



HAL
open science

Integration of Heterogeneous Spatial Databases for Disaster Management

Imen Bizid, Sami Faiz, Patrice Boursier, Jawahir Che Mustapha Yusuf

► **To cite this version:**

Imen Bizid, Sami Faiz, Patrice Boursier, Jawahir Che Mustapha Yusuf. Integration of Heterogeneous Spatial Databases for Disaster Management. SeCoGIS 2013 7th International Workshop on Semantic and Conceptual Issues in GIS, Nov 2015, Hong Kong, China. 10.1007/978-3-319-14139-8_10 . hal-01287169

HAL Id: hal-01287169

<https://hal.archives-ouvertes.fr/hal-01287169>

Submitted on 15 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Integration of Heterogeneous Spatial Databases for Disaster Management

Imen Bizid^{1,2}, Sami Faiz¹, Patrice Boursier², and Jawahir Che Mustapha Yusuf³

¹ Laboratoire LTSIRS, BP 37 Le Belvedere 1002, Tunis, Tunisia

² Laboratoire L3i, University of La Rochelle, Pole Sciences Technologies Avenue M. Crepeau, La Rochelle, France

³ UniKL-Malaysian Institute of Information Technology, University 2, Kuala Lumpur, Malaysia

Abstract. The response phase in a disaster case is often considered to be the most critical in terms of saving lives and dealing with irreversible damage. The timely provision of geospatial information is crucial in the decision-making process. Thus, there is a need for the integration of heterogeneous spatial databases which are inherently distributed and created under different projects by various organizations. The integration of all relevant data for timely decision making is a challenging task due to syntactic, schematic and semantic heterogeneity. This paper aims to propose a framework for the integration of heterogeneous spatial databases using novel approaches, such as web services and ontologies. We focus on providing solutions for the three levels of heterogeneity, in order to be able to interrogate the content of the different databases conveniently. Based on the proposed framework, we implemented a use case using heterogeneous data belonging to La Rochelle city in France.

Keywords: Heterogeneous Spatial Databases, Data Integration, Disaster Management, Ontologies

1 Introduction

Disaster management is a scope of great relevance as it aims to reduce the negative impact and consequences of a disaster and to provide a timely and clear rescue plan. Proposing powerful tools to generate efficient strategies for the decision makers gives rise to the delivery of disaster relief in the right moment. Two types of disasters are of interest natural and man-made. Natural disasters often strike without warning, such as earthquakes, tsunamis and tornados. Man-made disasters include every action due to technological hazards, sociological hazards or transportation hazards among others. Despite the difference between these two categories in specifics and complexity, both instances can cause irreversible damage if the rescue plan is not executed at time. In order to effectively manage such disasters, it is important to coordinate between the police, Red Cross, and other first responders to work together and extract the needed information to

maintain a rescue plan. However, **the data stored** in these distinct heterogeneous sources located in different departments, **are** maintained under heterogeneous formats, structures and **are** not semantically rich enough to express the emergency responders' needs. Experience suggests that the bottlenecks of emergency responses are, in most cases, the difficulties in making intelligent search and integration of heterogeneous geospatial information stocked in different departments [1]. This paper **has been working** on the heterogeneity levels that can block emergency responders to automate the search of geospatial data stored in different spatial databases. Although there were many works that all address important pieces of the major challenge to automatically search geospatial information from heterogeneous spatial databases, many issues still remain to be solved to fully realize this goal. This paper is structured as follows: In section 2, we describe the three heterogeneity levels. In section 3, we present briefly the recent progress made on the integration of heterogeneous spatial databases, Section 4 proposes a new solution represented in a framework dealing with the three heterogeneity levels. In section 5, we describe the different interoperability modules designed in our framework. Afterward, we analyze the evaluation results related to the framework in Section 6. Finally, we present our conclusions in section 7.

2 Spatial Databases Heterogeneity Levels

Each emergency response community has developed its own database using its personalized infrastructure and its own requirements. Hence, in order to maintain a rescue plan, emergency responders have to interrogate at the same time the overall data stored in heterogeneous databases through complex queries. This gives a rise to the problem of access to the required information in a critical time. As mentioned in [2, 3] that illustrate databases heterogeneity through three levels, i.e., semantic schematic and syntactic.

Syntactic heterogeneity refers to data format differences and problems related to the implementation of databases in different storage paradigms: relational or object orientated, and the geometric representation of objects: raster or vector. Many organizations contributing in a disaster case rely on a range of storage solutions. However, some organizations are motivated to make their data accessible through Open Geospatial Consortium (OGC) formats that provide syntactically unified formats and services such as Geography Markup Language (GML) [4] and Web Feature Service (WFS) [5].

Schematic heterogeneity regards the differences in data models **between** emergency response communities. Each spatial database schema reflects each community's abstracted view of data. Therefore, different hierarchical and classification structures are used by each community to refer to identical or similar objects. Schematic heterogeneity can be classified into three main schemes:

- Entity versus Entity, i.e., the same entity can exist in two different databases with different name or structure;
- Attribute versus Attribute, i.e., an attribute related to a class in one database can exist in another class related to another database;

- Entity versus Attribute, i.e., a class in one database can be designed as an attribute in another database.

Semantic heterogeneity can be divided into two types: the *naming heterogeneity* and the *cognitive heterogeneity*. The naming heterogeneity arises when the same data objects are named in a different manner or when different semantic data objects are named identically. The cognitive heterogeneity refers to the different domains assigned to each organization. Each community has different cognitive views of a word, which means that they describe similar real word objects from different perspectives.

It is against these complex backdrops that our paper applies. Hence, we offer a comprehensive approach providing an access to the whole information consistently and quickly.

3 Related Works

The integration of heterogeneous databases constitutes an old challenge for computer information systems communities; it has been extensively studied in the literature. Therefore, various approaches have been proposed that we divide into two broad categories; traditional approaches and novel approaches. For the first category, two main approaches can be cited: the virtualized approaches that include the *Federated Databases* approach [6] and the *Mediator / Wrapper* approach [7], and the *Data Warehouse* (or materialized) approach [8, 9]. Novel approaches deal with interoperability issues, and include *Web Services* [10], *Ontologies* [11, 12] and *Semantic Web* approaches [13]. Due to the large volumes of spatial data, the complexity of handling and analysing data and its importance in various domains, different researchers working on geographical information systems all over the world have focused on the adaption of the existing integration solutions to spatial databases. Thus, by using different techniques that integrate the spatial component, solutions have been proposed, such as *Spatial Data Warehouse*, *Federated Spatial Databases*, *Geo-web services*, *Spatial Data Infrastructure* (SDI) [14] among others.

Table 1. The interoperability position of the different integration approaches.

Approaches	Syntactic	Schematic	Semantic
Spatial datawarehouses		X	
Federated Spatial Databases	X	X	
Mediation systems	X	X	
Spatial Data Infrastructure	X		
Geo-Web Services	X		
Ontologies		X	X
Semantic Web	X		X

Traditional and innovative approaches presented in Table 1 provide partial solutions for the integration of heterogeneous spatial databases; they only afford one or two interoperability levels and most of them do not support the specification of the data semantics. However, the area related to heterogeneous databases and disaster management is still asking for new solutions and approaches. Currently, there exist some software packages, such as ArcGIS⁴ and Sahana⁵, that are used in real disaster scenarios in order to help emergency actors. However, they do not take into account the semantic interoperability and the entities and naming conventions in different databases are duplicated, ambiguous and imprecise. Hence, urgent solutions are required in order to solve these ambiguities that result from the heterogeneous spatial databases stored in diverse organizations.

4 A Hybrid Approach for Providing Intelligent Search in a Disaster Framework

The previous sections have introduced the different heterogeneity levels that could block emergency responders when they query various heterogeneous spatial databases. In response to these difficulties, we design a conceptual framework using Geo-web services and ontology based data integration approach to support the three heterogeneity levels. This framework is based on SOA (Service-Oriented Architecture), which is composed of a service provider, a service broker, and a client service as shown in Figure 1.

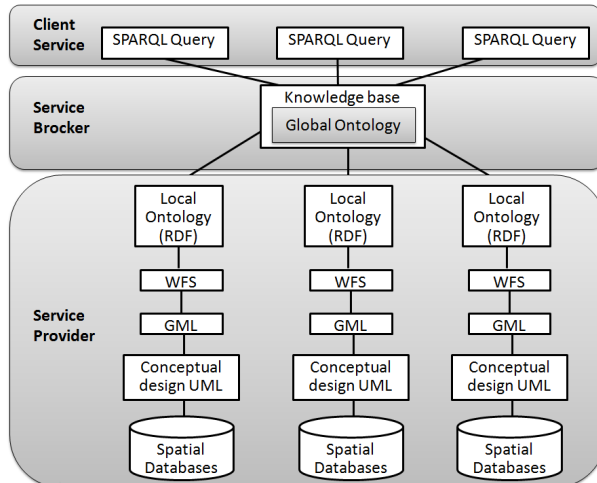


Fig. 1. The framework architecture

⁴ <http://www.esri.com>

⁵ <http://sahanafoundation.org>

Service Provider affords the syntactic interoperability, the data models are conceptualized in UML⁶ (Unified Modeling Language) and converted afterward into GML for sharing data. OGC WFSs (Web Feature Services) were used for publishing feature-level data from heterogeneous databases. For the schematic and semantic interoperability, two types of mappings are required: information source to local ontology mapping and local ontologies to global ontology mapping. Information source to local ontology mapping is obtained by translating GML (Geographic Markup Language) data to RDF (Resource Description Framework) using an automatic conversion system. Local ontologies to global ontology mapping is used for merging the different local ontologies built into one global ontology.

Service Broker is the main component of the architecture that contains the global ontology, a mapping directory, and a binding technique to apply the ontology hybrid approach by connecting the local ontology to the global ontology. The global ontology is used to create mappings of equivalent classes and attributes in the local ontologies. Hence, it describes the global model that represents local ontologies.

Client Service employs the RDF SPARQL query language to query the global ontology that contains all the properties defined in the different local ones. The results of SPARQL queries will be presented in a map using Google Visualization API and SPARQL (SPARQL Protocol and RDF Query Language) proxy.

5 The framework modules

In this section, we explain the two modules developed in our framework in order to ensure the three interoperability levels. Firstly, we describe the syntactic interoperability module that includes the conversion of the UML models into GML and the integration of these GML models into web feature services. Then, we analyze the schematic and semantic module used to build the different local ontologies and to merge them into a global one.

5.1 Syntactic Interoperability Module

This module deals with the syntactic heterogeneity problems. It contains an automatic converter from UML application schema to a common GML format. The application schema is expressed in UML, and the chosen exchange format is GML. A set of conversion rules have been identified and implemented in a tool named Shapechange that reads UML class diagrams and writes corresponding GML code. Therefore, in order to ensure the GML conversion, we have firstly to conceptualize the different UML models for each spatial database. Then, the UML application schemas have to be converted into XMI representations. Once the XMI documents have been designed using UML, conversion to GML schema

⁶ <http://www.uml.org>

specified by the OGC can be done easily by defining some conversion rules. The UML to GML Application Schema mapping tool supports this conversion automatically. After standardizing the different databases into GML files, they are published and integrated in WFS.

5.2 Schematic and Semantic Interoperability Module

In our framework, we opt to use the hybrid ontology approach for its various advantages [12]. In order to apply this approach, two types of mappings are required: information source to local ontology mapping and local ontologies to global ontology mapping.

Information Source to Local Ontology Mapping The local ontologies corresponding to each database are built using two languages, namely the Resource Description Framework (RDF) and the Web Ontology Language (OWL) for representing triples and modeling the relations between them. Such a representation is obtained by using the GML2RDF automated conversion system [15]. Each entity in the database is represented by its own configuration which contains a list of all the attributes. The following steps present briefly the building of the different local ontologies:

- Each spatial database is separately converted into its local ontology.
- Each entity in the database is converted to the class name of the local ontology. Each component related to a class is described using a set of triples. This triple comprises three parts: subject, predicate, and object.
- The subject is the result of the conversion of each class instance in the dataset.
- The predicate is the conversion result of each non-prime attribute. It expresses a relationship between the subject and the object.
- The object is obtained by converting each attribute value corresponding to an attribute.
- If two classes C1 and C2 take part in formal corresponding relationships, then class C2 will be declared as the subclass of class C1. The foreign key relationship between the classes is shown as the predicate where C2 will be the domain.

Local Ontologies to Global Ontology Mapping After building the different local ontologies, we have to map them in order to determine the semantic and schematic correspondences between the concepts of two ontologies. If two concepts correspond, then they mean the same thing or closely related things. The mapping process proposed in our research is based on the two approaches presented below.

N-Grams match [16] is a method used to compute lexical similarity between two words. Generally, the N-Grams algorithm takes as input two strings and computes the number of common substrings between them. An n-gram is a sequence of n characters; for each string, the set of all possible n-grams, are computed. Then, a pre-processing step is executed to normalize both strings.

The third step is to extract the n-grams from each string. Finally, the algorithm counts the matching percentage of the two strings.

Synonym match [17] is a technique for detecting semantic similarity between two words. In many cases, two strings may appear as different words but may represent the same meaning in a particular domain. To identify such semantically related words, we use the ontology and lexical database Wordnet.

Ontology Matching Steps To map the local ontologies and merge them into a global one, all the possible combinations of local ontologies should be detected and mentioned in a list. Then, we extract and match the different classes in each local ontology.

Class Name matching. The different classes are represented in prefixes. Hence, we extract all the names of classes from there without considering the predefined prefixes list. Then, we apply the class name match, each class in the first local ontology is paired with all other entities in the second one. The N-grams and the Synonym match scores are used to validate each step. So, the N-grams score should be greater than a threshold and the Synonym match returns a boolean result synonym (1) or not synonym (0). If the matching result is validated by the N-gram or the Synonym match, then we can proceed to next matching steps using the same approaches and techniques.

Attribute Type matching. This step is executed only if the previous one was validated by a matching approach. At this stage, the attribute type of the matching class is extracted from the list created for each class name containing the attributes name and their corresponding types. This list is obtained by the extraction of the attributes name from the predicate. The data type attribute is paired in such a way that one data type attribute related to a class in one local ontology is matched with all the data types of the second class in the other local ontology. If two types of attributes related to one pair are detected as similar then we proceed to the next step to compare their related names.

Attribute Name matching. This phase could be employed only if the pair of attribute names that will be matched have the same data types. In this case, the attributes of the matching entity names are executed.

Attribute Value matching. Until now, this approach is relatively susceptible to homonym-like pairs of strings: although syntactically very similar, they could mean entirely different things, N-Grams and N-Synonym will give them a high similarity score. To deal with this problem, we apply an attribute value match. This matching is executed only if the attribute name matching of one pair was validated, then three instances related to each attribute are matched to detect if the two classes are similar or not. Moreover, this solution is extensible and we could make an attribute-class matching, relations matching and many others using the same principle.

Local Ontologies Merging To merge the different local ontologies, we analyze the matching results using algorithm 1. If two classes are matched as similar,

then one of these classes will be added to the global ontology. Else, the two classes are saved in the global ontology. The same method is applied for the couple attribute-class and attribute- attribute merging.

Algorithm 1 Merging local ontologies

```

Require: ListOntologieN3 listOnt;
for (int i = 0; i < listOnt.size - 1 ; i=i+2) do
  ListClassMatching = matching(listOnt[i],listOnt[i+1])
  for (int j = 0; j < ListClassMatching.size-1 ; j++) do
    if classmatching[j][2] = true then
      if (!GlobalOntology.contains(classmatching[j][0])) then
        GlobalOntology.add(classmatching[j][0])
      end if
    else
      if (!GlobalOntology.contains(classmatching[j][0])) then
        GlobalOntology.add(classmatching[j][0])
      end if
      if (!GlobalOntology.contains(classmatching[j][1])) then
        GlobalOntology.add(classmatching[j][1])
      end if
    end if
  end for
end for
return GlobalOntology

```

After constructing our global ontology, we need to store it in an ontology server in order to be able to query it. Therefore, we stored the resulting global ontology in the joseki⁷ ontology server.

6 Experiments and Results

Based on the proposed framework, we have implemented a prototype for allowing emergency responders to automatically search from heterogeneous datasets. Therefore, we have used three heterogeneous datasets belonging to La Rochelle city in France. The first one represents the office of mayor data stored in KML files, the second one contains the emergency service data stored in MIF files and the third dataset belongs to the police service data stored in Shapefiles. Then, we designed the different UML models conceptualized for each dataset using Rational Rose Software.

The Unisys XMI tool is used to store the different UML models in a single XMI document. At that time, conversion to GML schema specified by the OGC can be done easily using the UML to GML mapping tool Shapechange⁸. To inte-

⁷ <http://www.joseki.org>

⁸ <http://www.shapechange.net/>

grate the different GML files, we use OGC web services specially WFS. Thus, we use Geoserver 2.2.1 and the GML plugin to publish the generated GML datasets. To build the local ontologies for each dataset, a conversion program named GML2RDF is used which makes use of two predeveloped open source APIs for parsing GML files and handling the creation of the RDF models. The GeoTools API is used for reading the data from GML files, and the Jena API is used to contain the converted RDF data in a memory model and eventually write the data into a N3 file. Similarly, the Jena semantic web framework for java is integrated in our application to access ontology definitions, analyze the WordNet thesaurus and to realize the different necessary mappings. To evaluate the mapping approaches applied in our framework, we visualize the different results provided by the mapping algorithm. Hence, these different results are analyzed using our three datasets and 90% of them are matched correctly. A spatial ontology server was developed based on Joseki to support the global ontology N3 file and to interrogate it using SPARQL forms. The query results are displayed into a satellite image using Google Maps API. Our experiments yield to promising results and show that, with the implemented prototype, it is possible to directly search geospatial information from heterogeneous sources using our framework. Therefore, to deliver the needed information in a short time, we only have to interrogate our global ontology using a simple query that returns the response as a map.

7 Conclusion and Future Work

In this paper, we propose a framework for the integration of heterogeneous spatial databases that enables emergency responders to easily query spatial heterogeneous databases and facilitates the automatic search of geospatial information in order to make quick and correct decisions and prompt actions. Our framework uses novel techniques maintained under SOA architecture to deal with three major tasks: Syntactic interoperability is analyzed by adopting GML as a common language; Schematic (Structural) interoperability is construed by the ontology mapping process and Semantic interoperability is assured by using the hybrid ontology approach.

Furthermore, by the extensible and autonomy characters hidden behind the ontology approach and OGC web services many interesting functionalities can be added. We plan to treat information from texts and comments posted in social networks. This will be helpful in the case of a lack of information.

References

1. Zhang, C., Zhao, T., Li, W.: Automatic search of geospatial features for disaster and emergency management. *International Journal Applied Earth Observation and Geoinformation* **12**(6) (2010) 409–418

2. Bishr, Y.: Overcoming the semantic and other barriers to gis interoperability. *International Journal of Geographical Information Science* **12**(4) (1998) 299–314
3. Stuckenschmidt, H., Harmelen, F.: *Sharing on the Semantic Web*. Springer Berlin Heidelberg (2005)
4. Cox, S., Daisey, P., Lake, R., Portele, C., Whiteside, A.: *Opengis geography markup language (gml) implementation specification, version 3.0* (2003)
5. Zhang, C., Li, W.: The roles of web feature and web map services in real time geospatial data sharing for time-critical applications. *Cartography and Geographic Information Science* **32**(4) (2005) 269–283
6. Heimbigner, D., McLeod, D.: A federated architecture for information management. *ACM Transactions on Office Information Systems* **3**(3) (1985) 253–278
7. Leclercq, E., Benslimane, D., Yetongnon, K.: Amun: An object oriented model for cooperative spatial information systems. In: *Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop*, Washington, DC, USA, IEEE Computer Society (1997)
8. Inmon, W.H.: *Building the Data Warehouse*. John Wiley & Sons, Inc., New York, NY, USA (1992)
9. Kimball, R.: *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley (1996)
10. Zhang, S., Gan, J., Miao, L., Lv, G., Huang, J.: Study on gml spatial interoperability based on web service. In: *Proceedings of the 31st Annual International Computer Software and Applications Conference - Volume 01*, Washington, DC, USA, IEEE Computer Society (2007) 649–656
11. Xu, W., Zlatanova, S.: Ontologies for disaster management response geomatics solutions for disaster management. In Li, J., Zlatanova, S., Fabbri, A.G., eds.: *Geomatics Solutions for Disaster Management*. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg, Berlin, Heidelberg (2007) 185–200
12. Mustapha, Y.J., Mohd Su'ud, M., Boursier, P.: Analyzing the ontology approaches and the formation of open ontology model: A step for organisational ontology employment. In: *Semaparo: The Fifth International Conference on Advances in Semantic Processing*, Lisbon, Portugal (2011) 6–13
13. Zhang, C., Li, W., Zhao, T.: Geospatial data sharing based on geospatial semantic web technologies. *Journal of Spatial Science* **52**(2) (2007) 35–49 Zhang, C. Li, W. Zhao, T.
14. Maguire, D.J., Longley, P.A.: The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems* **29**(1) (2005) 3–14
15. Bulen, A., Carter, J., Varanka, D.: A program for the conversion of the national map data from proprietary format to resource description framework (rdf). Technical Report U.S. Geological Survey Open-File Report 20111142., U.S. Department of the Interior (2011)
16. Cavnar, W., Trenkle, J.M.: N-gram based text categorization. In: *Proceeding of 3rd Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas (1994) 161–175
17. Lin, F., Sandkuhl, K.: A survey of exploiting wordnet in ontology matching. In Bramer, M., ed.: *Artificial Intelligence and Practice II*. Volume 276. Heidelberg, Springer, Berlin, Heidelberg (2008) 341–350