

Ontologies in the Geographic Information Sector
Authors: Billen, Roland; Nogueras-Iso, Javier; López-Pellicer, F. Javier; Vilches-Blázquez, Luis M.
In: Ontologies in Urban Development Projects
Editors: Falquet, Gilles; Métral, Claudine; Teller, Jacques; Tweed, Christopher
Book Series Title: Advanced Information and Knowledge Processing
Copyright: 2011
Publisher: Springer London
Isbn: 978-0-85729-724-2
Start Page: 83
End Page: 103
Url: http://dx.doi.org/10.1007/978-0-85729-724-2_6
Doi: 10.1007/978-0-85729-724-2_6

Ontologies in the Geographic Information sector

Roland Billen¹, Javier Nogueras-Iso², F. Javier López-Pellicer² and Luis M. Vilches-Blázquez³

¹Geomatics Unit, University of Liege, Belgium
rbillen@ulg.ac.be

²Computer Science and Systems Engineering Department, University of Zaragoza, Spain
{jnog,fjlopez}@unizar.es

³Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
lmvilches@fi.upm.es

Keywords: Geographic Information (GI), Spatial Data Infrastructures (SDI), Geographic Information System (GIS), ontology-based integration of geographic information

1. Introduction

Geographical information (GI) or geoinformation describes phenomena associated directly or indirectly with a location (coordinates systems, address systems...) with respect to the Earth's surface. Such phenomena can be either spatially discrete (represented by geometric primitives like points, lines, regions, etc.) such as a municipality, a road axis, etc. or spatially continuous (represented by interpolation on an image grid for example) such as terrain's elevation, pollution diffusion, etc. GI is created by

manipulating geographic data (or geospatial data) in a computerized system. Geospatial data can be acquired by different means; topographic survey, remote sensing, aerial photographs, GPS, laserscan, and all other types of sensors or survey techniques. Traditionally, these data are the core component of Geographic Information Systems (GIS), which is the term commonly used to refer to the software packages that allow to capture, store, check, integrate, manipulate, analyze and display them.

Geographic information is therefore used in a wide variety of domains; indeed, in any application dealing with spatial or geographical frame of reference. Typical applications are land registration, hydrology, cadastre, land evaluation, planning or environmental observation. The link between urbanism applications and GI domain is obvious as most of information treated in urban applications is indeed GI (maps or spatial databases including information about buildings, networks, terrain, etc.). Therefore, it is reasonable to depict the use of ontologies in the GI sector in the framework of the Towntology project.

The potential of GI as an instrument to facilitate decision-making and resource management in diverse areas (e.g., natural resources, facilities, cadastre or agriculture, urban planning) of government or private sectors has led to the evolution of GIS into the broader concept of *Spatial Data Infrastructure* (SDI). According to the *Global Spatial Data Infrastructure Association Cookbook* (Nebert 2004), “the term Spatial Data Infrastructure is often used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data”. The European Committee for Standardization (CEN) defines the SDI concept as a platform-neutral and implementation neutral technological infrastructure for geospatial data and services, based upon non-proprietary standards and specifications (CEN 2006).

From the previous definitions of SDI it can be derived that one of the main objectives of SDIs is to make the work with geospatial data more efficient (McKee 2000; Nebert 2001), avoiding problems that occur with conventional GIS technology and geographic data sets. Bernard *et al.* (2004) remarks that there are two major problems with traditional GIS stand-alone applications: first, data sets exist in a plethora of different data formats (datasets in different formats often have to be converted in order to be used in a different system); and second, these data are often not (sufficiently) documented (it is difficult or even impossible for outside users to discover data sets and to assess whether a given data set is useful for their tasks). In other words, what these authors are meaning is the inability of

isolate GIS tools to deal with interoperability issues in the current context where GI must be shared between online systems. As mentioned by (Egenhofer 1999) with respect to GI interoperability, “the goal of interoperating GISs is to achieve an automated process that will allow to use data and software services across the boundaries that their collectors and designers envisioned”.

Going a bit further with GI interoperability issues, the main obstacle for the interoperation of systems is the heterogeneity in data and services managed by these systems. In order to determine whether two systems are heterogeneous, one must analyze their different features and this yields different types of heterogeneity as well as different types of interoperability levels. A commonly made distinction is that between syntactic (solving syntactic heterogeneity) and semantic interoperability (solving semantic heterogeneity) (Kolodziej 2003). The syntactic interoperability is concerned with the technical level, i.e. it refers to the ability for a system or components of a system to provide information portability and inter-application as well as cooperative process control. It comprises intercommunication at communication level protocol, hardware, software, and data compatibility layers. The semantic interoperability, in contrast, deals with the domain knowledge necessary for informatics services to "understand" each other's intentions and capabilities.

In order to overcome interoperability problems, GI standards have been developed by organizations and standardization bodies such as the Open Geospatial Consortium (OGC) or ISO/TC211 (ISO technical committee for geographic information and geomatics). The use of GI standards has gradually eliminated many of the difficulties resulting from incompatibility of data structure and syntax but it is not enough to solve completely the problems derived from semantic heterogeneity. According to Bishr (1998), semantic heterogeneity is defined as the consequence of different conceptualizations of a real world fact. Because of different perspectives on the same real world facts, there may not be a common base of definitions of the underlying facts between two disciplines (domains). Derived from these different perspectives, Bishr distinguishes two main subtypes of semantic heterogeneity: cognitive heterogeneity and naming heterogeneity. Cognitive heterogeneity occurs when the same term is used in different domains for representing different concepts. On the other hand, naming heterogeneity occurs when the same real world facts are understood in the same way but are named differently.

Semantic interoperability problems arise in different scenarios of GI interoperability, ranging from discovery and retrieval of GI to the integration of data from different sources. For instance, in the case of GI discovery, though there are standardized interfaces for catalogue services operations (e.g., OGC Catalogue services specifications), the conformance to the specifications does not prevent from having GI catalogues with semantic heterogeneity problems. Catalogue implementations based on simple word-matching between user queries and metadata holdings suffer from typical naming and cognitive heterogeneities in the form of synonymy and homonymy problems respectively (Bernard et al. 2004). And as reported in section 3, similar problems of semantic heterogeneity occur in the case of GI retrieval or integration of data from different sources.

The objective of this chapter will be the study of GI ontologies as a possible approach to facilitate semantic interoperability and overcome the problem of semantic heterogeneity. The explicit definition of knowledge by means of ontologies is commonly used as a mechanism to understand and solve the semantic heterogeneity arisen when interoperating between two systems (Wache *et al.* 2001). Defining, building and using ontologies have become a key research topic in Geographical Information Sciences (GISc). A lot of work has been dedicated to the definition of geographical ontologies and to the use of them in practical applications.

Apart from this introduction section about GI, SDIs and interoperability issues, the remaining parts of this chapter are structured as follows. Section 2 describes the features of geographical information related ontologies. After, we focus on the role of ontologies to facilitate GI interoperability (section 3). Section 4 presents three study cases discussing ontology design methodologies and ontology's uses in the geographical information context. Finally, section 5 gathers conclusions, and section 6 points out open problems and research perspectives.

2. Ontologies in GI

In the GI sector, and more especially in spatial database community, the term ontology is often associated to (Yeung and Hall 2007):

- A *concept* of using formally and explicitly defined terminology and vocabulary to describe real world features or phenomena associated with a specific discipline, domain or application.

- A *systematic collection and specification* of spatial entities, their properties and relations, which are commonly stored in a hierarchical structure and shared by users in a particular discipline or domain.
- An emerging *approach* to designing spatial database systems that has several advantages over conventional methods of systems development, including:
 - Allowing the establishment of correspondence and interrelation among different domains of spatial entities and relations.
 - Contributing to create better information systems by improving communication between systems developers, managers and users.
 - Enabling a user-centred approach to systems development.
 - Providing the underlying concept and technology for interoperable database systems.
 - Designing spatial databases from a perspective beyond the map metaphor that views the real world as independent layers of information that can be combined and overlaid.

Let's put aside for now the spatial database design approach and focus on the first two aspects. As seen in chapter 1 (Roussey *et al.*, 2009), ontologies can be implemented using various markup languages (e.g., RDF¹ or OWL²) and can be managed formalized using specific tools (e.g., Protegé³). Ontologies can also be recorded graphically using entity-relationship or UML diagrams. As stated by (Yeung *et al.* 2007) and deeply discussed by Fonseca *et al.* (2002, 2003), the process of ontology building and documentation is comparable to database conceptual data modelling because both processes aim to identify and define real world features and determine their relationships. However, although the processes are similar, the end products are not the same. While the purpose of a conceptual schema is to describe the intended database structure at a high level of abstraction, an ontology represents a consensual agreement on the meanings of and relations between the vocabulary of terms used to represent data. There is not necessarily direct correspondence between the structure of an ontology and the structure of the database as it is represented by a conceptual database model. This point is illustrated in the case study 2 in section 4.2.

¹ Resource Description Framework (RDF), see (Manola y Miller, 2004)

² Web Ontology Language (OWL), see (McGuinness and van Harmelen, 2004)

³ <http://protege.stanford.edu/>

As discussed in chapter 1 (Roussey *et al.*, 2009), ontologies can be obtained through top-down, bottom-up or middle out approaches. Just recall that a top-down approach builds ontology from upper level ontologies, bottom-up extract ontology from implemented systems, and middle out approach is a combination of the two others. Case study 1 (section 4.1) is an example of top-down approach when case study 2 (section 4.2) is a bottom-up case.

Generally speaking, Ontologies are created by consensus among the experts of data pertaining to a particular domain. These experts are sometimes collectively referred to as an *information community*, using a series of ontology building activities (Auxilio and Nieto 2003). These activities include extraction from existing database schemas (in the case of bottom-up approaches) and a formal data modelling process, called *semantic modelling*, that focuses on identifying and defining relevant terms. In the ontology building processes, it is often necessary to solicit the help of subject matter experts to ensure accuracy and precision of definitions.

Ontology as an approach to database design and implementation serves several useful purposes. The ability of ontologies to provide unambiguous meanings of and structured relationships among the terminology used to describe the real world makes them a useful tool to address the problem of semantic heterogeneity in database design and application, but it is also a crucial medium of communication by providing precise notions that can be used to describe an application domain. It also provides the means to help define the semantics of database fields in a clear and unambiguous manner.

When focusing on semantic heterogeneity and interoperability, the greatest value of ontology is its role in supporting database interoperation strategies by means of query translation and schema integration. Query translation is the process of translating or mapping heterogeneous field names used in different data sets to an ontology in order to query them simultaneously using a single operation, for example by one SQL statement. Schema integration, on the other hand, makes use of the concept of ontology to combine the schemas of individual data sources into one global schema. The next section focuses on the specific roles of ontologies for resolving problems resulting from semantic heterogeneity.

3. Ontologies as a way to solve interoperability issues

This section reviews the state of the art in the use of ontologies in three typical GI interoperability scenarios. Firstly, section 3.1 describes the use of ontologies to help in the discovery and retrieval of GI resources. Secondly, after GI resources are available, section 3.2 presents how ontologies can contribute to solve the problems involved in data integration from heterogeneous sources. Thirdly, section 3.3 describes the role of ontologies as the conceptual model that guides the design and development of information systems in the GI context.

3.1. Ontology-based discovery and retrieval of geographic information.

Discovering and retrieving geographic information is obviously one of the main goals of developing interoperable systems, and by extension of SDIs. It is also crucial to discover suitable geoprocessing services to handle these data. Conventionally, discovery and retrieval for geographic information and geoprocessing services is carried through based on keywords. However, keywords are not always sufficient to find exactly suitable geographic information because they lack semantics, there are ambiguities in natural language and inference mechanisms cannot be applied. The emergence of ontology provides possibility to enhance discovery and retrieval; it solves problems of semantic heterogeneity between user's search and description of geographic information in SDI.

SDIs provide catalogue services for discovering appropriate data and services for a specific task. Searches in these catalogues are currently mainly based on string-matching keywords with metadata entries (Lutz 2005). Keyword-based search can have low recall if different terminology is used and/or low precision if terms are homonymous or because of their limited possibilities to express complex queries (Bernstein and Klein 2002 cited by Lutz 2005). A way to overcome these limitations is to use ontologies to improve matching processes.

For instance, (Bernard et al. 2004) describe the architecture of an ontology based discovery and retrieval system of geographical information. In this system, different Web Feature Services are described with metadata which includes a reference to an application ontology that describes the feature types in terms of a shared domain ontology. The user queries are processed as follows: the user states their queries in terms of the shared

domain ontology; then the system expands the user query restrictions with the names of the stored features. Lutz and Klien (2006) show the evolution of the previous system. This latter version defines a query language and provides a user interface that helps users to formulate queries using a well-known domain vocabulary. In this system, the names of the elements of the Geography Markup Language – GML (see section 3.2) returned by the Web Feature Services are mapped to a shared vocabulary that is used to expand the user queries using a Description Logic reasoner.

Other works in this line are the ones proposed by (Hübner et al. 2004) and (Navarrete 2006). The first one describes an ontology based reasoning system that allows integrating heterogeneous geographical information by resolving structural, syntactic and semantic heterogeneities. The query system supports the specification of queries of the type *concept@location in time*. The user selects a set of registered domain-specific application ontologies (in the thematic, spatial, and temporal domains) based on a common vocabulary and use them to select search terms that are expanded by selecting all equivalences and subconcepts (for the thematic search term), spatially related place names (for the spatial search term), and relevant time periods (for temporal ones). The second one provides a framework to represent semantic relations among the concepts from different datasets of a repository. The system is based on a high level ontology constructed by merging the knowledge provided by the datasets of the repository that describe in a precise and formal way the content of the repository. This ontology is then used to define semantic services or queries that enable agents find and integrate thematic information. It specifically focuses on finding datasets containing information on a particular theme (including theme subclasses if they are considered of interest); translating the content of a dataset to another compatible vocabulary; and integrating heterogeneous content from different datasets.

With respect to the discovery and retrieval of geographic information services, similar approaches based on ontology-based descriptions of queries and service advertisements can be adopted. By using ontologies to enrich services' description, their semantics become machine-interpretable, and users are enabled to pose concise and expressive queries. Furthermore, logical reasoning can be used to discover implicit relationships between search terms and service descriptions. Lutz (2005) proposes ontology-based descriptions of operations consisting of a semantic signature, which contains Description Logics (DL) concepts (instead of datatypes) to represent inputs and outputs, and a specification of pre- and postconditions in First Order Logic (FOL). The operation descriptions and the associated on-

tologies occur at two levels: At the domain level, they describe the generic operations of the domain and thus provide a shared vocabulary (preferentially related to existing standards or agreements within the domain rather than designed from scratch, e.g. the 19100 series of ISO standards published by ISO/TC 211), on which, at the application level, service providers (or requesters) can base the descriptions of (or queries for) a particular operation.

3.2 Data integration in heterogeneous spatial databases

Geographic applications are an example of the need to bring data integration to a big scale. This is the case for the studies of weather, environment, sustained development, terrain use (ground use), mobile applications and more. Semantic understanding is necessary to discover and extract the essential information into a structure suitable for integration from the sources of data. Researchers show the need to focus on a specific domain to achieve the main goal of semantic understanding.

Ontologies define semantics independently of data representation and reflect the relevance of data without accessing them. Such a high-level description of the semantics of geographic information provides more and new means for comparing and integrating spatial data. In addition, ontologies enable knowledge reuse by semantically describing data that were derived from consensus reached by different GIS communities.

Kashyap and Sheth (1996) present a semantic taxonomy to demonstrate semantic similarities between two objects and related this to a structural taxonomy. At present days, intelligent integration has been applied to heterogeneous database integration. From artificial intelligence world often it is achieved by means agents or mediators that provide intermediary services by linking data resources and application programs.

Within the SDI context, several ontologies have been built in last years with the purpose of facilitating integration of data. Some of them are the following:

- *Ontology for Geography Markup Language*⁴ provides an ontology-based representation of the Geography Markup Language (GML) version 3.0 using OWL as ontology language. GML is an OGC specification for

⁴ <http://efe.ege.edu.tr/~unalir/MK/gml30.owl>

the encoding and exchange of GI. The motivation for defining this ontology, developed at the Drexel University in 2004, was to define a core ontology that could be reused and extended in other ontologies for specific application domains.

- *Geospatial Resource Description Framework* (GRDF) (Alam et al. 2008) is another OWL ontology whose concepts and properties extend also the definitions found in GML. The purpose of this ontology is to define an expressive language in the geospatial domain making profit of the advantages provided by Web semantic languages.
- *OntoSensor* (Russomanno et al. 2005) is an ontology based on the IEEE Suggested Upper Merged Ontology (SUMO)⁵, which is a top-level ontology defining general concepts and associations. The purpose of OntoSensor is to provide an ontological perspective of SensorML, the language specified by OGC to represent sensor data collected from remote dispositives. SensorML is also a language derived from GML.

3.3. Ontology-driven Geographic Information Systems

Ontology has been proposed to play a central role in information system's life cycle, leading to ontology-driven information systems (ODIS) (Guarino 1998). In this case the ontology drives all aspects and components of the system. In ODIS the ontology is called application ontology and it is a specialization of a domain ontology and a task ontology (Guarino 1998). The difference between ontology-driven and other types of information systems is that the ontology is made explicit before the information system is even designed. As explained by Fonseca (2007), using an ontology during the development stage enables designers to practice a higher level of knowledge reuse than is usually the case in software engineering. The use of a common vocabulary across heterogeneous software platforms provides for the reuse and sharing of the application domain knowledge. Thus, designers can focus on the structure on the domain instead of being overly concerned with implementation details. Developing and using ontologies should be a prerequisite to conceptual modeling, ontologies being by definition broader than conceptual schemas. At run time, an ontology may enable the communication between software agents or be used to support information integration. Complementary information on ODGIS can be found in Roussey *et al.* (2009).

⁵ <http://www.ontologyportal.org/>

The approach of Fonseca is also connected with a recent approach to software engineering that is called Model Driven Engineering (MDE) or Model Driven Development (MDD). MDD focuses on models as the primary artefact in the development process, with transformations as the primary operation on models. This new approach allows to concentrate the efforts on modelling system functionalities, instead of platform specific details. The successive application of model transformations facilitate the conversion of the original model (based on systems functionality) into a platform-specific application. (Grangel et al. 2007) describe the main issues for the adoption of this MDD approach within the urban domain.

4. Practical case studies

The last section will present three particular case studies on ontology design methodologies and ontology's uses within the GI context. The first two concern a top-down ontology design approach applied to hydrology (*core reference and formal ontology*) and a bottom-up ontology design approach applied in the case of urban spatial database reengineering project (*local and software ontology*), respectively. The third case concerns the use of ontologies for the semantic annotation of geocoding services in the field of urban management systems (*domain and formal ontology*).

4.1. Development of a domain ontology to facilitate interoperability in the context of hydrography

The first case study of the three above mentioned is a project launched by the Spanish National Geographic Institute (IGN-E) to facilitate the semantic harmonization of hydrographic information among data producers at different levels (national, regional and local). IGN-E developed a common reference model by means of a core reference ontology, called *hydrOntology*.

hydrOntology is an ontology that follows a top-down development approach. Its main goal is to harmonize heterogeneous information sources coming from diverse cartographic agencies and other international resources.

Initially, this ontology was created as a local ontology to establish mappings between different IGN-E data sources (feature catalogues, gazetteers, etc.). Its purpose was to serve as a harmonization framework among Spanish cartographic producers. Later, the ontology evolved into a global do-

main ontology, and now it attempts to cover most of the hydrographical features found in a map. The final version of this ontology was finished in the mid-2008.

The statistical data (metrics) and its different taxonomic relations appearing below provide an overview of the *hydrOntology* characteristics.

hydrOntology has 150 classes, 34 object properties, 66 data properties and 256 axioms. Some examples of the four taxonomic relations defined in the Frame Ontology (Farquahr et al. 1997) and the OKBC Ontology (Chaudhri et al. 1998), namely, Subclasses, Disjoint-Decomposition, Exhaustive-Decomposition and Partitions, have been implemented in the ontology. Further details are shown in (Vilches-Blázquez et al. 2007). The ontology documentation is exhaustive, thus, definitions and their definition sources can be found in each concept (class). The ontology has an important amount of labels with alternative names (synonyms) as well as concept and synonym provenances.

In order to develop this ontology following a top-down approach, more than 20 different knowledge models (feature catalogues of IGN-E, the Water Framework European Directive, the Alexandria Digital Library, the UNESCO Thesaurus, Getty Thesaurus, GeoNames, FACC codes, EuroGlobalMap, EuroRegionalMap, EuroGeonames, different Spanish Gazetteers and many others) have been consulted; additionally, some integration problems of geographic information and several structuring criteria (Vilches-Blázquez et al. 2007) have been considered. The aim was to cover most of the existing GI sources for building an exhaustive core reference ontology. Thus, the ontology contains 150 relevant concepts related to hydrography (e.g. river, reservoir, lake, channel, and others).

Regarding methodological issues, *hydrOntology* was built following METHONTOLOGY, a widely-used ontology building methodology. This methodology emphasises the reuse of existing domain and upper-level ontologies and proposes to use, for formalisation purposes, a set of intermediate representations that can be later transformed automatically into different formal languages. A detailed description of this methodology can be found in (Gómez-Pérez et al. 2003).

hydrOntology has been developed according to the ontology design principles proposed by (Gruber 1995) and (Arpírez et al. 1998). Some of its most important characteristics are that the concept names (classes) are sufficiently explanatory and rightly written. Each class groups only one concept and, therefore, classes in brackets and/or with links (“and”, “or”) are avoided. According to some naming conventions, each class is written with a capital letter at the beginning of each word, while object and data properties are written with lower case letters.

With respect to databases, it should be added that this project handles various information databases, both Spanish and European. These databases are created at different scales (from 1:1,000,000 to 1:5,000) and come from diverse institutions or producers. A common component of these databases is that all sources have hydrographical information related to Spanish geographical feature instances.

As commented before, this project handle two European databases (EuroGlobalMap, and EuroRegionalMap), and four Spanish databases that belong to IGN-E. The Spanish databases have information at different scales; of the four Spanish databases, two are Numerical Cartographic Databases (Numerical Topographic Database (BTN25) and Numerical Cartographic Database (BCN200)), and two are gazetteers (Conciso Gazetteer and National Geographic Gazetteer). Finally, with regard to the local databases, the project employs two, one developed by a local producer (Cartographic Institute of Andalusia) and other, by a thematic producer (Hydrographical Confederation of Ebro River).

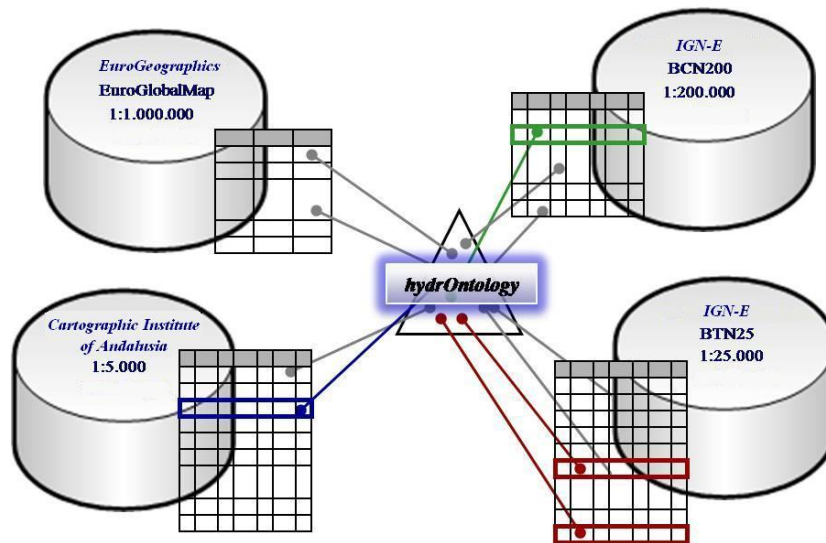


Fig. 1 An overview of wrappers between *hydrOntology* and databases

Within this context of databases, semantic understanding is achieved by setting wrappers between *hydrOntology* and various databases with R2O language (Barrasa et al. 2004). The wrappers, which are still in progress, build and improve relationships between features (from ontology) and instances (from databases). An overview of this work is shown in Fig. 1.

Once *hydrOntology* is consolidated as a harmonization framework for the community of GI producers, the second phase will involve a complex integration framework of databases and ontologies. An overview of this integration approach is shown in Fig. 2. This approach is related to the hybrid approach proposed in (Wache et al. 2001). In the hybrid approach, *hydrOntology* will provide the global shared vocabulary and each producer (European, regional and local) will have a local ontology that sets mappings with the global ontology and its databases. The application of this approach to the Spanish Spatial Data Infrastructure's gazetteer web service⁶ will provide better and richer answers.

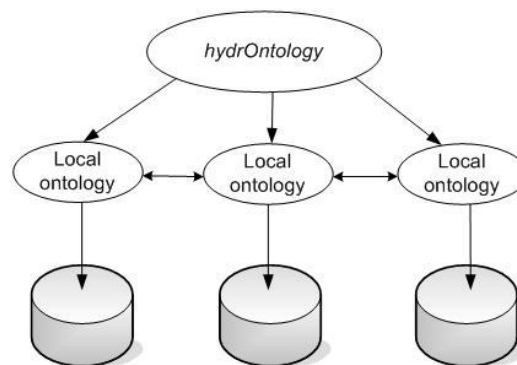


Fig. 2 Hybrid approach of *hydrOntology* (inspired by Wache et al. 2001)

4.2. An ontology extraction bottom-up approach in a spatial database reengineering project

The second case concerns a bottom-up ontology extraction approach within a spatial database reengineering project (Chaidron *et al.* 2007). In Belgium, the *Centre Informatique pour la Région Bruxelloise* (CIRB) manages spatial databases (SDBs) that cover the Brussels Region. This particular set of SDBs is known as Brussels UrbIS 2 ©. At the end of the nineties, it became obvious that a complete reengineering of the databases was needed. A collaboration between the CIRB and the Geomatics Unit of the University of Liege started in 1998 to provide the necessary support to achieve the reengineering process of part of the SDB (the ADM base containing 33 classes and 830000 instances mostly related to geographical

⁶ <http://www.ideo.es/gazetteerIGN/indexLayout.jsp?PAGELANGUAGE=EN>

administrative information), i.e. bringing the DB to its second operational version.

The objective was to create *a posteriori* a feature catalogue and conceptual data models. One of the first step was the (re)-definition of *local software ontologies* of the original database (Fonseca et al. 2003). In order to fulfil project's objective, a bottom-up ontology extraction approach has been adopted. It can be divided in several steps (figure 3):

1. The first step consists in analysing the existing database documentations and then extracting a draft version of the ontologies. Local ontologies can be extracted from data catalogues or data dictionaries and semantic nets can be derived from CDMs (examples of extraction are presented below). The derived ontology should be expressed in an ontology-language like KIF or OWL, or even in UML.
2. At this stage, two options are possible depending on DB designer collaboration.
 - a. The relevance of extracted ontologies can be checked by comparing them to the related populated DB. Final ontologies can be then obtained and the extraction process ends.
 - b. If it is possible, the next step is to submit the draft ontologies to the DB designers. An important issue at this stage is to ensure that both "teams" use the same language, the same concepts. A definition is provided for each concept. This definition includes a textual description as well as a formal expression of its relations with other concepts (ISA, part of and possible topological relations).
3. Remarks formulated by the DB designers must be included in the ontologies extraction process and new ontologies have to be provided until final acceptance.

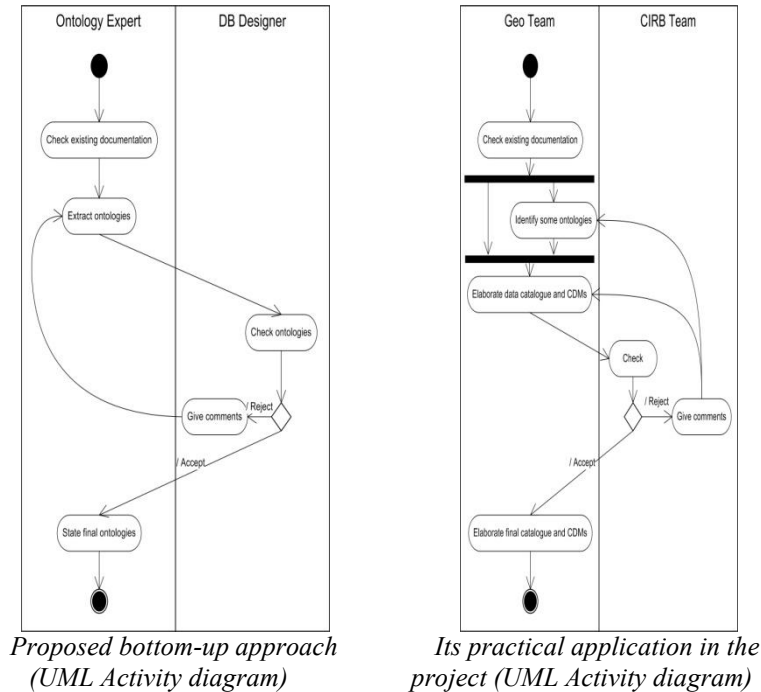


Fig. 3. A theoretical bottom-up approach and its practical application (Chaidron *et al.* 2007)

Some difficulties occurred during practical application of this approach. First, the existing documentation was incomplete and non-standardised; specific relational schemes, a simple data list, data acquisition specifications (for photogrammetric and land surveying measurements). As a result, only some hierarchic and thematic links have been deduced from this documentation. Then, the DB designers failed at the beginning to validate the draft outputs. Tools and methods to formalize their knowledge had to be provided to them and more especially a common *spatial language*. For this purpose we have used first a “natural” language expressed within and Entity / Relationship (E/R) formalism, and later we adopted a more specialized *geo-formalism* (i.e. CONGOO formalism, Chaidron *et al.* 2007).

One of the most important aspects of the submission / acceptance process was the establishment of objects spatial properties: object representation and spatial relationships between objects. By identifying spatial (topological) relationships between objects, this stage revealed object’s definition inconsistencies. It appeared to be the most crucial element of the extraction approach (see Chaidron *et al.* 2007 for further details).

As presented in figure 3, the practical application of the bottom-up approach was slightly different to the theoretical approach as the expected outputs were feature catalogue and CDMs when full documented ontologies were not. Deriving a semantic from the reengineering E/R model is possible. However, such CDM are not ontologies because it has been designed for a specific information system, describing the contents of a specific database, i.e. the specifications of one possible “world” (Bishr and Kuhn 2000, Fonseca *et al.* 2003). That means that we would have to operate an intermediate step to build a kind of semantic net (figure 4); a richer model (global-transposable-sharable) than the database conceptual schema, capturing the semantics of information in a formal way, and usable as a possible way for data integration (Morocho *et al.* 2003).

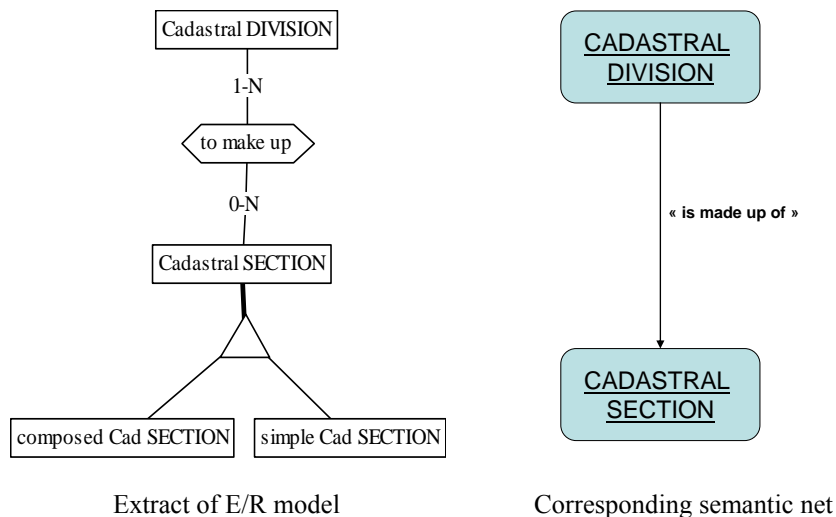


Fig. 4. An extract of Urbis2© E/R model and its corresponding semantic net (Chaidron *et al.* 2007)

This study clarifies the role of ontologies in SDB’s design and reengineering. If the ontology level is necessary for DB’s design (and interoperability) (Frank 1997, Smith and Mark 1998), related ontologies are not always formalized. Therefore, local SDB ontologies are usually hidden in SDBs and associated documentations (feature catalogues and CDMs). In this case, it is possible to extract them from the documentation by applying a bottom-up approach. This process could be improved by a good collaboration with DB’s original designer when the DB is poorly documented.

From our experience, extracting local ontologies (and associated objects definitions) implies a very good knowledge of spatial relationships between DB's objects; we believe that a comprehensive analysis of spatial relationships between instances should be the first stage of local ontologies extraction.

4.3. Enabling geolocating via ontologies

The third case concerns the use of ontologies for the semantic annotation of geocoding services in a system that integrates different geocoding services. This use case, described in detail in (Florczyk et al. 2009), deals with the geocoding of urban addresses using different geocoding services such as a local council geocoding service, a national cadastre geocoding service and a national gazetteer service. Ontologies are used here to solve the semantic heterogeneity between the results retrieved from the different services in terms of address organization.

In Spain, the Zaragoza city council launched its local SDI in 2004 named IDEZar. This SDI has been created in collaboration with the University of Zaragoza (Lopez-Pellicer et al. 2006). IDEZar has as a mandatory requirement the implementation of new geocoding services because many urban related datasets were only georeferenced with street addresses. Two use cases were defined: an on-line geocoder in the SDI web portal to geocode input text addresses and a batch geocoder for large files containing address names.

Urban management systems need geocoding functionality support to enable the assignation of geographic coordinates to location description such as "about 100 meters south of a park, and near a coffee shop". Usually, available geocoding services work on absolute locations and are not appropriate for this kind of task (Hutchinson and Veenendall 2005) and should be enhanced with other services such as a point of interest (POI) service. A system that integrates several geocoding services and other similar would join the functionality of them to provide a location (e.g. the geocoded results of an environment geocoder and a POI geocoder are applied to constrain the query to a third geocoder). However, this approach introduces a high level of complexity in the use of services and data integration. Domain ontologies such as an administrative units ontology (Lopez-Pellicer et al. 2008) might help to fuse the data models. However, the key issue here is the selection of the geocoder services applied to solve the user query.

The service description is composed of a description of the geocoder data model based on domain ontologies, such as an administrative units ontology, and a set of service attributes. (Florczyk et al. 2009) distinguishes the following attributes for the geocoding service description: coverage, content type, spatial object type, result accuracy, reliability, precision and granularity. Some of them are linked to an appropriate ontology. For example, coverage, that defines the data location area, is linked to a concept provided by the appropriate administrative unit ontology.

The architecture of the geocoder integration service consists of the following components (see Fig. 5):

1. The first component consists of an **input data processor** that is responsible for pre-processing of input data that uses the typical geocoding strategies.
2. The **decision maker** is the core component. It hides the process of service selection and the evaluation of the query results.
3. The **mediator** component that contains:
 - a. Pluggable service connector responsible of the invocation to service providers.
 - b. Data integration component that hides the mapping process.

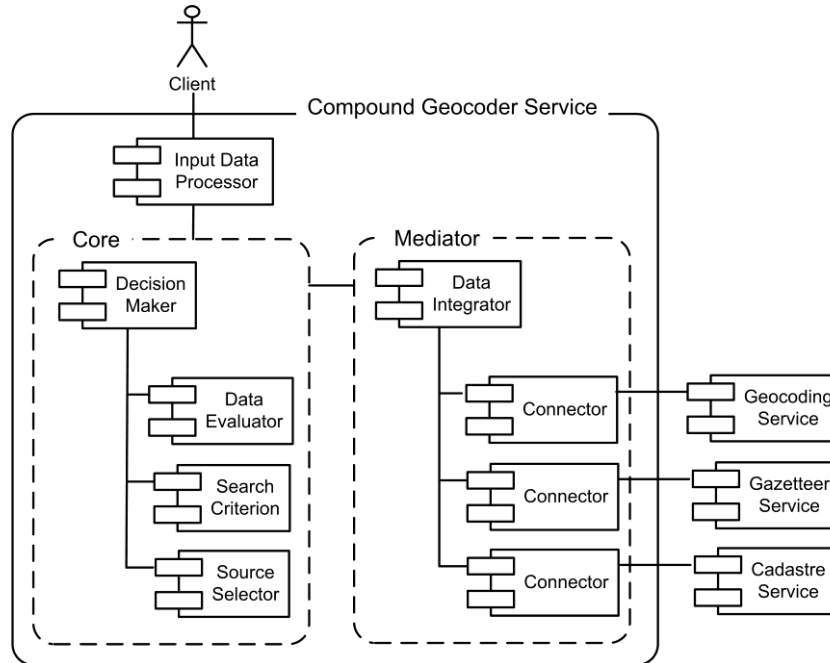


Fig. 5. Compound Geocoder Architecture

The adequate description of each service with the help of domain ontologies determines the behaviour of the whole system mainly because the service characteristics are clues for service selection. For example, the administrative unit ontology plays fundamental role in service selection. This ontology is responsible for defining the relations among the administrative units that provide the basis for source selection according to the correlation between the query constraints and the service coverage. Also, when data from different sources should be integrated in a response, ontologies are applied. For example, the administrative unit ontology helps to build an extensible data model suitable for the representation of the spatial data relationship in the context of administrative units that is used to merge administrative units found in each response.

This approach was applied to the two geocoding use cases defined in IDEZar. The first was implemented as a part of the system in charge of advertising (through the IDEZar web portal) daily incidents on the urban network. The application georeferences input addresses from forms and returns a list of proposals that are visualized on the associated map. The compound geocoder builds the list of proposals according to an internal

ranking based on text comparison, the street type correspondence if available, and, above all, the characteristics of each source geocoding service. The second case is a batch geocoder tool for large CSV (Comma-separated values) files containing address names. The logic of the tool is highly configurable as a result of the characteristics of the chosen architecture.

This experience shows that the usefulness of ontologies in service description and selection in the field of urban management systems. Selecting the best service is a hard task that might be leveraged with the use of service descriptions annotated with semantic descriptions. However, today service descriptions lack these descriptions. Moreover, data and data model behind these services fail to have a semantic description. Defining ontologies and processes to automatically create these descriptions from services should be the first stage of the use of ontologies for integrating services.

5. Conclusions

This chapter has presented roles, types, uses and design processes of ontologies within the Geographic Information sector. One has focused on solving interoperability issues which is especially crucial when dealing with SDIs. The use of ontologies in three typical GI interoperability scenarios have been presented; discovery and retrieval of GI, data integration in heterogeneous spatial databases and development of GIS. In all of these cases, the heterogeneous nature of GI (syntax and semantic) makes the use of ontologies especially important.

Then, three real cases discussing ontology design methodologies and ontology's uses in the GI context have been presented. The first two concerns respectively a top-down ontology design approach applied to hydrology and a bottom-up ontology design approach applied in the case of urban spatial database reengineering project. The third case concerns the use of ontologies for the semantic annotation of geocoding services in the field of urban management systems.

The use of ontologies is growing in the GI community; it is a consequence of development of SDIs and of global services needing various types of GI. Ontologies play a central role in system development, information retrieval and data integration. Knowing that urban information is often of spatial or geographical nature, it is necessary to consider GI ontologies and their uses when approaching urban ontologies context.

6. Open problems and research challenges

Beside general research issues such as the evaluation of ontologies quality (Guarino and Welty 2004), there are some specific research challenges for the GI community (Albrecht et al., 2008; Bucella et al., 2009).

As already mentioned by Mark et al. (2004), there is a need to continue to develop geo-ontologies. Indeed, although the use of ontologies in the GI sector is widely discussed (mainly in academia), there are rather few ontologies on geographic relations and processes. It appears that we are short of ontologies of geographic processes and ontologies are much easier translated into a database schema than into process model. A practical ontology of process that is both proven to be formally correct and at the same time well enough developed to reach to the level of real world applications is still missing (Albrecht et al., 2008). Additionally, a huge work has still to be done to cover concepts such as spatial relations, vagueness or geo-object's changes. It is also rather clear that due to the strong interaction between space and time, spatio-temporal ontologies are a key issue for further model integration.

Further to the need for more geo-ontologies, some authors have also pointed out some technical development needs. For example, Albrecht et al. (2008) raise the issue that moving from static GIS repositories to GIS-based process modelling systems requires the development of reusable libraries of process specifications. They also identify a rather important technical drawback which is that current ontology editors are far from allowing a straightforward connection to GIS; there is usually a long way towards linking original geospatial ontology development with the creation of professional GIS database schemata.

Finally, another future challenge is to ensure integration with other domain ontologies (construction, historical, etc.), which are notably crucial in the urban context.

References

- Alam A, Khan L, Thuraisingham B (2008) Geospatial Resource Description Framework (GRDF) and security constructs. In: IEEE 24th International Conference on Data Engineering Workshop, ICDEW 2008, pp 475-481
- Albrecht J, Derman B, Ramasubramanian L (2008) Geoontology tools: the missing link. *Transaction in GIS*, 12(4), pp 49-424

- Auxilio, M, Nieto, (2003) An overview of ontologies, technical report. Center for research in information and automation technologies, interactive and cooperative technologies Lab, Universidad de las americanas puebla, Mexico
- Arpírez JC, Gómez-Pérez A, Lozano A, Pinto HS (1998) (ONTO)2Agent: An ontology-based WWW broker to select ontologies. In: Gómez-Pérez A, Benjamins RV (eds) ECAI'98 Workshop on Applications of Ontologies and Problem-Solving Methods. Brighton, United Kingdom, pp 16–24
- Barrasa Rodríguez J., Corcho O., Gómez-Pérez A. (2004) R₂O, an extensible and semantically based database-to-ontology mapping language. In: Bussler C, Tannen V (eds) Second Workshop on Semantic Web and Databases (SWDB'04), Toronto (Canada)
- Bernard L, Einspanier U, Haubrock S, Hübner S, Klien E, Kuhn W, Lessing R, Lutz M, Visser U (2004) Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures. Geotechnologien Science Report No. 4
- Bernstein A, Klein M (2002) Towards high-precision service retrieval. In: I. Horrocks and J. Hendler (Eds.), The Semantic Web—First International Semantic Web Conference (ISWC 2002), Springer: Sardinia, Italy, pp. 84–101
- Bishr Y (1998) Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science*, 12(4): 299–314
- Bishr Y, Kuhn W (2000) Ontology-based modelling of geospatial information. In: Third AGILE Conference on Geographic Information Science, Helsinki
- Bucella A, Cechich A, Fillottrani P (2009) Ontology-driven geographic information integration: a survey of current approaches. *Computers & Geosciences*, 35, pp 710–723
- Chaidron C, Billen R, Teller J (2007) Investigating a Bottom-up Approach for Extracting Ontologies from Urban Databases, In: Teller J, Lee J, Roussey C (eds) *Ontologies for Urban Development, Studies in Computational Intelligence*, vol61, Springer-Verlag, Berlin Heidelberg, pp. 131–141
- Chaudhri VK, Farquhar A, Fikes R, Karp PD, Rice JP (1998) Open Knowledge Base Connectivity 2.0.3. Technical Report KSL-98-06, Knowledge Systems Laboratory, Stanford, CA, <http://www.ai.sri.com/okbc/okbc-2-0-3.pdf>
- Cohen W, Ravikumar P, Fienberg SE (2003) A Comparison of String Distance Metrics for Name-Matching Tasks. In: Proc. IJWeb 2003 (IJCAI 2003 Workshop), pp 73–78
- Egenhofer M (1999) Interoperability: Theory and Concepts. In: Goodchild M, Egenhofer M, Fegeas R, Kottman C (eds) *Interoperating Geographic Information Systems*, Kluwer Academic Publishers, Norwell, MA, pp. 1–4
- European Committee for Standardization (CEN) (2006) *Geographic information – standards, specifications, technical reports and guidelines, required to implement spatial data infrastructures*, CEN/TR 15449
- Euzenat J, Shvaiko P (2007) *Ontology Matching*. Springer Berlin Heidelberg New York

- Farquhar A, Fikes R, Rice J (1997) The Ontolingua Server: A Tool for Collaborative Ontology Construction. *International Journal of Human Computer Studies*, 46 (6): 707–727
- Florczyk AJ, Lopez-Pellicer FJ, Rioja R, Nogueras-Iso J, Zarazaga-Soria FJ (2009) Enabling Geolocating via Ontologies. In: *Urban Ontologies for an improved communication in urban development projects (Proceedings of the Final conference of the COST Action C21 – Towntology: Urban Ontologies for an improved communication in urban development projects, Liège, 9&10 March 2009)*, J. Teller, A.F. Cutting-Decelle, R. Billen (Eds.), Les Editions de l’université de Liège, pp 85-94
- Fonseca, F (2007) The Double Role of Ontologies in Information Science Research. *Journal of the American Society for Information Science and Technology*, 58(6): 786-793
- Fonseca, F, Davis C, Câmara G (2003) Bridging ontologies and conceptual schemas in geographic information integration. *GeoInformatica*, 7(4): 355-378.
- Fonseca F, Egenhofer M, Agouris P, Câmara C (2002). Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS* 6(3)
- Frank A (1997) Spatial Ontology. In: *Spatial and Temporal Reasoning*, O. Stock (Ed), Dordrecht: Academic Publisher, pp. 135-53.
- Ganter B, Wille R (1999) *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin-Heidelberg
- Gómez-Pérez A, Fernández-López M, Corcho O (2003) *Ontological Engineering*. Springer-Verlag, London (United Kingdom)
- Grangel R, Metral C, Cutting-Decelle A, Bourey J, Young R (2007) Ontology based communications through model driven tools: feasibility of the MDA Approach in Urban Engineering. In: *Ontologies for Urban Development*, vol. 61 of *Studies in Computational Intelligence*, Springer, pp 181-196
- Gruber TR (1995) Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, vol 43 n.5-6
- Guarino N (1998) Formal Ontology in Information Systems. In: *1st International Conference on Formal Ontology in Information Systems (FOIS’98)*. IOS Press, pp 3-15
- Guarino N, Welty C (2004) An overview of ontoClean. In: Staab S, Studer R (Eds), *Handbook on Ontologies*. Springer, Berlin, Heidelberg, pp 151-172
- Hübner S, Spittel R, Visser U, Vogege T (2004) Ontology-based search for interactive digital maps. *IEEE Intelligent Systems*, 19(3):80-86
- Hutchinson M, Veenendall B (2005). Towards using intelligence to move from geocoding to geolocating. In: *Proceedings of the 7th Annual URISA GIS in Addressing Conference*, Austin, TX
- Kashyap V, Sheth A (1996) Semantic and Schematic Similarities between Object in Databases: A Context-based approach. *VLDB*, 5, 1996, pp. 276—304
- Kolodziej K, ed. (2003) *Open GIS Web Map Server Cookbook*. Version: 1.0.1. Stage: Draft. Open Geospatial Consortium Inc (Open GIS Consortium Inc), OGC Document Number 03-050r1
- López-Pellicer FJ, Álvarez P, Lacasta J, Muro-Medrano PR (2006) IDEZar: procesos, herramientas y modelos urbanos aplicados a la integración de datos

- municipales procedentes de fuentes heterogéneas. In: Avances en las Infraestructuras de Datos Espaciales. Treballs d'informàtica i tecnologia. Castelló de la Plana: Universidad Jaime I de Castellón, pp 105-113
- López-Pellicer FJ, Florczyk AJ, Lacasta J, Zarazaga-Soria FJ, Muro-Medrano PR (2008) Administrative units, an ontological perspective. Lecture Notes in Computer Science (LNCS), vol. 5232: 354-363
- Lutz M, Klien E (2006) Ontology-Based retrieval of geographic information. IJGIS, 20/3: 233-260
- Lutz M (2005) Ontology-based service discovery in spatial data infrastructures. In: Workshop on Geographic Information Retrieval (GIR 2005), Bremen, Germany, ACM Press: New York
- Manola F, Miller E (2004) RDF Primer W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-primer/>
- Mark D, Smith B, Egenhofer M, Hirtle S (2004) Ontological Foundations for Geographic Information Science, A Research Agenda for Geographic Information Science, In: McMaster R, Userly L (Eds.), CRC Press, Boca Raton, FL, pp. 335-350
- McGuinness DL, van Harmelen F (2004) OWL: Web Ontology Language Overview (W3C Recommendation 10 February 2004). <http://www.w3.org/TR/owl-features/>
- McKee L (2000) Who wants a GDI?. In: Groot, R. & J. McLaughlin (eds) Geospatial Data Infrastructure - Concepts, cases, and good practice, New York, Oxford University Press, pp 13-24
- Morocho V, Pérez-Vidal L, Saltor F (2003) Semantic integration on spatial databases. In: Proceeding of VIII Jornadas de Ingenieria del Software y Bases de Datos, Alicante, pp 603-612
- Navarrete A (2006) Semantic integration of thematic geographic information in a multimedia context, Phd Thesis, Universidad Pompeu Fabra
- Nebert D. (ed) (2004) Developing Spatial Data Infrastructures: The SDI Cookbook v.2.0, Global Spatial Data Infrastructure (GSDI). Available from: <http://www.gsdi.org>
- Roussey C, Pinet F, Ah Kang M, Corcho O (2009) How ontologies are used for software interoperability. In:
- Russomanno DJ, Kothari C, Thomas O (2005) Building a Sensor Ontology: A Practical Approach Leveraging ISO and OGC Models. In: Proc. 2005 International Conference on Artificial Intelligence, Las Vegas, NV, pp 637-643
- Smith B, Mark D (1998) Ontology and geographic kinds. In: Proceedings of the Tenth International Symposium on Spatial Data Handling, T. Poiker and N. Chrisman (Eds.), Burnaby: Simon Fraser University, pp. 308-320
- Stumme G, Maedche A (2001) FCA-MERGE: Bottom-up merging of ontologies. In: Proc. 17th IJCAI. Seattle (WA US), pp 225-230
- Vilches-Blázquez LM, Bernabé-Poveda MA, Suárez-Figueroa MC, Gómez-Pérez A, Rodríguez-Pascual AF (2007) Towntology & hydrOntology: Relationship between Urban and Hydrographic Features in the Geographic Information Domain. In: Ontologies for Urban Development. Studies in Computational Intelligence, vol. 61, Springer, pp 73-84

- Wache H, Vögele T, Visser U, Stuckenschmidt H, Schuster G, Neumann H, Hübner S (2001) Ontology-Based Integration of Information — A Survey of Existing Approaches. In: IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, pp 108-117
- Yeung AKW, Hall GB (2007) Spatial Database Systems. Design, Implementation and Project Management. Springer