An Ontology based Approach to Integrate Data and Maps
In the Government Enterprise Architecture: A Case Study

Daniela Giordano, Alfredo Torre, Carmelo Samperi, Salvatore Alessi and Alberto Faro
Department of Electrical, Electronics and Computer Engineering
University of Catania, viale A.Doria 6, 95125, Catania, Italy

Keywords: e-Government, Government Enterprise Architecture, Geographic Information Systems, Ontology Technologies.

Abstract: The problem of managing data and maps within an ontological approach is little studied in the Government Enterprise Architecture. Aim of this paper is to present a methodology to solve this problem in case we would join municipal and cadastral data bases. In particular, we aim at linking the information contained in the local taxation registry to the urban territory to allow the Public Administration managers to check if the taxes have been paid, and the citizens to compute the correct amount to pay. We plan to extend this methodology to manage other relevant location based services that need to interconnect public and private data stores of city interest to vector drawings derived from the Cadastre or other CAD systems to define territorial plans immediately understandable to all the involved organizations.

1 INTRODUCTION

The Semantic Web is a mesh of information linked up in such a way to be processed easily by machines on a global scale (Sheth, 2005). The Semantic Web is built generally on syntaxes which use *International Resource Identifiers* (IRIs) to represent resources, i.e., subjects and objects, linked by properties. Subject-predicate-object relations are represented by triples, also called semantic web statements. Let us recall that IRI is an extension of the *Uniform Resource Identifier* (URI) that provides an encoding for Unicode character sets.

The semantic web statements are usually formalized by the Resource Description Framework (RDF), i.e., a directed multi-graph consisting of subjects, predicates and objects (Hayes, 2004). The RDF graph can be queried by means of the SPARQL query language to retrieve and manipulate the stored data (Prud'hommeaux, 2008). The RDF Scheme (RDFS) is a collection of RDF resources that behaves as a vocabulary of terms and properties related to application-specific domain. Such vocabularies may range from controlled lists of terms to taxonomies and thesauri depending on the type of terms and relationships that can be expressed (e.g. parent-child relationships in a taxonomy).

Ontology refers to a formal specification of a shared vocabulary and allows us to define formally a set of terms, interconnections, constraints and rules of inference on a particular domain (Zhai, 2008), (Faro, 2003). A logical formalism is needed to represent an ontology such as Description Logic (DL) (Baader, 2003).

Ontology, with rule definition language and description logic, can also provide a new kind of data retrieving and mash up with the "backward chaining" concept to make possible the inference of data structures not present in the knowledge base at the moment of the query.

The use of an ontology with the intention of describing a particular aspect of reality, provides information reusable for all parties in the given domain. Regarding the *e-government* activities, the Linked Data group at the W3C and the Government Linked Data (GLD) are publishing data sets and knowledge bases (often in the form of light-weight ontologies and vocabularies) to support e-government services involving different organizations. These ontologies are under test and will be refined in the next future by incorporating novel global and local vocabularies.

The problem of managing data and maps within the mentioned ontological approach is little studied because it is necessary to study more complex

Giordano D., Torre A., Samperi C., Alessi S. and Faro A..
An Ontology based Approach to Integrate Data and Maps - In the Government Enterprise Architecture: A Case Study.
DOI: 10.5220/0004558403560362
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (ICEIS-2013), pages 356-362
ISBN: 978-989-8565-61-7

problems that involve location based information, and because unifying terms of proprietary vocabularies such as road and street in view of a shared vocabulary implies only an equivalence between symbols, whereas equivalent drawings even if they are labelled by the same name, e.g. building, needs to be processed by complex conversion procedures when passing from the adopted Geographic Information System (GIS) to the one used by another organization to be sure that they deal with the same physical thing.

Therefore, in problems starting from personal and cadastral data based on maps, as well as escape routes in case earthquake or traffic light optimization, we have to adopt not only a standard vocabulary that behaves as a bridge between equivalent terms used in the proprietary systems, but also conversion procedures to ensure that a physical vector in a GIS is the same in another one.

Of course, such problem would disappear if one adopt the same vocabulary and the same GIS in all the computing systems, however this is not only unrealistic but also not useful since proprietary codification of data and drawings may be more effective than the standard ones to carry out some basic operations such as storing and updating.

Aim of this paper is to present an ontology based methodology to solve this problem by illustrating how it works in practice by a case study dealing with the computation of local taxes from municipal and cadastral data. In particular, we aim at linking the information contained in the local taxation registry to the urban territory by geographic points (*Points Of Interest* - POI) to allow the Public Administration (PA) to check if the taxes have been paid, and the citizens to determine the correct amount to pay.

This work aims at supporting the transition of the PA information systems from their current structure, often consisting of separated silos of data, towards a *Government Enterprise Architecture* (GEA) able to integrate all the administration data and maps, to optimize internal procedures and to improve the relationships with citizens.

Therefore, to speed up this transition we have to increase the levels of impact of GEA to the Public Administrations with the final aim of integrating all their data in a sort of *Connected Government* Model that enables "governments to connect seamlessly across functions, agencies, and jurisdictions to deliver effective and efficient services to citizens and businesses" (Saha, 2010)**.**

To identify what is needed in practice to favour this transition, let us recall that the main dimensions of a GEA are: Citizen centricity, Common infrastructure and interoperability, Collaborative services and business operations, Public sector governance, Networked organizational model, Social inclusion. This implies that to support the above transition we have to increase the PA web applications and to adopt data standardization and integration technologies to organize the PA work according to the business model most suitable for the organization at hands.

Since the PA organization model is outside the scope of the paper, the work illustrated in the paper may be reused only to improve dimensions such as citizen centricity, common infrastructure and inter-operability, collaborative services and business operations, whereas public sector governance, networked organizational model and social inclusion are for further study.

Let us note that even if the paper discusses a specific case, i.e., the computation of local taxes from municipal and cadastral data, the proposed methodology may be followed to manage other problems involving data and maps.

Sect. 2 illustrates the ontologies and technologies that allow the taxation registry and the land registry to be interconnected in a single RDF framework.

Sect. 3 points out the main steps to convert the proprietary SQL codification of the original data bases into standard RDF statements, as well as the map conversion to allow the physical entities associated to the terms of the ontologies to be represented by the same physical thing in almost all the available open source GISs.

Sect. 4 presents the SPARQL queries that allow the citizens to extract the geo-referenced reports on how much they should pay to PA for their estates and support the PA employers to check if the citizens are in arrears.

2 JOINING E-GOV DATA&MAPS

As pointed out in the introduction, the final purpose of this work is to develop a new kind of distributed system architecture capable of aggregating heterogeneous data from multiple data sources that have their own storage and representation format. In particular, the paper aims at integrating municipal and cadastral data bases by using suitable ontologies and technologies as illustrated in the following sections.

2.1 Ontologies

To identify the relevant ontologies of an egovernment problem, the first step is the one of classifying the entities involved in the specific domain of interest. With reference to the mentioned local taxation problem at the centre of the case study, the main elements are: the *taxpayer and* her/his *personal data*; the *property tax data* referred to a specific period of time; and the *waste tax data*.

The element enabling the right connections among the various entities in the database is the taxpayer identification number subdivided in people and organization identification number. Therefore, the main taxation concepts are as follows:

- County
  - **City**
- Taxpayer
  - **Citizen**
  - **Organization**
- Report
  - **property tax**
  - **waste tax**

In the above classification, the Report class is connected to the cadastral geographic entities to compute the local taxes. Thus, we have to analyse how the Cadastre is organized. In Italy, the Cadastre consists of two main sections: the Cadastre of Land Properties and the one of Real Estates. The data deal with owners and holders of the estates or the lands as well as geographical location, size, intended use, earning capacity and consistency.

Since in the paper we are interested in the Real Estates Cadastre (REC), we have to deepen its structure. Thus, we found that it is divided into "*pages*", each of which includes *parcels* associated to the basic entities, i.e., the Urban Real Estate Unit (UREU), defined as a portion of a building, an entire building or set of buildings that is capable of producing an independent income.

It should be emphasized that the UREU must have a range of income and a functional autonomy; however, there is not any constraint in the definition of UREU that it has to belong to a single owner. Consequently, a Real Estate Unit, belonging for example to two owners, will be reported with a single contextual registration and double identification. Each UREU is characterized by a set of cadastral codes that give rise to a unique identification.

In order to define the ontologies needed in our scenario, we selected the following identifiers: *Cadastral municipality*, i.e., the municipality where the property is located; *Administrative Section*, i.e., a portion of the municipality; *Page*, i.e., a section of the municipality that the Cadastre Registry represents in its cartographic maps; *Parcel*, i.e., a piece of land or building and any area of relevance within the Page; *Subordinate*:, i.e., the UREU.

Generally, each UREU is identified by its own subordinate, but, if the building is made up of a single UREU, then the subordinate may be missing.

Considering the case of our interest and the above cadastral data structure, we attached the cadastral main entities (*page, parcel, subordinate*) to the fundamental geometrical ontology, as shown in fig.1, with the future goal of providing a wider ontology scheme including other classes.

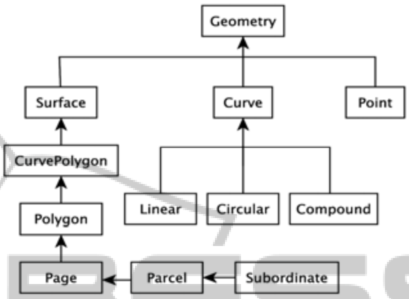

Figure 1: Geographic and cadastral entities.

2.2 Technologies

Since our aim is to expose RDF data structures through a SPARQL endpoint, first we carried out the data normalization of the original relational database, and then we mapped these data in RDF triples using the D2RQ Platform (http://d2rq.org), that is an Open Source system that offers RDF based access to the contents of the relational databases. Fig.2 presents how users may query the RDF triple store and the technologies involved in the deployment of the proposed distributed architecture.

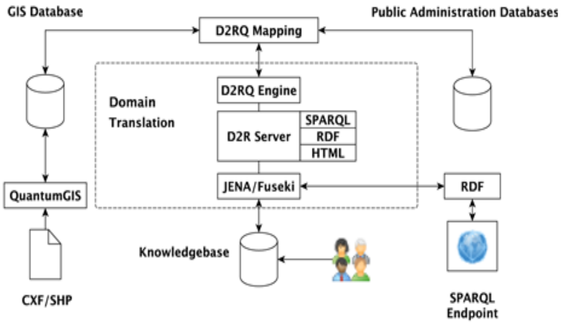

Figure 2: Overview of the multi-tier system architecture and of the technologies adopted.

Each technology is a free and open source software to allow the Public Administration to reduce running expenses and maintenance costs. Also this choice allows us to follow the cornerstone philosophy of the Open Data movement, whose description is given at http://opendefinition.org/.

As shown on the left of fig.2, our model suggests that the data sources from the cadastral domain should be converted from CXF format to *shapefile*; then they are imported into Quantum GIS to be exposed as vector drawings provided with geo-referenced data.

These latter data, together with the public administration data, are mapped through the D2RQ Platform into RDF schemas in order to allow a multi-tier querying by RESTful Web Services, RDF marshalling and HTML/XHTML visualization.

A Jena/Fuseki framework is also contained into the domain translation so that the above-mentioned tiers may be attached to external RDF sources (such as external SPARQL endpoints). The responses provided to the users are stored in a knowledge base to be used to speed up the future queries.

3 CASE STUDY

In this section we point out the main problems that arise when one tries to implement the previous architecture to provide in practice a specific e-government service, i.e., the local estate taxation depending on both personal data of the owners and cadastral data of the real estate units.

3.1 Municipal Databases

The datasets about citizen and taxation were provided by the local administration in the form of IBM DB2 Databases. Therefore, before using the D2RQ Platform these databases were converted into MySQL databases to work on an open source format. Then we selected the data of interest, normalizing and exporting them into an additional MySQL database. Finally the last step of the process was the one of mapping the contents of the databases into RDF triples through the on-the-fly translation obtained by the above mentioned D2RQ Platform.

Although the above procedure seems easy, several problems were encountered such as the lack of semantics in the definition of the tables analysed in this database. Indeed, it was made up of 58 tables and 19,170,374 records containing many repetitions, thus we have to cut off the tables to a large extent, and to select only a subset of the data that would be useful for the problem at hands.

Since the taxation process has to be applied to citizens who own a property, only table rows corresponding to estate owners were selected. Also, only relevant columns of the above tables were imported in a MySQL database, with the intent of mapping them in RDF triples to be used for supporting the taxation payment and checking using SPARQL queries.

To this aim, we have carried out a connection of the resulting municipal triple stores to the information stored in the cadastral database, treating the GIS data bases not as MySQL or PostGIS data bases, but as a virtual RDF graph obtained using a custom D2R mapping system, such as G2R (Della Valle, 2010). Thanks to these structured semantic connections we obtained a unique SPARQL endpoint where it is possible to attach many kinds of end-user application logic.

3.2 Cadastral Databases and GIS

Let us note that the cadastral datasets were provided by the provincial Land Administration as an extraction of the national cadastral map database from WEGIS, that is a licensed closed source powered by SOGEI http://www. sogei.it) and used by the Italian Land Administration. In this extraction each Page is represented by a pair of ASCII files: a CXF file (*Cadastral eXchange Format*) containing all the graphical elements that compose the cadastral map, and a homonymous SUP file containing statistical data and parcel surfaces.

Thus, a suitable conversion of the CXF and SUP files should be done to store the cadastral information within relational GISs that can be interconnected to the municipal data, as suggested in the previous section.

For this reason, these two files were converted into ESRI *shapefile*, i.e., a geospatial vector data format for geographic information system software developed by ESRI using *CXFToShape,* i.e., a free CXF to *ESRI shapefile* converter. Then, the *shapefile* was imported into Quantum GIS, i.e., a cross-platform free and open source GIS application that provides capabilities of data visualization, editing, and analysis and may be viewed as a virtual RDF store using the mentioned D2R Platform.

Let us note that in this way we may use the textual information contained in the Cadastre as RDF triples, but for using the drawings in any GIS it is necessary to represent them into a reference system known by all the GISs available on the market. Thus, we have geo-referenced the *shapefile* imported into Quantum GIS by means of the *Cassini-Soldner* geo-coordinate system through the definition of a custom projection algorithm.

In this way the cadastral data may be processed as triples and the vector drawings can be overlapped to any raster data layers (*Google Street Maps* *satellite*, *Bing Map Aerial* layers and cartographic regional data provided by Province Bureau) or added to existing vector data layers (such as *OpenStreetMap*, *Google Streets* and *Bing Road*).

Fig.3 shows the perfect overlapping obtained in Quantum GIS of the Cassini-Soldner layer dealing with the coast area derived from the Cadastre over the OpenStreetMap raster data layer.

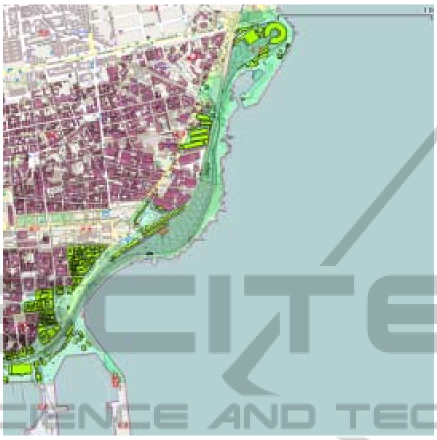

Figure 3: Overlapping of the Cassini-Soldner layer over the OpenStreetMap raster data layer.

This means not only that a SPARQL query may allow us to join the municipal and cadastral data using the relevant shared fields, e.g., the fiscal codes of the owner or the UREU codes of the estates, but also that the vector drawings of the cadastral entities involved in the problem at hands (e.g., buildings, roads, zones) may be represented in any GIS as entities coloured depending on the theme. For example, we may visualize on Google Maps in red the buildings that contain commercial activities that are in arrears if one is checking tax evasions, or in yellow the roads that are not covered by regular waste collection and in orange the downsized schools if one is studying the quality of the services offered to the citizens.

4 QUERYING E-GOV DATA & MAPS

The semantic web gives to a developer two major choices when using distributed databases for querying in a semantic manner: the former is to copy the entire amount of data in a unique knowledge base, the latter is to perform a distributed query. Each of them is affected by two disadvantages: latency and scale, but for the Public Administrations that have usually to manage large databases and big amount of data, copying an entire data store is not feasible. Thus in our approach we adopted the distributed model already shown in fig.2 to allow the users to query two or more RDF knowledge bases using a formula expressed in SPARQL, or to combine relational databases to be queried with a D2RQ mapping language that translates the SPARQL query to SQL on the fly. Fig.4 shows the result of the following SPARQL query to obtain the drawing of the buildings of an area whose centroid returns, when clicked, the building tax payment status:

```
 SELECT ?Person ?Surname ?Name
?Street ?PropertyTaxReport ?Parcel ?Centroid
 WHERE {
   ?Person a foaf:Person;
     vocab:TaxpayerCode ?code;
     foaf:lastName ?Surname;
     foaf:firstName ?Name;
     vcard:street-address ?Street .
   ?PropertyTaxReport a
     vocab:propertyTaxReport;
     vocab:TaxpayerCode ?code;
     vocab:Page ?Page;
     vocab:Parcel ?Parcel.
   ?Centroid a geo:point;
     vocab:TaxpayerCode ?code;
     vocab:Parcel ?Parcel .
   FILTER (?Page = 8)
        } ORDER BY ?Surname.
```

Let us note that the same result may be obtained by executing the above query in either the former and the latter scenario with similar performance. Better performance may be achieved if the query may reuse previous results stored on an RDF cache.

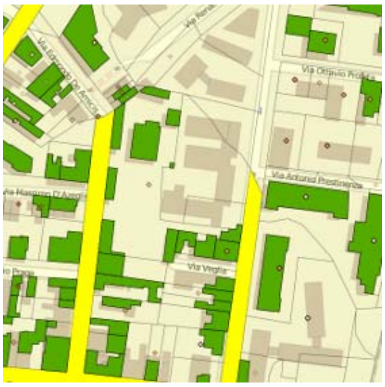

Figure 4: Results of the SPARQL query as a vector data layer on Quantum GIS. The raster data is derived from OpenStreetMap, the green polygons represent buildings, the yellow highlighted sub-layers of the parcel represents the streets. The points return, if clicked, geo-referenced tax information.

5 CONCLUSIONS

Related works to the subject of the paper deal mainly with cadastral system interconnection and introducing spatial dimensions to RDF schemas.

An ontology architecture for the land administration domain, targeted to achieving semantic interoperability between cadastral systems is proposed in (Bošković, 2010). The architecture complies with both Geospatial and Land Administration standards. An example of extension dealing with the specificities of their national cadastre is also provided. In (Hay, 2010), a Semantic Web approach is proposed to customize applications in the Land Administration Domain. Also, an OWL layered architecture adaptable across jurisdictions is outlined. Both these works differ from our proposal in the conversion from relational data models to ontology and in the presence of an explicit ontology alignment step.

The work of introducing spatial dimensions to the semantic web is described in (Auer, 2009) where it is demonstrated how crowd sourced geographical data transformed into RDF can be interlinked (mapped) with other (spatial data) sets to enable spatial data web applications. A similar work, i.e., (Della Valle, 2010) focuses explicitly on some computational issues influencing the use of GIS from the semantic web standpoint, and proposes to treat GIS as virtual RDF graphs instead of re-implementing GIS functionalities in semantic web frameworks. This is achieved through an extension of the D2RQ mapping language to include spatial data types.

Therefore, the application scenario addressed in this work, i.e., integration of cadastral systems with citizen data, has not been tackled before.

Currently, the proposed ontology based methodology to manage data and maps using RDF schemas is being experimented in a project promoted by the Sicily Region, named K-metropolis, for supporting the municipal Governments not only in defining the local taxation policy but also in the transition from the current not interoperable GIS platforms to the implementation of a spatial data infrastructure to integrate data and maps of different municipalities.

For example, we plan to support the municipal Governments to define city master plans and civil protection policies by interconnecting both public and private data stores to the drawings derived not only from the Cadastre but also from any relevant institutional CAD system.

This will allow the municipal Government not only to draw the intended land use and the emergency plans over any raster background, such as aerial photogrammetry or satellite images, but also to represent these plans by a standard graphical notation immediately understandable by any other involved organization.

REFERENCES

Auer S., Lehmann J., Hellmann S., 2009. LinkedGeoData: Adding a Spatial Dimension to the Web of Data, in A. *Bernstein et al. (Eds.): ISWC 2009*, LNCS 5823

Bizer C., Heath T., Berners-Lee T., 2009. Linked data - the story so far, *Int. J. Semantic Web Inf. Syst*., vol.5(3).

Baader F., Calvanese D., McGuinness D.L., Nardi D., Patel-Schneider P.F., 2003. The Description Logic Handbook: Theory, Implementation, Applications, *Cambridge University Press, Cambridge*, UK, 2003.

Boškovic D., Ristic A., Govedarica M., Przulj D., 2010, Ontology Development for Land Administration", IEEE 8th Int. Symposium on Intelligent Systems and Informatics, SISY.

Corlan M., 2009. Flash Platform Tooling: Flash Builder, *Adobe*.

Costanzo A., Faro A., Giordano D., Venticinque M., 2012. Wi-City: A federated architecture of metropolitan databases to support mobile users in real time, Int. Conf. on Computer and Information Science, ICCIS, *A Conference of World Engineering, Science and Technology Congress, ESTCON.*

Costanzo A., Faro A., Giordano D., 2013. WI-CITY: living, deciding and planning using mobiles in Intelligent Cities, *3rd International Conference on Pervasive and Embedded Computing and Communication Systems, PECCS*, Barcelona, INSTICC.

Crisafi A., D. Giordano D., C. Spampinato C., 2008. GRIPLAB 1.0: Grid Image Processing Laboratory for Distributed Machine Vision Applications, Proc. 17th *IEEE Int Conf on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE '08, IEEE.*

David M., 2011. Developing Websites with jQuery Mobile, Focal Press.

Della Valle E., Qasim H.M., Celino I., 2010. Towards Treating GIS as Virtual RDF Graphs. Proceedings of the *1st International Workshop on Pervasive Web Mapping*, Geo-processing and Services, WebMGS.

Faro A., Giordano D., Musarra A., 2003. Ontology Based Mobility Information Systems" Proc. *of Systems, Men and Cybernetics Conference, SMC'03, vol.3, 4288-4293, IEEE.*

Faro A., Giordano D., Spampinato C., 2008. Evaluation of the Traffic Parameters in a Metropolitan Area by Fusing Visual Perceptions and CNN Processing of Webcam Images. *IEEE Transactions on Neural Networks, Vol. 19 (6), IEEE.*

Faro A., Giordano D., Spampinato C., 2011. Integrating Location Tracking, Traffic Monitoring and Semantics in a Layered ITS Architecture, *Intelligent Transport Systems, vol.5(3), IET.*

Faro A., Giordano D., Spampinato C., 2011. Adaptive background modelling integrated with luminosity sensors and occlusion processing for reliable vehicle detection", *IEEE Transactions on Intelligent Transportation Systems*, Vol.12(4)

Hartl M., 2011. Ruby on Rails 3, Addison Wesley.

Hay G:C., and G. B. Hall G.B., 2010. A Semantic Web Approach to Application Configuration in the Land Administration Domain, FIG Congress 2010 *Facing the Challenges - Building the Capacity Sydney.*

Hayes. P., 2004. RDF Semantics. W3C Recommendation 10, Available on line at http://www.w3.org/TR/rdf-mt

Nebot. V., Berlanga. R., 2012. Building data warehouses with semantic data, Decision Support Systems, Vol.52(4). Available at http://sparql-wrapper. sourceforge.net/.

Prud'hommeaux E., Seaborne A., 2008. SPARQL Query 1Language for RDF", W3C Rec. 15.1.2. Available on http://www. w3.org/TR/rdf-sparql-query/.

Saha P., 2010. Government Enterprise Architecture Research Project, NUS Systems Science Inst., http://unpan1.un.org/intradoc/groups/public/documents/unpan/unpan039390.pdf

Sheth, A., 2005. Enterprise applications of semantic web, *IFIP International Conference on Industrial Applications of Semantic Web (IASW2005),* Jyväskylä, Finland.

Zhai, J., Jiang, J., Y. Yu, Y., Li J., 2008. Ontology-based Integrated Information Platform for Digital City", *IEEE Proc. of Wireless Communications, Networking and Mobile Comp., WiCOM '08.*