

AN APPROACH TOWARDS A HARMONIZED FRAMEWORK FOR HYDROGRAPHIC FEATURES DOMAIN

Luis Manuel Vilches Blázquez¹, Antonio F. Rodríguez Pascual², Julio Mezcua Rodríguez³, Miguel Ángel Bernabé Poveda⁴, Oscar Corcho⁵

^{1, 2, 3} Subdirección General de Aplicaciones Geográficas. Instituto Geográfico Nacional.
e-mail: {lmvilches|afrodriguez|jmezcua}@fomento.es

⁴ ETSI Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid.
e-mail: ma.bernabe@upm.es

⁵ School of Computer Science. University of Manchester. e-mail: ocorcho@cs.man.ac.uk

Abstract

We describe an ontology in the hydrographic domain, *hydrOntology*, which can be used to overcome the semantic heterogeneity of different databases and other information sources that deal with geographic information. We show that ontologies are richer than other means commonly used to represent geographic information, such as feature catalogues and thesauri, and how ontologies could be used to overcome some of the problems associated with them.

1. Introduction

Nowadays, Geographic Information (GI) is increasingly captured, managed and updated by different cartographic agencies with variable levels of granularity, quality and structure. This approach causes in practice the building up of multiple sets of spatial databases with a great heterogeneity of feature catalogues and data models. That means a coexistence of a great variety of sources with different information, structure and semantic without a general harmonization framework. This heterogeneity, combined with the sharing needs of miscellaneous users and information overlaps from different sources, cause several and important problems to link similar features, to search, to retrieve and to exploit GI data.

Ontologies are probably the most appropriate instruments to capture and share the semantics of any domain, including that of geospatial information. Ontologies are defined as “formal and explicit specifications of a shared conceptualisation” [6], and define a “set of relations between terms, as well as the rules that combine terms and relations and improve the definitions provided in the vocabulary” [7]. Ontologies can prove to be the solution to the aforementioned problems, not only by improving the structure of the world of classical cartography, computer-assisted cartography, GIS (Geographic Information System) and SDI

(Spatial Data Infrastructure), but also by providing support for the shared use of agreed names, codes, and attributes that reflect this world.

In this paper we focus on *HydrOntology*, a domain ontology of hydrographic features, which is a starting point to relate different feature catalogues corresponding to the geospatial databases generated by diverse cartographic agencies in Spain at national, regional and local levels, from 1:1.000.000 national scale to 1:1000 local scale. Moreover, hydrographic features have relationships with other knowledge areas, such as the legal framework (international law), the geological domain (hydrogeology) and urban civil engineering (COST UCE Action C21, *Towntology* project).

The paper is organised as follows. In section 2 we describe the current structuring forms of GI. In section 3 we describe some of the main problems related to the use of geospatial data. The main benefits of applying an ontology-based approach, and more specifically of the use of *hydrOntology*, are commented in section 4. Finally, in section 5, conclusions and future research lines are indicated.

2. Current organisation of Geographical Information

The Open GeoSpatial Consortium (OGC) [16] says that a geographic feature is the starting point for modelling geospatial information. For that reason, the basic unit of GI within most models is the ‘feature’, an abstraction of a real world phenomenon associated with a location relative to the Earth, about which data are collected, maintained, and disseminated [11]. Features can include representations of a wide range of phenomena that can be located in time and space such as buildings, towns and villages or a geometric network, geo-referenced image, pixel, or thematic layer. This means that traditionally a feature encapsulates all that a given domain considers about a single geographic phenomenon in one entity [15].

Currently, the most common way to structure geographical feature names, codes, attributes and any other characteristics, including their graphical representation, is the use of feature catalogues [11]. They consist of a list of geographical features grouped into various feature classes that have instances with common characteristics (e.g., river, church, road, etc). Each geographical feature is identified by a unique code, which represents a discrete phenomenon that is associated with its geographic and temporal coordinates and may be portrayed by a particular graphic symbol [11] (e.g., the Burgos cathedral, the Teide vulcano or the city of Madrid). On the other hand, sometimes these features have definitions and different attributes.

Another common way to organise geographical information is the use of thesauri. However, these are not commonly used by Spanish GI providers. A thesaurus is a controlled vocabulary that contains natural language terms, which are used to represent briefly document areas. Thesauri are formally organised in order to explicit relationships between terms. For instance, “more generic than” or “more specific than” [9, 10]. Thesauri do not formalise as much information as feature catalogues, but they provide a better organisation for feature classes and instances, since they describe relationships between instances and at the same time avoid ambiguity and lack of precision in the vocabulary used to describe features.

3. Some problems of current Geographic Information

Nowadays, all GI is setup, developed, analyzed and stored at different levels of detail and by different kind of cartography producers. As a result, available geographical data has heterogeneous content and structure, derived from the lack of consensus and from independent provisioning processes, provoking the problems described below.

As aforementioned, the current forms of geographic information representation (feature catalogues and thesauri) are not structured in an optimal way, and are usually limited to simple hierarchical representations, sometimes determined by the identifier of each feature. Besides, the simplicity (and sometimes lack) of relations between terms has a negative impact on the efficiency of searches, accesses and information retrieval. Finally, the formal semantics of these representation forms is not clearly defined, what makes it difficult to use them for the semantic integration of a set of geographic information sources.

3.1. Vocabulary heterogeneity and Databases

In OGC term features are not fixed in their class but have application-oriented views that are classed [16], i.e. depending on the domain classification, a feature instance may be classified one way or another. Therefore, it is apparent that features are not the atomic units of GI as the phenomena they represent, encapsulating different human concepts resulting in multiple types [15]. It is recognized that geospatial information is subjectively perceived and that its content depends upon the needs of particular applications. The needs of particular applications determine the way instances are grouped into types within a particular classification scheme, so that the same geographic features are usually abstracted differently from a feature catalogue to another. Hence different cartography producers develop spatial data with different points of

view, interest and necessity, using different vocabularies, organization forms and models. It causes.

This heterogeneity poses difficulties for the retrieval and integration of information from different sources, since different databases may provide different answers for a given query (since they may use different terms for the same phenomenon).

Another common problem in the access to geographical databases is related to a poor implementation model. Some databases have only one table where they store all information. Some others are organised in different tables without relationships between them and without explicit primary and foreign keys.

3.2. Scale factor

Scale factor acts as a filter in the cartographic representation such as catalogues and dictionaries in the geospatial information. For this reason, it is very important to consider information at several scales (local, regional and national), what affects geometric and semantic resolutions. Moreover, the existence of different phenomena at different scale factors is an added problem in the generalisation of cartographic information, due to the difficulties found in the overlaps of features from different catalogues.

3.3. Language problems and semantics differences

Problems related to language ambiguity are also related to the previous heterogeneity. Language problems refer to polysemy, synonymy, hyperonymy and homonymy, and are present in many concepts in geographic information, due to that fact that there is not a harmonized semantic framework. As a usual case, we can find different feature catalogues with various concepts associate to the same feature (e.g. river, water flow, waterway).

Existing semantic differences in GI domain are numerous. A repetitive example is the river definition. The Water Framework Directive (WFD) defines a river feature as “*a body of inland water flowing for the most part on the surface of the land but which may flow underground for part of its course*” [17], while the Ordnance Survey defines it as “*water flowing in a definite channel towards the sea, a lake or into another river*” [5]. The IGN-E considered the river a “*natural freshwater stream*”. Nowadays, the IGN-E has decided to adopt the WFD proposal because it is a continuous phenomenon, although it would lack a cartographic representation when the flowing occurs underground.

These problems mainly arise because each producer community is typically focused on specific needs [14], which means that a semantic harmonisation between the different producers has not been achieved, in spite of there is a increasingly need to share, to exchange, and to use for different purposes geographic datasets.

In summary, these problems show the difficulties of achieving semantic interoperability¹ in the context of geographic information, and consequently the difficulties in the tasks of querying, retrieval, exploitation, update and visualisation of GI, when users require simplicity and efficiency.

4. *hydrOntology*: A harmonized framework for hydrographic features domain

For the time being, International standards promoted by organisations like ISO and OGC enable users to access and process geographic data from a variety of sources across a generic computing interface within an open information technology environment. This has allowed achieving syntactic interoperability between GI systems, but does not solve any of the aforementioned problems.

As described in the introduction, ontologies can simplify the tasks of GI access and distributed search in different GI databases, by easing accessibility and providing a common structure of the GI data [8]. Existing ontologies can be used as the starting point for the development of new ones (hence improving knowledge reuse). Besides, mappings between different ontologies can be used to describe the correspondences between phenomena (e.g., *rio*, *river*, *rivière* and *fleuve*), and mappings between these ontologies and databases can be also used to describe further correspondences.

hydrOntology characteristics

The purpose of building *hydrOntology* is to serve as a harmonization framework among the Spanish cartographic producers. However, we also aim at making this model available for other international GI producers. With this ontology we intend to provide the necessary steps to obtain a better organization and management of the hydrological features, which are spread over into the different projects, documents and directives in this field.

The development of this ontology has been based on METHONTOLOGY [2]. This methodology enables the development of ontologies at the knowledge level, and has its roots

¹ Semantic interoperability assures that the content is understood in the same way in both systems, including by those humans interacting with the systems in a given context [1]

in the main activities identified by the IEEE software development process [3] and in other knowledge engineering methodologies.

Information sources of hydrOntology

In the development of *hydrOntology* we have considered information sources at several scales (local, regional and national), mainly feature catalogues and data dictionaries (Numerical Cartographic Databases of the Instituto Geográfico Nacional of Spain (IGN-E), catalogues and data dictionaries from other local Spanish cartographic agencies, EuroGlobalMap and EuroRegionalMap), thesauri (UNESCO, GEMET, Getty TGN, Feature Type Thesaurus of Alexandria Digital Library), the project SDIGER [12], different gazetteers (National Geographic Gazetteer of IGN-E, Conciso Geographical Gazetteer of Spain, Digital Alexandria Library Gazetteer, etc.), trying to cover the greatest amount of IG sources, in order to build a complete domain ontology. This glossary contains more than 100 relevant concepts related to hydrography as river, reservoir, lake, channel, pipe, water tank, siphon, etc.

Furthermore, in the *hydrOntology* development we have taken into account some concepts about feature capture that depend exclusively on different Spanish geographic regions. Among these features appear “ibón”, “lavajo”, “chortal”, “bodón” and “lucio”. These concepts are designated by their local name and they are synonymous to the feature “Charca”²

Due to the diversity in semantic concepts within the domain, the definition of the characteristics and the context has been restricted, adapting it to the topographic database, as the Numerical Cartographic Database of IGN-E. Every definition will take the cartographic representation into account through map, GIS or SDI, no matter what the intrinsic reality of these phenomena is.

Modelling criteria

The criteria for this ontology structuring are governed mainly by four criteria [8]:

1. The European Directive to set up *a communitarian frame of performance in the scope of the water policy (WFD)* [17]. That contributes to the modelling of more abstract features that make up the *hydrOntology*.
2. On the other hand, as a consequence of the aim of implementation of this ontology in the Spanish Spatial Data Infrastructure (IDEE), we are taking into consideration the classification

² “Charca” is a small lake of shallow water. The above mentioned terms are Spanish local names.

worked out by the SDIGER Project³ [12]. That project was chosen by Eurostat⁴ as a pilot project of the applicability of INSPIRE [4].

3. Being aware of the importance of the establishment of a taxonomical order, several semantic criteria have been added. Thus the hydrographic feature classification is in accordance with the meaning of each feature.

4. Finally, an important matter should be added to those criteria, namely the presence of the inheritance of different sources in the modelling of this ontology, on the one hand to facilitate the possible information mapping and on the other to be consequent with the hierarchy of the features carried out by the expert in the domain.

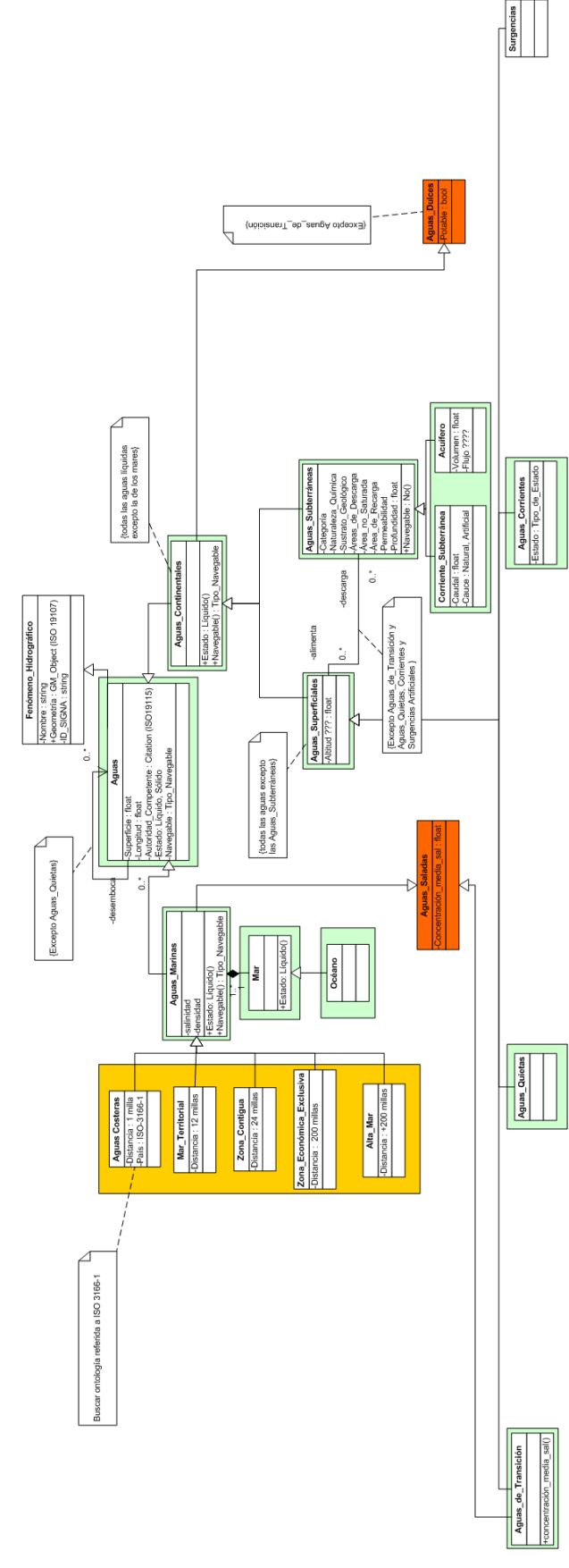
hydrOntology model

In this section we show a graphical overview of the upper level of hydrOntology (see below), so as to show the degree of completeness of the ontology.

This ontology is divided into two levels: the upper level, which contains the most abstract phenomena in the ontology, and the lower level, where a set of well-known hydrographic phenomena are described. The upper level contains the concept “Hydrographical Feature”, with more specialised concepts like “Inland Waters” and “Sea Waters”. There is a different degree of specialisation in each of these concepts, since the current focus of this ontology is on “Inland Waters”. These are divided, following WFD, into “Superficial Waters”; with the following classes: “Transitional waters”, “Stand Waters”, “Flowing Waters” and “Sources”; and “Groundwaters”. For each of these classes we have identified concepts in the lower level, where a detailed set of hydrographic phenomena is provided.

³ An SDI created to support the access to GI resources concerned with the WFD within an inter-administration and cross-border scenario that involves two countries, France and Spain as well as the two main river basin districts on both sides of the border, the Adour-Garonne basin district, managed by the Water Agency for the Adour-Garonne River Basins (*L'Agence de l'Eau Adour-Garonne*) and the Ebro river basin district, managed by the Ebro River Basin Authority (*Confederación Hidrográfica del Ebro*) Available on <http://www.ideo.es/sdiger/>

⁴http://epp.eurostat.cec.eu.int/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL



Moreover, several features are considered which will be used to relate to other domains, such as the legal framework (international law). Concepts like “territorial waters”, “contiguous zone”, “high seas”, etc., or the geological domain (hydrogeology) “underground currents”, “aquifers”, etc., or the urban civil engineering (COST UCE Action C21, *Towntology* project) “pipe”, “reservoir”, “channel”, etc., may be considered as an example. This fact enriches the domain ontology because it provides a gradual increase in knowledge and added benefit for GI users on the web.

5. Conclusions and future work

The diversity of problems that arise in the management of Geographical Information, together with the need for integrating information available in different sources, make it necessary to work in order to achieve semantic interoperability. Ontologies are used as instruments to deal with heterogeneity problems [13]. We have created hydrOntology in order to solve structural problems of geographical information in the context of hydrographic information, so as to provide a common and shared view of different databases, as well as a clear structure of the domain. As part of our future work, we will work on providing mappings of this ontology with other databases, on providing multilingual support for the ontology, and on working on relating this ontology with ontologies from other domains.

6 References

- [1] ISO (2002) ISO/DIS 19119 Geographic Information Services. Geneva.
- [2] Fernández-López M, Gómez-Pérez A, Juristo N (1997) METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, pp 33–40
- [3] IEEE (1996) IEEE Standard for Developing Software Life Cycle Processes. IEEE Computer Society. New York. IEEE Std 1074-1995.
- [4] The European Parliament and The Council of the European Union (2007) Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).
- [5] Ordnance Survey (2001): Master Map real-world object catalogue.
<http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/faqs/Docs/realWorldObjectCatalogue.pdf>

- [6] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications". Knowledge Acquisition, 1993, 5(2), pag.199-220.
- [7] Neches et al., "Enabling technology for knowledge sharing". AI Magazine, 1991, 12(3), pag.16-56.
- [8] Vilches-Blazquez, L.M., Angel Bernabe-Poveda, M., Suarez-Figueroa, M.C., Gomez-Pérez, A., Rodríguez-Pascual, A.F. (2007) Towntology & hydrOntology: Relationship between Urban and Hydrographic Features in the Geographic Information Domain. In: Ontologies for Urban Development: Interfacing Urban Information Systems. Volume 61 of Studies in Computational Intelligence. Springer 73-84.
- [9] ISO 5964 (1985) Guidelines for the establishment and development of multilingual thesauri.
- [10] ISO 2788 (1986) Guidelines for the establishment and development of monolingual thesauri
- [11] ISO 19110 (2005) Geographic information—Methodology for feature cataloguing
- [12] M.A.Latre, F.J.Zarazaga-Soria, J.Nogueras-Iso, R. Béjar, P.R.Muro-Medrano. SDIGER: A cross-border inter-administration SDI to support WFD information access for Adour-Garonne and Ebro River Basins. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework. 2005. pag 5-7. Alghero, Sardinia (Italia).
- [13] Bucella, A.; Cechich, A. (2007) "Towards Integration of Geographic Information Systems" ScienceDirect. Electronic Notes in Theoretical Computer Science 168, 45-59.
- [14] Bermudez L, Piasecki M. (2004) Role of Ontologies in Creating Hydrologic Metadata. International Conference on HydroScience and Engineering, Brisbane, Australia.
- [15] Greenwood J, Hart G. (2003) Sharing Feature Based Geographic Information - A Data Model Perspective. 7th Int'l Conference on GeoComputation. United Kingdom.
- [16] OGC (2003) OpenGIS Reference Model, Version 0.1.2, OGC Inc. Wayland, MA, USA.
- [17] European Parliament and the Council of the European Union (2000) Directive 2000/60/EC of the European Parliament and of the Council for establishing a framework for Community action in the field of water policy. Brussels, p 72.