, *Proceedings of the International Workshop on Spatio-Temporal Database Management*, Böhlen, M. H., Jensen, C. S., Scholl, M. O. (Eds.) (STDBM'99), LNCS 1678, p. 60-78, Springer-Verlag Berlin Heidelberg.

Uitermark, H.T.; Vogels, A. y van Oosterom, P. (1999b): “Semantic and geometric aspects of integrating road networks”, *Interoperating Geographic Information Systems, 2nd. International Conference*, INTEROP '99 (LNCS 1580), Vckovski, A., Brassel, K. E., and Schek, H.-J., (eds.), pp. 177–188.

Uitermark, H.T.; van Oosterom, P.; Mars, N. y Molenaar, M. (2005): “Ontology - based integration of topographic data sets”, *International Journal of Applied Earth Observation and Geoinformation*, vol. 7, no. 2, pp. 97-106.

Uitermark, H.T. (2001): *Ontology-Based Geographic Data Set Integration*, Ph.D. thesis, 155 pp. University of Twente, The Netherlands.

Vilches-Blázquez, L.M.; Rodríguez Pascual, A. F.; Mezcuca Rodríguez, J.; Bernabé Poveda, M.A. y Corcho, O. (2007a), “An approach towards a harmonized framework for hydrographic features domain”, *Proceedings of XXIII International Cartographic Conference*, 4-10 August 2007. Moscow, Russia.

Vilches-Blázquez, L.M.; Martins, B.; Wytenbach, A.; Bernabé, M.A.; Álvarez, M.; Luzio, J. y Borbinha, J. (2007b): *Geographical and historical thesauri: The state of the art, DIGMAP Deliverable D2.1*. ECP-2005-CULT-038042, DIGMAP. (Disponible en: http://www.digmap.eu/lib/exe/fetch.php?id=wiki%3Aproject_documents%3Adeliverables&cache=cache&media=wiki:ec:digmap-d2.1_v1_2.pdf – Último acceso 2012-02-07)

Visser, U.; Stuckenschmidt, H.; Schuster, G. y Vögele, T. (2002): “Ontologies for geographic information processing”, *Computers & Geosciences*, vol. 28, pp. 103-117.

Wache, H.; Vögele, T.; Visser, U.; Stuckenschmidt, H.; Schuster, G.; Neumann, H. y Hübner, S. (2001): “Ontology-based integration of information – A survey of existing approaches”, In *IJCAI-01 Workshop: Ontologies and Information Sharing*, pp. 108-117.

Xunta de Galicia (2004): *Diccionario de datos de base topográfica 1:5.000 v_1.0*. Noviembre 2004, Xunta de Galicia, Consellería de Política Territorial, Obras Públicas e Vivenda.

TABLAS

Tabla 1: Extracto de dos bases de datos

ID	Nombre	Longitud	Superficie	Coordenadas
190	Río Ebro	2121,33	2121,33	42,7612972; 42,7686864
3077	Río Miño	53574,76	53574,76	43,1791692; 43,1608346

ID	Nombre	Vertiente	Cuenca	Longitud
11	EBRO	ATLÁNTICA	EBRO	910
15	MIÑO	ATLÁNTICA	MIÑO	310

FIGURAS

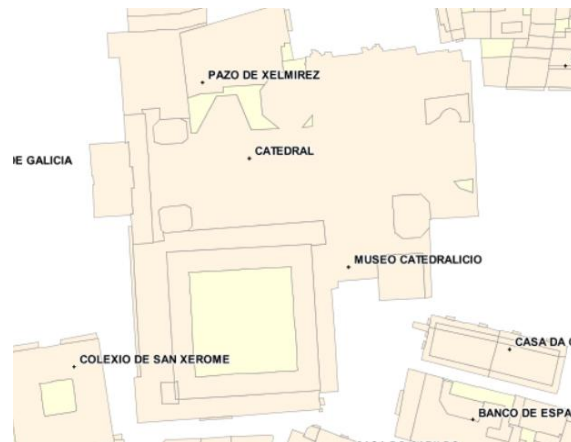


Figura 1: Representación gráfica de la Catedral de Santiago de Compostela.

Fuente: www.ideo.es

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

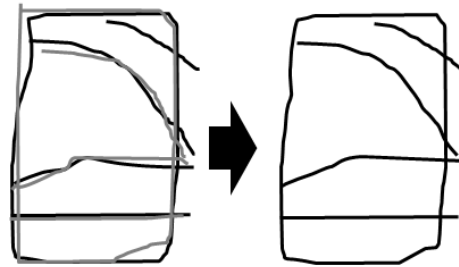


Figura 2: Conflación de dos conjuntos de fenómenos lineales (izquierda) unificados en un único conjunto (derecha)

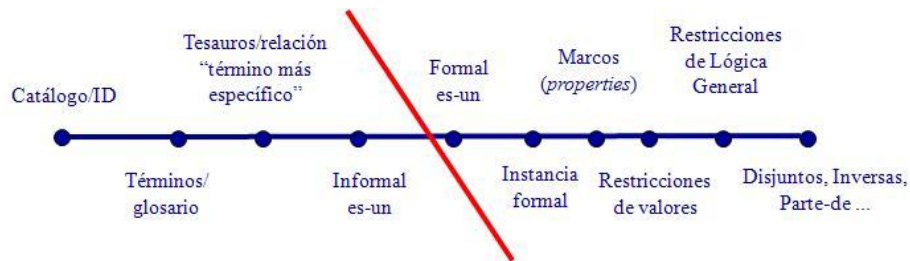


Figura 3: Clasificación de tipos de ontologías (Lassila y MacGuinness, 2001)

Tipo	dgn	Entidad	T_i	Grupo	Código	Cer.	Trato	[!comentario]
02000900	104	1	0	090101	1	!I	Marco de hoja	
02300902	104	2	0	100200	0	!	Base Geodésica de Madr	
06003900	104	3	0	025102	0	!	Acantilado	
06006900	104	4	0	025302	0	!	Costa rocosa no acanti	
06009900	104	5	2	037402	1	!	Playa fluvial de guiija	
06012900	104	6	0	025501	1	!	Lavas. Contorno	
06015900	104	7	0	058303	0	!I	Dique de hormigón >15 :	
06018900	104	8	0	058304	0	!I	Dique de hormigón < 15	
07013400	104	9	0	058302	0	!I	Dique de tierra	
07016400	104	10	0	055401	1	!	Vertedero. Contorno	
11003003	104	11	1	062202	0	!	Autopista. Enlace	
11012000	104	12	0	056091	1	!I	Patio. Contorno	
13003300	104	13	1	060101	0	!	Autopista. Eje	
13303300	104	14	1	060131	0	!	Autopista en Contruucci	
14002401	104	15	1	066901	1	!I	Puesto de S.O.S.	
14003301	104	16	1	067901	1	!I	Peaje	
15003003	104	17	1	062204	0	!	Autovía. Enlace	
15003004	104	18	1	060701	0	!	Autovía	

Figura 4: Fenómenos en BCN25

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

CODIGO	LV	COL	PS	LC	SIMB.	NOMBRE
01.01.01	01	000	00	6	00006	LIMITE_MUNICIPAL
01.01.02	01	015	00	6	03846	LIMITE_MUNICIPAL_PROVISIONAL
01.02.01	01	030	03	4	07708	LIMITE_PROVINCIAL
01.03.01	01	045	06	2	11570	LIMITE_AUTONOMICO
01.04.01	01	060	06	3	15411	LIMITE_NACIONAL
01.05.01	01	075	00	1	19201	AGUAS_JURISDICCIONALES
01.51.01	01	090	00	6	23046	MUNICIPIO_CONTORNO
01.51.31	01	105	00	6	26886	MUNICIPIO_ANEJO
01.51.91	01	120	00	6	30726	MUNICIPIO_ENCLAVE
01.52.01	01	135	03	4	34588	PROVINCIA_CONTORNO
01.52.31	01	150	03	4	38428	PROVINCIA_ANEJO
01.52.91	01	165	03	4	42268	PROVINCIA_ENCLAVE
01.53.01	01	180	06	2	46130	AUTONOMIA_CONTORNO
01.53.31	01	195	06	2	49970	AUTONOMIA_ANEJO
01.53.91	01	210	06	2	53810	AUTONOMIA_ENCLAVE
01.54.01	02	000	06	3	00051	NACION_CONTORNO
01.54.31	02	015	06	3	03891	NACION_ANEJO
01.54.91	02	030	06	3	07731	NACION_ENCLAVE
01.55.01	02	045	00	6	11526	MUNICIPIO_EN_CONJUNTO_CONTORNO
01.55.31	02	060	00	6	15366	MUNICIPIO_EN_CONJUNTO_ANEJO
01.55.91	02	075	00	6	19206	MUNICIPIO_EN_CONJUNTO_ENCLAVE
01.56.01	02	090	00	6	23046	MUNICIPIO_EMPRIVIANO_CONTORNO
01.56.31	02	105	00	6	26886	MUNICIPIO_EMPRIVIANO_ANEJO
01.56.91	02	120	00	6	30726	MUNICIPIO_EMPRIVIANO_ENCLAVE
01.57.01	02	135	00	6	34566	MUNICIPIOS_SIN_DESLINDE_CONTORNO
01.57.31	02	150	00	6	38406	MUNICIPIOS_SIN_DESLINDE_ANEJO
01.57.91	02	165	00	6	42246	MUNICIPIOS_SIN_DESLINDE_ENCLAVE
01.58.01	02	180	00	6	46086	TERRITORIO_EMPRIVIANO_CONTORNO
01.58.31	02	210	00	6	53766	TERRITORIO_EMPRIVIANO_ANEJO
01.58.91	02	225	00	6	57606	TERRITORIO_EMPRIVIANO_ENCLAVE
01.59.01	03	000	03	3	00027	TERRITORIO_ESTATAL_CONTORNO
01.59.31	03	015	03	3	03867	TERRITORIO_ESTATAL_ANEJO
01.59.91	03	030	03	3	07707	TERRITORIO_ESTATAL_ENCLAVE
01.60.01	03	045	03	3	11547	COMARCA_CONTORNO
01.60.31	03	060	03	3	15387	COMARCA_ANEJO
01.60.91	03	075	03	3	19227	COMARCA_ENCLAVE
01.61.01	03	090	03	1	23065	PARQUE_NACIONAL_CONTORNO
01.61.31	03	105	03	1	26905	PARQUE_NACIONAL_ANEJO
01.61.91	03	120	03	1	30745	PARQUE_NACIONAL_ENCLAVE
01.62.01	03	135	03	1	34585	PARQUE_NACIONAL_CONTORNO
01.62.31	03	150	03	1	38425	PARQUE_NACIONAL_ANEJO

Figura 5: Fenómenos en BCN 200

Comunidad Autónoma
Provincia
Comarca administrativa
Capital
Población 1
Población 2
Sierra
Pico
Puerto
Área geográfica
Río
Canal
Lago
Embalse
Cabo
Golfo
Playa
Isla
Mar
Lugar protegido
Aeropuerto
Puerto marítimo

Figura 6: Fenómenos del Nomenclátor Conciso

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, n° 12, p.147-171. ISSN: 1578-5157

NOMENCLÁTOR GEOGRÁFICO
ENTIDADES
Nación
Región geográfica
Capital de Nación
Elevación orográfica
Comunidad Autónoma
Llanura/Raso
Ciudad con Estatuto de Autonomía
Depresión orográfica
Capital de Comunidad Autónoma
Accidente costero
Provincia
Accidente marítimo
Capital de Provincia
Accidente hidrográfico
Coprincipado
Corriente fluvial
Capital de Coprincipado
Canal
Comarca
Embalse
Capital de Comarca
Lago/Laguna
Isla Humedal
Capital de Isla
Isla fluvial
Municipio
Isla marítima
Capital de Municipio
Garganta/Hoz
E.Á.T.I.M.
Lugar/Paraje
Capital de E.Á.T.I.M.
Paso/Collado
Población
Puerto de montaña
Comunidad de Municipios
Puerto comercial
Enclave
Helipuerto comercial
Territorio anejo
Aeródromo/Aeropuerto
Territorio autonómico
Estación de ferrocarril
Zona neutral

Figura 7: Fenómenos del Nomenclátor Geográfico Nacional



Figura 8: Diferente representación geométrica

Vilches-Blázquez, L. M. y Ramos Gargantilla, J.A. (2012): "Conflación semántica: un estudio sobre la integración de información geoespacial basada en ontologías", *GeoFocus (Artículos)*, nº 12, p.147-171. ISSN: 1578-5157

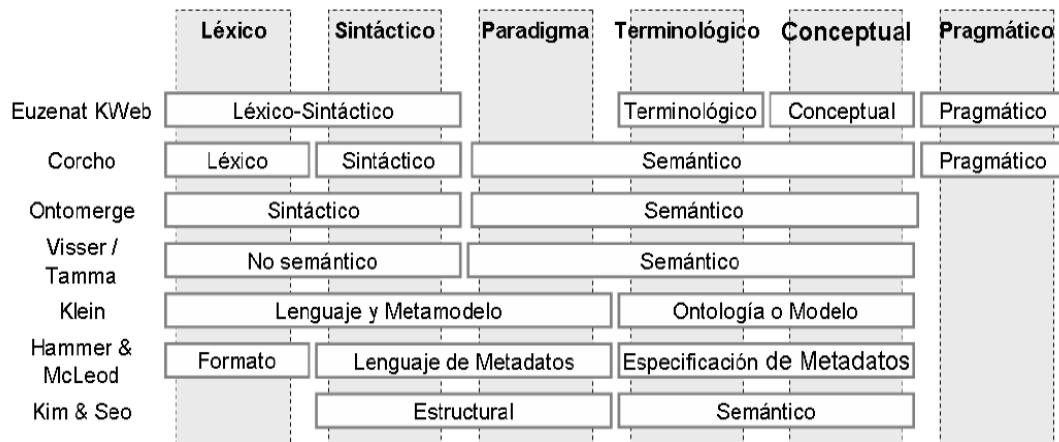


Figura 9: Diferentes tipos de heterogeneidad organizados en niveles conforme a sus caracterizaciones (Barrasa, 2007)

- ¹ http://www.eurogeographics.org/eng/00_home.asp
- ² http://www.eurogeographics.org/eng/03_projects_euroregionalmap.asp
- ³ http://www.eurogeographics.org/eng/03_projects_EGM_overview.asp
- ⁴ <http://www.ideo.es/gazetteerIGN/indexLayout.jsp?PAGELANGUAGE=es>
- ⁵ http://www.ideo.es/show.do?to=pideep_pidee.ES
- ⁶ <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=950>
- ⁷ Ontolingua. <http://www.ksl.stanford.edu/software/ontolingua/>