A Semantic Geodatabase for Environment Analysis - Extraction, Management and Sharing of Earth and Water Information in GIS.

A Semantic Geodatabase for Environment Analysis
Extraction, Management and Sharing of Earth and Water Information in GIS

Andrea Lingua and Francesca Noardo
Politecnico di Torino, DIATI, c.so Duca degli Abruzzi, 24, 10129 Torino, Italy
{andrea.lingua, francesca.noardo}@polito.it

Keywords: Geographic Information System, Geoprocessing, Ontology, Digital Maps, Geodatabase.

Abstract: The great potential of GIS to manage and analyse georeferenced information is well-known. The last several years of development of ICT (Information and Communication Technologies) saw a necessity of interoperability arise, from which the Semantic web standards and domain ontologies are derived. Specific application field ontologies are often insufficient for representing the information of multidisciplinary projects. Moreover, they are often aimed at the representation of homogeneous data formats (alphanumeric data, vector spatial data, raster spatial data, etc.). In this scenario, traditional GIS often have a limit: they implement personal data models, which are very difficult to exchange through different systems. In this study we structured a GIS for the monitoring project ALCOTRA ALIRHYS according to parts of two different self-integrated ontologies, from the perspective of the major interoperability of the system and the sharing of data through a web-GIS platform. The two standard models chosen (SWEET ontology and INSPIRE UML model) have been integrated in a unique conceptual model useful both for geometric and cartographic data, and for thematic information. In this case, the implemented schemas are published on the project website, and are available for other users who want to produce similar studies. Since user-friendly results were desirable, some integrated commercial widespread software programs have been used even if their abilities to manage such a GIS are suboptimal.

1 INTRODUCTION

Traditionally, Geographic Information Systems (GIS) are powerful instruments for storing geospatial data in digital archives and managing these data for inferring further knowledge, extracting spatial information and realizing geometric and morphological analysis. Another basic capability of GIS is the archiving of dynamic data related to monitored values in a "many to one" relation with reference to spatial objects. These can be queried and represented on the land, with additional consideration of the time at which the data have been acquired (the 4th dimension managed by GIS).

Since the mid-1990s some research has further developed these systems. In particular, a need for interoperability arose; the solution was found in ontologies, in order to better manage the semantic content of spatial data (Freksa and Barkowsky, 1996; Fonseca et al., 2002), and in object-oriented GIS (Scholl and Voisard, 1992), which permitted the implementation of some important characteristics of the semantic structure (Mennis, 2003). Today, the achievement of the evolution of the World Wide Web, the Semantic Web, offers theories and structures that make the theorized interoperability a reality (Waters et al., 2009). Standard languages and services foster this aim: OGC (Open Geospatial Consortium) published standards for spatial data exchange, such as WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), WPS (Web Processing Services) and so on. Moreover, the semantic web offers more standards and tools aimed at reaching maximum data integration and exchangeability. Structuring data by means of these tools offers an enhanced ability to run automated reasoning and evolved queries on published spatial data. In several fields of application some standard ontologies have been developed, as references for the management of the semantic content of represented data (Pundt and Bishr, 2002). However, these ontologies often seek to represent alphanumeric data, without concern for spatial implications. In the meantime, other ontologies, such as CityGML (OGC, 2008) or INSPIRE UML (Unified Modelling Language) model

Lingua A. and Noardo F..
A Semantic Geodatabase for Environment Analysis - Extraction, Management and Sharing of Earth and Water Information in GIS.
DOI: 10.5220/0005379002130220
In *Proceedings of the 1st International Conference on Geographical Information Systems Theory, Applications and Management* (GISTAM-2015), pages 213-220
ISBN: 978-989-758-099-4

(http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/), consider the cartography, but are insufficiently able to manage the data of particular application fields in a GIS. For this reason the need for ontology integration arises.

In this paper, the environmental data (geometric and dynamic data) studied during the European project ALCOTRA ALIRHYS have been managed in a GIS and in an external database structured according to a data model designed by integrating existing ontologies. The project aims to study subterranean water resources through several monitoring activities (www.polito.it/alirhys). In this way the data produced are codified to be uniquely interpreted, and to be effectively shareable on the web without losing part of their meaning. Moreover, when a GIS or a database is constructed in the traditional way, the conceptual model used is shared, together with other metadata, in order to limit interpretation ambiguity. However, sharing of implemented structures is not common. In this project, both the models and the XML schema of the constructed geodatabase in ESRI ArcGIS and the related database tables in MS Access are published on the project's website (www.polito.it/alirhys) in order to foster the reproduction and the reuse of the method for similar studies, even in a stand-alone environment.

The advantages of data management in GIS have been exploited by running different analyses on the data archived in order to acquire further information, useful for the monitoring of springs; this work is described in the final part of this paper.

1.1 Semantic Data Management: Ontologies and Standards

Representation in many fields of application has been influenced in recent years by new paradigms established by the web and informatics technologies. The Semantic Web is intended to realize a common framework that allows the meaningful content of data to be shared, beyond the boundaries of applications, and to be easily reused for automatic processing. Permitting an effective semantic interoperability, documents are associated to metadata that specify the semantic context in standardized language, and data are structured according to pre-formatted and known data models. One of the most powerful tools for reaching these objectives is the formulation of domain ontologies, primary conceptual models that provide completely solution-independent schemas in order to support modelling in producing suitable representation structures. A set of standard languages is available for defining interoperable ontologies (http://www.w3. org/). Almost all these languages have as their origin structure the eXtensible Markup Language (XML), which is then specified in a set of more evolved structures developed to better express the logic structure and correcting semantics of data.

W3C (World Wide Web Consortium) is the organization developing the structures to enable the Semantic web. It published some specific languages to model the data in domain ontologies, in order for the data to be effectively shared. These evolve from the original RDF (Resource Description Framework) (http://www.w3.org/RDF/) to more suitable ones for representing knowledge structures: OWL (Ontology Web Language) (http://www.w3.org/OWL/), and some evolutions of it (OWL 2, DAML+OIL, OWL1 DL). These languages are best suited for representing information in the form of alpha-numeric data, but they have some lacks in the suitable management of geometric information.

In this scenario, many domain ontologies are being developed. Among these, we find the NASA (National Aeronautics and Space Administration) SWEET (Semantic Web for Earth and Environmental Terminology) ontology, intended for the representation of Earth and Environmental Sciences (EES) (Raskin, 2004). This ontology is in modelled in the OWL language, and includes a very high variety of concepts about both the object studied and the methodology aspects (research, analysis, measurements, residuals, etc.); these could be effectively transposed to other application fields. SWEET has become the *de facto* standard for data management in EES (Di Giuseppe, 2014).

Concurrently, some particular formal languages have been developed for spatial data. The most widespread of these are published by the OGC (Open Geospatial Consortium), an organization founded to develop publicly available geo-enabled standards. In particular, it encoded GML (Geography Markup Language, ISO 19100 series of International Standards and OpenGIS Abstract Specification), which is an XML schema for the description of the application, transport and storage of geographic information. This has been applied to the formulation of some standard models - such as CityGML, the common semantic information model for the representation of 3D urban objects – and allows these data to be shared over different applications. Another fundamental application of GML is the INSPIRE conceptual model (http://inspire-twg.jrc.ec.europa.eu/data-model/draft/r4530/), which is the pre-specified ontology for formulating harmonic and homogeneous maps in the European Union.

A point of contact between these two fields can be found in considering the W3C language SPARQL (SPARQL Protocol and RDF Query Language) (http://www.w3.org/TR/rdf-sparql-query/), a query language for RDF structured data. Starting from this language, OGC developed the geoSPARQL language (http://www.opengeospatial.org/standards/geosparql); it defined a vocabulary for representing geospatial data in RDF, as well as an extension to SPARQL for processing geospatial data. This language offers the ability to transition between the languages GML and RDFS/OWL. In the future, this will provide a good solution for managing and sharing data, as some research has already outlined (Tschirner, 2011).

In this study, the available entities have been extracted from the SWEET ontology in order to model the environmental objects studied during the project ALIRHYS. Concurrently, the spatial properties of objects have been managed by integrating the selected part of the SWEET ontology with the entities extracted from the INSPIRE UML model, to complete the representation for suitable management in the GIS.

1.2 INSPIRE

The INSPIRE Directive (http://inspire.jrc.ec.europa.eu/) is a European Union directive aimed at creating a European spatial data infrastructure, able to establish a unique network for sharing environmental spatial information among European nations. The directive has been in force since 15th May 2007, and after passing through various stages will be fully implemented by 2019.

Two of the main principles of INSPIRE are: “It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications; […it should be] easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.” (INSPIRE, 2007) As in some other works related to cartography harmonisation (Tóth; 2007; Afflerbach, 2004) and also in the ALIRHYS project, our purpose was to respect these two principles by producing some cartographic data and sharing them on a WebGIS together with every useful metadata file.

2 THE EUROPEAN PROJECT ALCOTRA –ALIRHYS

The European project ALCOTRA – ALIRHYS (2013-2015) has been developed by the partnership among the Politecnico di Torino, Politech Nice Sophia, Université Nice Sophia - Antipolis, Regione Piemonte and the Nice - Côte d’Azur Metropole, with the aim of studying and monitoring the subterranean hydric resources in the cross-border mountainous territory between the provinces of Cuneo and Nice (Fig.1). These activities are required in light of climatic changes that have occurred over the last several years and that increase the risks caused by climate variability: periods of drought alternated with flooding are becoming more and more frequent. Subterranean hydric resources feed several springs in the territory from which the local hydrographic network originates, and they also feed several aqueducts that supply drinking water to users. For this reason, it is essential to know the chemical composition of the water and to foresee the influence of rainfall and snow fusion on the behaviour of the springs in order to optimize their management (Bellone et al., 2014). These parameters can only be studied from a cross-border point of view as the geological assets influence both sides of the Alps. Therefore, studies must be carried out in a comprehensive manner. Furthermore, this cross-boundary case study justifies the use of standards for the management of the project’s data; these standards ensure that data can be shared, queried and used in analysis.

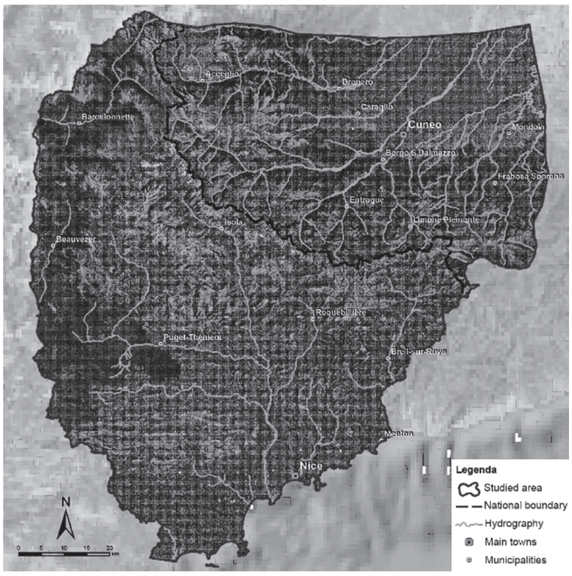

Figure 1: Study Area of the project ALIRHYS.

In this article, the part of the project regarding the management, analysis and integration of data in GIS is presented.

3 APPLICATIONS OF THE SEMANTIC MODEL IN GIS FOR ALIRHYS PROJECT

The data of the springs monitoring project ALIRHYS have been managed in a GIS. As usual, a conceptual schema for modelling the data is used, but particular care was taken in choosing entities useful for the specific needs of the study and appropriate for the two reference ontologies: the SWEET ontology and the INSPIRE UML model. The resulting conceptual model is shown in Figure 2. It shows the entities present in the INSPIRE conceptual model (in blue), which have been used mainly for the harmonisation of digital maps (Noardo et al., 2015); the entities extracted from the SWEET ontology (bordered in red) are used to manage the remaining concepts present in the system. These last ones and some additional entities are added because the needs of the project are in turn divided in spatial entities (in yellow), dynamic data tables (in violet), and geoprocessing products (in pink). The integration between the two ontologies is indicated by the green arrows.

In the implementation, different logic data models are useful for different data management requirements. It would be suitable to manage spatial data in an object-oriented database, as is also implicit in the structure of standard models. This kind of logic model enables some meaningful constructs for cartographic objects, such as polymorphism and inheritance (Worboys, 2004). On the other hand, reams of dynamic data can be managed effectively in a relational database, in which they can be automatically imported and suitably queried. Software implementing the hybrid model, an "object-relational database management system" (ORDMBS), could have good performance in both cases (examples of these are PostgreSQL and Oracle). However, these system are often less widespread than others that use a relational database model, such as ESRI ArcGIS or MS Access; these are more commonly known and, consequently, more user-friendly. In this situation ease of use is important, so these last two software programs have been chosen to permit an easy re-employment of the data and of the schemas implemented by without advanced qualifications.

Since the implementation with this kind of software is not the most ideal, the model has been divided in three main parts. The first segment concerns the harmonisation of digital maps (ESRI ArcGIS); the second manages the representation of useful georeferenced data of the project (ESRI ArcGIS), and the third deals with dynamic data tables, which are managed in an external DBMS (Database management system), MS Access.

Ontologies and data models are oriented towards the publication of data on the web and the sharing of these data through specific web service interfaces. During this project the aim was to conduct some analysis on the springs and on the studied area, even in a stand-alone environment. Only in a final part of the project have results been published in a WebGIS using the open-source platform Geonode. The requirement of respecting defined data models was useful for obtaining structured data in this case; it could potentially be implemented in other similar systems, or shared on the web to enhance the system's functionality for future work.

3.1 The ALIRHYS Geodata

3.1.1 The Harmonisation of Digital Maps

As a first step in building a geographic database for the project map, GIS tools were used to harmonize the available national cartographic products by exploiting both their geoprocessing capabilities and their database characteristics (as better explained in Noardo et al., 2015).

In this paper we focus on the part of the harmonisation processes concerning the digital maps. Analysing the national digital maps, one notices some differences in the geometric visualisation of objects due to the origin of the data, the plotting methods, and the map's nominal scale. These geometric differences are too difficult to solve, because doing so would require the re-plotting of maps or the acquisition of new homogenous data for the whole area; this is not within the scope of this project. What is interesting in this context are the data structures, which have been analysed and transformed in order to make them harmonic. To pursue this aim the part of the defined conceptual model including INSPIRE entities has been used as a reference.

The national databases were analysed in order to extract a simplified version of the conceptual model structuring the maps, considering only entities useful for our representation needs. The next step was the mapping of each entity into the selected part of the INSPIRE model, using a transformation to make the data homogeneous (Noardo et al., 2015).

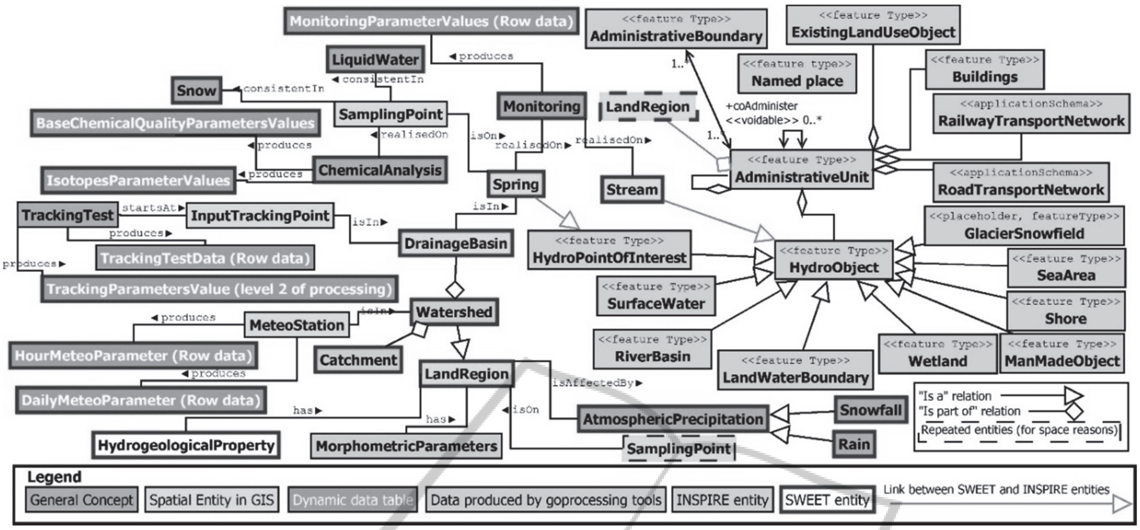

Figure 2: Conceptual reference model used for the GIS.

3.1.2 The Construction of an Interoperable Geodatabase for ALIRHYS Data

On the harmonised base maps some original data were produced. These have been mainly structured following the part of the model designed on the SWEET ontology. Entities with spatial consistency have been archived as feature classes in a geodatabase, while the entities mainly represented by dynamic data tables are managed in a related external database.

Using ESRI tools, a geodatabase is created in ArcGIS. This is useful in order to define the static entities of the system, to establish relationships and topological rules between them and to impose constraints including taxonomies and subtype definition (tables 1-3).

Table 1: Entities (feature classes) of the geodatabase.

Springs	Points
MeteoStations	Points
SamplingPoints	Points
InputTrackingPoints	Points
Stream	Points
LandRegion with HydrogeologicalProperties	Polygon
DrainageBasin	Polygon

Table 2: Relationship classes of the geodatabase.

S_Comparison_MS	Relation Springs - MeteoStations 1-n relation
S_Comparison_SP	Relation Springs - SamplingPoints n-n relation
S_Comparison_SS	Relation Springs - Streams n-1 relation

Through these associations it is possible to query and access the data tables in a cross-referenced way for easy comparison.

Table 3: Topology rules stated in the geodatabase.

SamplingPoints	Must be Properly Inside - LandRegion
Spring	Must be Properly Inside – DrainageBasin

The external database is defined in MS Access for the representation of the dynamic entities in Table 4 (for which extensive records are needed). These are represented in violet in the reference conceptual model. Once the tables are related to the geodatabase feature classes, they can be queried; it is then possible to carry out analysis and statistics on them while also considering their spatial reference (Fig.3). The updating of the tables in the external software is automatically translated to the GIS.

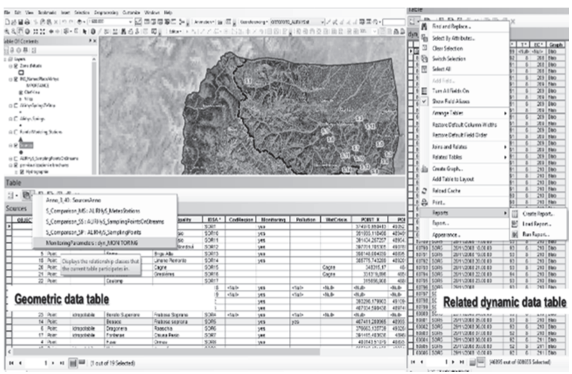

Figure 3: Example of resulting GIS with related dynamic tables.

Some difficulty is encountered in the limited compatibility of the two software (ESRI ArcMap and MsAccess) in their most recent versions, and by the limitations on data recording for memory occupation reasons. Therefore, in subsequent projects we will use open-source platforms to work towards general interoperability and better performance.

The data are then exported and published on a WebGIS, which can be accessed from www.polito.it/ALIRHYS for viewing the data; downloading is available through the Web Map Service (WMS), Web Coverage Service (WCS) and Web Feature Service (WFS). Since the data are semantically structured in a known and shared model, it will be easier to correctly interpret them and.

On the other hand, the implemented structures can be extracted from both software programs as an XML file (with or without data) and can be shared and easily imported by other users in the same software for obtaining similar researches and comparable data. The XML files are also shared on the project web site to facilitate the use of the same method.

Table 4: Dynamic data tables stored in the external DB.

DailyMeteoParameters (Row data)
HourMeteoParameters (Row data)
MonitoringParameterValues (Row data) on Springs
IsotopesParametersValues on SamplingPoints [LiquidWater]
IsotopesParametersValues on SamplingPoints [Snow]
BaseChemicalQualityParametersValues on SamplingPoints [LiquidWater]
BaseChemicalQualityParametersValues on SamplingPoints [Snow]
IsotopesParameterValues on Springs [LiquidWater]
BaseChemicalQualityParametersValues on Springs [LiquidWater]
MonitoringParameterValues (Row data) on Streams
TrackingTestsData (Row data)
TrackingParametersValues (Level 2 of processing)

3.2 Geoprocessing Tools of GIS for Subterranean Resources Analysis

The other important capability of GIS is represented by the geoprocessing algorithms, which can be used for analysing terrain models in order to extract important information for the interpretation of the ground, and, in this case, for the study of water resources.

This capability has been used for the automatic extraction of morphologic maps for the study of flows. Moreover, the extracted maps together with some data derived from satellite imagery have been used for estimating the snow volume in the study area in particular periods. This can be useful in order to manage the springs' water by foreseeing the possible flow rate in the snow fusion period.

The algorithms implemented in GIS management software are able to extract a number of morphologic and hydrologic data, starting primarily from the DTM. This can be extremely useful when discussing the geological aspects and water flows. In Figures 4-6 there are some examples of information extracted by a simple DTM. The processed objects correspond to the pink entities of the conceptual model.

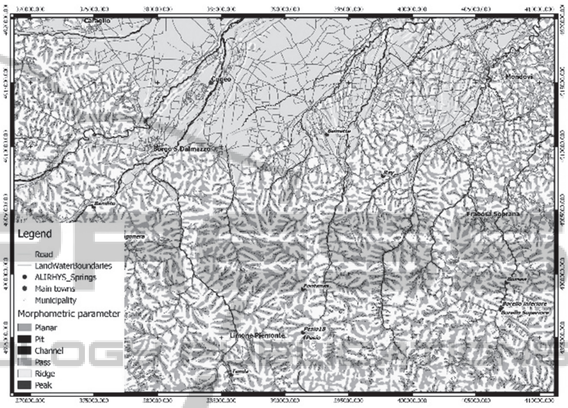

Figure 4: Morphometric parameters.

Figure 5: Catchment computed with the DTM as input file (in black). The stream's location appears, as visible from the comparison with its complement in the digital map (in red).

Some of these maps, such as the contour, aspect and slope maps, are simply extracted by the DTM; other algorithms analyse these basic maps to produce modelling maps of flow dynamics that permit researchers to locate the rivers, basins and other such features. This can help in the hydrologic and morphologic analysis. Another interesting tool permits the calculation of groundwater flow, starting from the DTM, the surface model of the aquifer, and some other data. This could an interesting avenue of future work linked to this project.

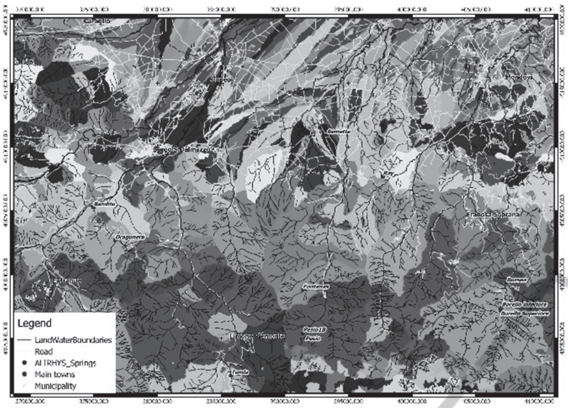

Figure 6: Sink watershed, allowing for the isolation and visualization of each basin for conducting more specific analysis on the hydric resources (tracking test hypotheses, rain and snow monitoring, pollution source monitoring, etc.).

4 CONCLUSIONS

Modelling the conceptual schema of a GIS using affirmed ontology makes the information less ambiguous, and data could be more easily sharable on the web in line with the goals of the Semantic Web. Using the integration of different application field ontologies makes it possible to exhaustively represent the environmental information, which traditionally has been an enormous part of documentation and monitoring data; this data is inherently related to the Earth's surface, a geometric object.

The GIS conceptual model includes several kinds of data: geometric vector data, dynamic data and raster datasets. It is the precondition for real integrated management of all these kinds of data, even if in this paper this work is still quite incomplete. Future goals will include the automatic implementation of such structures to achieve a unique framework for data sharing. Some limits of the system built are surely in the software used, because these programs do not include some essential characteristics of the ontological models. For example, an ORDBMS (Object-Relational Database Management System) such as PostgreSQL could better manage both the geometric and the dynamic data by also using some object-oriented properties that are extremely useful for expressing the object's structure.

Such approaches could be worse for the usability by inexpert users. This problem would be solved by managing the data directly on the web, through some user-friendly interfaces, even with a greater requirement for informatics contribution. On the other hand, this could effectively be a step towards the realization of the Semantic Web's aims of sharing structured data and processing services.

Moreover, a limit to interoperability is the use of commercial software, chosen for the user-friendly interfaces and widespread adoption; these programs work mainly with their own formats. In subsequent studies, this issue will be solved through the use of open source and self-integrated software. Furthermore, it is well-known that GIS data management is used for the geoprocessing and data-modelling of archived data, in particular for environment applications. The relationship between GIS and environmental modelling has been studied from a semantic point of view as well (Fallahi et al., 2008; Kiehle et al., 2006; Argent, 2004). This will be object of future in-depth analysis, with an ultimate goal of a truly complete and integrated system.

ACKNOWLEDGEMENTS

This project ALCOTRA ALIRHYS was possible due to Prof. Geol. Bartolomeo Vigna.

REFERENCES

Afflerbach, S, Illert, A, and Sarjakoski, T, 2004. The Harmonisation Challenge of Core National Topographic Databases in the EU-Project GiMoDig. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV(B4/IV), pp.129-134.

Bellone, T., Dabove, P., Manzino, A. M., Taglioretti, C., 2014. Real-time monitoring for fast deformations using GNSS low-cost receivers. In: *Geomatics, Natural Hazards & Risk*, pp. 1-13. ISSN 1947-5705

Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I., 2000. The Semantic Web: the Roles of XML and RDF. In: *IEEE Internet Computing*, 15, 3, 2000, 63-74.

DiGiuseppe, N., Pouchard, L. C., Noy, N., 2014. SWEET ontology coverage for earth system sciences. In: *Earth Sci Inform*. DOI: 10.1007/s12145-013-0143-1.

Disabato, F., 2010. *Remote sensing data analysis for environmental and humanitarian purposes. The automation of information extraction from satellite data*. PhD Thesis, Politecnico di Torino.

Dozier, J., Elder, K., Michaelsen, J., 1991. Snow accumulation and distribution in an alpine watershed. In: *Water Resources Research*, 27, 7, pp. 1541-1552.

Fallahi, G. R., Frank, A. U., Mesgari, M. S., Rajabifard, A.,

2008. An ontological structure for semantic interoperability of GIS and environmental modelling. In: *International Journal of Applied Earth Observation and Geoinformation* 10, Elsevier, pp.229-246.

Fonseca, F., Egenhofer, M., Davis, C., Camara, G., 2002. Semantic granularity in ontology-driven geographic information systems. In: *MAI Annals of Mathematics and Artificial Intelligence* 36 (1e2).

Freksa C., Barkowsky T., 1996. On the relation between Spatial Concepts and Geographic Objects. In: *P. Burrough & A. Frank, Geographic objects with indeterminate boundaries, London.* Taylor & Francis, pp. 109-121.

Gröger, G., Kolbe, T. H., Czerwinski, A., Nagel, C., 2008. OpenGIS City Geography Markup Language (CityGML) Encoding Standard. Open Geospatial Consortium Inc., OGC 08-007r1.

Kiehle, C., Greve, K., Heier, C., 2006. Standardized geoprocessing - taking spatial data infrastructures one step further. In: *The 9th AGILE Conference on Geographic Information Science, Visegràd, Hungary*.

Mennis, J. L., 2003. Derivation and implementation of a semantic GIS data model informed by principles of cognition. In: *Computers, Environment and Urban Systems* 27 (2003), pp. 455-479.

Noardo, F., Lingua, A., Aicardi, I., Vigna, B., 2015. Cartographic data harmonisation for a cross-border project development. In: *Applied Geomatics.*

Pundt, H., Bishr, Y., 2002. Domain ontologies for data sharing – an example from environmental monitoring using field GIS. In: *Computers & Geosciences*, 28(1), pp. 95-102.

Raskin, R., Pan M., Mattmann C., 2004. Enabling semantic interoperability for earth science data. In: *4th NASA Earth Science Technology Conference*. http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/

Scholl, M., Voisard A., 1992. Object-oriented database systems for geographic applications: an experiment with O2. In: *Proc. Int. Workshop on Database Management Systems for Geographical Applications,(Capri, 1991),* Springer.

Tóth, K, Smits, P, 2007. Infrastructure for spatial information in Europe (INSPIRE): from cartography to spatial objects and network services, In: *Proceedings of the 23rd Cartographic Conference – Cartography for everyone and for you,* Roskartografija, p.1-11.

Tschirner, S, Scherp, A, Staab, S, 2011. Semantic access to INSPIRE. In: *Terra Cognita 2011 Workshop Foundations, Technologies and Applications of the Geospatial Web* (p. 75).

Waters, J., Powers, B. J., Ceruti, M. G., 2009. Global Interoperability Using Semantics, Standards, Science and Technology (GIS3T). In: *Computers Standards & Interfaces*, 31, Elsevier, pp. 1158-1166.

Worboys, M. F., Duckham M., 2004. GIS: A computing perspective. CRC press.