

RDB Data Triplification vs. Ontologies: The Case of Municipal Registry Data

Stamatios A. Theocharis*, George A. Tsihrintzis
Department of Informatics, University of Piraeus, Greece.

* Corresponding author. Email: stheohar@unipi.gr
Manuscript submitted August 11, 2015; accepted January 20, 2016.
doi: 10.17706/ijcee.2016.8.1.44-56

Abstract: Nowadays the individualized provision of electronic services according to the needs of each user has become particularly important. Particular interest is acquired by the "intelligent" information retrieval with the application of innovative primary data management methods. Motivation for this work, has been the data located mainly in databases developed and managed by the public sector and associated with the development of e-Government. Its integration relies heavily on the integration of innovations that support intelligent information search and data interface in order to feed the various public organizations and governmental portal. In particular, this paper examines the possibility of converting information stored in relational databases in RDF triples that may subsequently be treated with semantic web technologies and fed into cognitive connected databases. We also present a case study, namely the development of an ontology for describing the assets held in the national population register of the Ministry of Interior of the Greek Government and the municipalities of Greece. The knowledge base associated with this ontology can be fed with the transformed data of the existing relational databases using appropriate software.

Key words: Municipal registry, ontologies, RDB triplification.

1. Introduction

Nowadays, the majority of citizens and businesses use internet applications in order to have access to primary electronic data or information published on various websites. Moreover, most of the public sector has digitized raw data held and has isolated internet users (citizens, businesses, and public organizations) as open data, processed at relative governmental portals. In most cases, information retrieval is achieved with the help of mainly relational databases and related applications supporting access to these databases through the existing Internet.

At the same time, the interest of the scientific community focuses on further development of the Internet and the evolution of the so-called semantic web. Initially, the semantic web is expected to contribute to smarter access and manage information handled on the internet through new technologies and development of corresponding new applications. The overall vision is that of the transition from the current Internet of static pages into a dynamic network of providers of services (Web services) that automatically discover the information sought, taking into account the semantics of the concepts referred to by the user. This will benefit in improving the quality of the answers given to specific questions-searches in each case. Thus, the research interest is focused on developing tools and methods based on the development and management of common vocabularies and new standards for representing information

and knowledge produced.

Our interest focuses on the opportunities that can provide the technologies of the semantic web in the further development of e-government and support of open government. The field of e-government can be one of the main consumers of smart services based on the Semantic Web. This is because the modern trend of open government is supported both by the publication of "open data" to the public and also by the connection of digitized data and information in the back-office of public administration. By using semantic web technologies, it is expected that the search and data sharing between stakeholders will improve by the use of common terminology and semantics.

More specifically, in this work we examine the usefulness of processing the data stored in existing relational databases with the help of semantic web technologies. Our objective is to develop ontologies based on common vocabulary for describing each field of interest and then to feed them into the knowledge base of existing data. This is particularly important because until now, a huge volume of digitized data stored in relational databases has been developed and their depreciation and their replacement by other technologies may prove not to be economic in terms of time and money.

2. Motivation and Previous Related Work

Motivation for this work was provided by problems of the existing system when searching data from the database of the national population register used by municipalities of the country and the Ministry of Interior. Our interest is mainly focused on exploitation of the data held in existing relational databases to extract information based on dynamic queries. An example of such a query is the search for the nearest relatives of someone, excluding his/her immediate family. Existing database systems that support the management of public information are currently using relational databases. Searches in these databases are usually based on standardized and inflexible questions that have been designed beforehand and may not be personalized/individualized [1]. Their configuration and adaptation to new needs usually requires the assistance of managers/developers of the database. The process is time-consuming and uneconomical. Another serious problem is the compliance of local databases in municipalities with the central database in the Ministry of Interior. Although procedures of information synchronization are provided by the central database of the Ministry, there is still a possibility of existence of conflicting, incomplete or incorrect data in two or more public bodies.

The situation described above can be handled by technologies of the Semantic Web in which information is encoded in the RDF triples format and can be accessed by software-based questions in SparQL. This way, we can implement intelligent and dynamic questions that seek information based on specific properties and rules and not simple data control carried out in traditional relational databases. Our interest focuses on making full use of existing databases and data and not redesigning a system from scratch.

The number of semantic web tools is widening and there is already great interest in developing data conversion technologies from traditional forms into formats that can be managed by the technologies of the Semantic Web. These technologies are mainly based on programming languages like Java and Python. From the published work on the conversion of data from relational schemas in RDF triples, we indicatively present the following.

In [2], the authors analyze the various ontology construction approaches from relational databases and identify the advantages and disadvantages of these techniques, so that an enhanced and efficient approach can be proposed. The authors have performed detailed comparative analysis of various ontology construction techniques from relational database (RDB) based on database schema analysis (meta-data, cardinality restrictions and datatype information) and stored data (through data mining). The authors in [3] propose a particular process that integrates a design rationale approach with a triplication strategy. The process

supports the reuse of standard RDF vocabularies recommended by W3C for publishing datasets and automatically collects the entire rationale behind the ontology design, using a specific vocabulary. In [4], the authors present advanced features of RDB2OWL mapping specification language that allows expressing RDB-to-RDF/OWL mappings in a concise and human comprehensible way. Relative with the previews is [5] which describes a practical approach about the transformation from relational database to RDF/OWL format. In [6], an Eclipse plug-in is introduced that supports the standard RDB to RDF Mapping Language (R2RML) to produce Direct Mappings in RDF. On the other hand, [7], [8] survey current techniques, tools and applications for mapping between Relational Databases (RDB) and the Resource Description Framework (RDF). In [9], a direct mapping is defined from relational data to RDF. This definition provides extension points for refinements within and outside of this document. This document has been reviewed by W3C Members, software developers and other W3C groups and interested parties and is endorsed by the Director as a W3C Recommendation. It is a stable document and may be used as reference material or cited in another document. W3C's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability of the Web.

Finally, in [10] links are provided to applications that are essential references of the thematic data conversion to relational databases in RDF triples, among which the application D2RQ which we used in this study.

3. Semantic Web Technologies and Linked Data vs. Relational Databases

3.1. Semantic Web and Linked Data

The Semantic Web started as an initiative of the World Wide Web Consortium - W3C [11] and the tools that have been developed are now expanding the existing Internet. The basic achievement of tools supporting the semantic web is the support of automated discovery, access, processing and sharing data by humans and computers. The key to achieving this objective includes metadata and ontologies. The metadata clarifies data, providing the information necessary for the importance they have to users. Ontologies are the most important tool in the attempt to describe the knowledge and establish a common vocabulary in each field.

The term Linked Data sets out the basic principles governing the publishing and connecting structured open data online. The principles introduced by Tim Berners-Lee [12] include a) the use of URIs as identifiers for objects of the natural world, b) the use of HTTP URIs as identifiers in order to ease the search by humans, c) the provision of useful information under specific standards (RDF, SPARQL) when someone searches for a URI and d) the connection between URIs order to enable transition and hence finding new information.

The interconnected (linked) open data is a method of publishing structured data the main objective of which is to connect via their real relationship or their importance seemingly "irrelevant" information pieces. Many times the same concept is used in different verbal ways (vocabulary) in different sources, resulting in inefficient information search. The success of the project is based on wide acceptance and use of technologies used (HTTP, URIs) to achieve effective retrieval of information between websites in a way processable by computers.

The dominant idea of Interconnected Data is the application of the existing architecture of the Internet in sharing structured data worldwide. The current form of the Internet is based on the concept of URIs (Uniform Resource Identifiers) as a global mechanism for unique identification of objects and on the HTTP protocol (Hypertext Transfer Protocol) as a global mechanism for access to online documentation and the HTML language as the dominant format for the submission of data. Furthermore, there is the idea of linking web documents. i.e. files that can be stored in different locations (servers). The links between documents convert the scattered content to a global information space. The URIs are not used for identification only of digital content, but also as identifiers of people, places, or even abstract concepts. Thus, as is now accepted

that almost every public organization has its own URI, (for example, the Interior Ministry uses <http://www.ypes.gr>), it is possible to identify a person with a relative URI, such as his personal website. Also, abstract concepts can have their own URI. For example, the concept of "know someone" may be determined via the URI <http://xmlns.com/foaf/0.1/knows>. In a similar way, we can state any other relations between people, such as the relationships recorded in national registries and the corresponding relationships that arise.

The HTTP protocol is a widespread mechanism for access to online content. In the current form of the Internet HTTP, URIs are used to uniquely identify all forms of available information. Therefore, Linked Data encourage the use of HTTP URIs to identify the entities of the physical world. In fact, the HTTP protocol makes it possible to search for these entities.

The third principle of interconnected data requires the representation of structured information in a common format. This form is usually the Resource Description Framework (RDF) which is a model based on the logic graph.

The connection of the digital data is the fourth principle of interconnected data. Thus, the general direction is not only the connection between online documents, but also between different types of objects. In contrast to existing unstructured internet connections without semantic content, the links between the entities identified by URIs become important. This means that links can describe relationships of the natural world. For example, it is possible to connect a man and an organization in order to describe the employment relationship between the person and the organization. Links to Linked Data are called RDF links in order to separate out the links between the online documents. Therefore, in correspondence with the global information space of the current Internet, we are led into a global data space.

3.2. Semantic Tools

The Semantic Web is based on a specific architecture that includes various languages and standards such as:

- XML (Extensible Markup Language): It is a data description language for data which is easy to be read and edited by humans and programs. It imposes no semantic restriction in describing the data.
- XML Schema : It is a language that defines the structure of XML documents.
- RDF: This is a metadata description and processing model
- RDF Schema: It is a mechanism for describing resources and the relationships between them and a semantic extension of RDF.
- OWL: Provides a way of describing the terms and relationships around an area of interest, providing more powerful syntax of RDF and RDF Schema and stronger semantics based on logic (logic-based semantics).

3.2.1. RDF files

In particular, the Resource Description Framework (RDF) is a W3C standard for simple description of metadata on the Internet, which enables the implementation of data graphs and their sharing between users and machines. The basic concept of the representation of knowledge is the resource. In this sense we mean any entity of the Web, such as a website, a part or a set of web pages, electronic files or even objects that are not readily available on the internet, such as a book for example. The description of the resources of the Internet with RDF is primarily focused on the performance of metadata such as the name of the citizen who is registered in the municipal registry, the recording date in the municipal registry, the name of the municipality in which the family register belongs etc.

The general idea is that each resource has one or more properties which have specific values or they may themselves be resources of the Internet. In RDF, such information is given as suggestions in the form of

triples subject — property — object. In this triad, the subject is a resource on the Internet, which is identified by a specific unique URI. The property attributed to the subject is specifically described in another resource of the internet (and therefore is itself a resource). The object can be either a literal or another resource itself. A set of RDF triples can be understood as a graph. In this graph, the objects and subjects play the role of nodes while properties play the role of the edges connecting them. We present detailed examples in the following section.

3.2.2. SPARQL

The SPARQL (SPARQL Protocol and RDF Query Language) is the query language for RDF documents which has become a standard of the W3C since January 2008 [11]. Its characteristics include the requirement that the RDF data be treated as simple triads without some shape or another ontological information unless this has been explicitly defined in the RDF document. This language is SQL-like, i.e. it imitates the ANSI SQL syntax style used in relational databases. The basic building block of the SPARQL is the triple pattern and it is an individual (elementary) question. A standard triplet is actually a RDF triplet (subject — predicate — object) which may contain a variable in one or more of the three positions. A prerequisite for the question to give a result, in other words for the variable to get a value, is the existence of pairing with some RDF triple of the document to which the question is posed.

A typical example is the question of the form :

SELECT? Subject? Predicate? Object

WHERE {? Subject? Predicate? Object}

which returns all RDF triples containing the relevant RDF document.

3.2.3. D2R server

In this paper, we used the D2R Server as a tool for publishing Linked Data views of relational databases. The D2R Server [13] is a tool for publishing the content of relational databases in the Semantic Web. The content of the database is mapped into RDF data with a declarative mapping that specifies how the resources are identified and how they produced the values of properties in the content database. Based on this mapping, the D2R Server allows in RDF and HTML browsers to browse the non-RDF database contents and also allows applications to submit queries to a database using the SPARQL query language of submission. The server receives requests from the web and transforms them "on-the-fly" to corresponding SQL queries for submission to the database. The following Fig. 1 illustrates schematically the operation architecture of D2R Server [13].

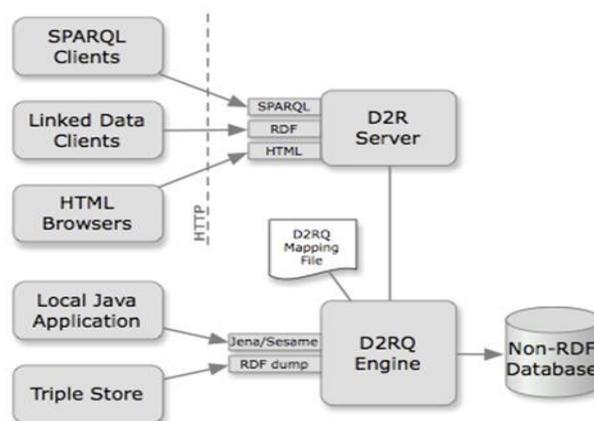


Fig. 1. Operation architecture of D2R server.

3.2.4. D2RQ mapping language

The D2R Server uses D2RQ mapping language to capture mappings between database shapes and RDFS or OWL ontologies shapes. A D2RQ mapping determines how resources are identified and how the values of properties in the content database are produced. An example of the structure of a D2RQ map file is shown in Fig. 2. The central object in D2RQ is ClassMap. A ClassMap represents a mapping from a set of entities described in the database to a class or group of such classes of resources. Each ClassMap has a set of PropertyBridges, which determine how the resource descriptions are created. The values of the properties can be created directly from the database values or by using "molds" or translation tables. The D2RQ supports conditional mappings at ClassMap and PropertyBridge, mapping $n: m$ relations and handling highly normalized table structures where entities descriptions extend to several tables.

The D2R Server includes a tool that automatically generates a D2RQ file mapping the structure of a database table. The tool generates a new RDF vocabulary for each database using the names of tables as class names and column names as property names. The mapping may be adjusted subsequently, replacing the automatically generated terms with terms of familiar RDF vocabularies [14].

4. Triplification — The Case of Production of RDF Triples from Data in Relational Databases in National Municipal Registry

4.1. Presentation of the Field of Interest

As already mentioned above our interest focuses on finding information on semantics using data from existing databases. This is because the retention of data in relational databases is the most widespread solution in most cases of public data regardless of their size and type. Their main disadvantage is that the stored data does not have the semantics of the original information so the questions submitted to databases are static and non-personalized to the needs of each user. In contrast, the transformation of this data into a format that supports the performance semantics will support the presentation of dynamic queries personalized to each user.

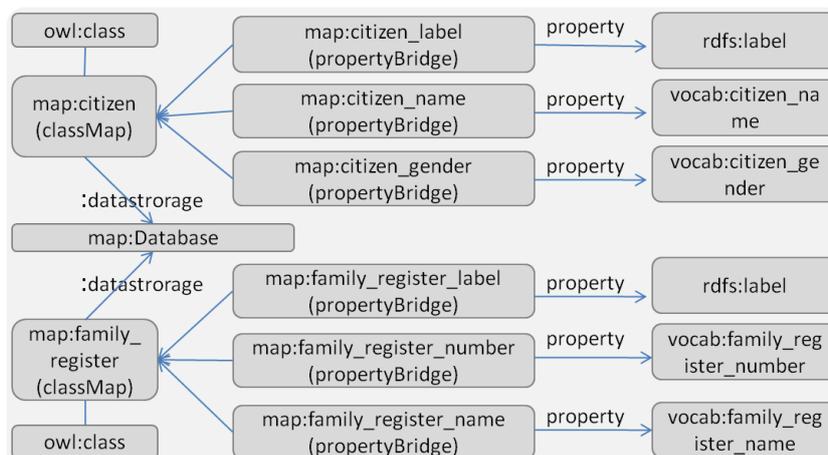


Fig. 2. The structure of an example D2RQ map [15].

In this study, we examined the case of the National Population Registry kept by the Ministry of Interior and which is accessible by the country's municipalities. The system that supports the management of features of Registry has the architecture shown the Fig. 3.

According to this architecture a central database is maintained in the Ministry of Interior while databases are also maintained in the municipalities. It is the responsibility of the municipalities to synchronize local databases with the central database kept by the Ministry. More specifically, the Ministry of Interior hub also provides all the Web-Services to public bodies and citizens calling its services electronically via a web form

that is installed in the portal of the Ministry of Interior. The server host has the ultimate oversight of the system and checks every kind of transaction of the citizens with the Population Register. The data of all the country's citizens are stored in a central database which communicates with the application of the Ministry of Interior server. The central database operates independently of the existence of other databases in the various municipalities-clients and does not require any kind of communication with these bases. It is the responsibility of the municipalities to update their information by periodically calling for updates and downloading service provided by the Ministry of Interior. In this scenario, all transactions start with the initiative of the customer who makes a request for a service from the Ministry of Interior and receives in response the results of the application.

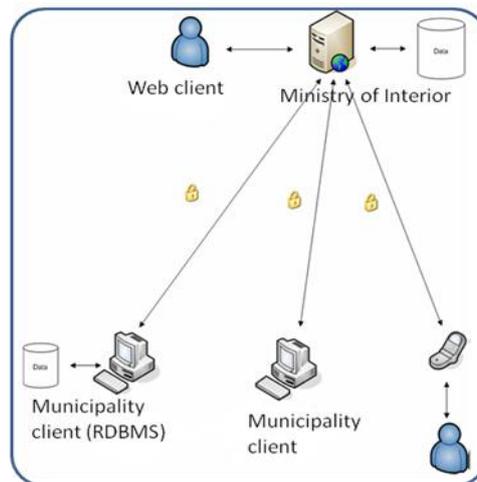


Fig. 3. The architecture of municipality registry RDBMS.

Experience to date of the databases on specific areas of e-government, such as the system introduced earlier, presents serious problems. Specifically, there are problems related to compliance with common vocabulary and common rules in the management of records in local bases. Thus, for example problems are commonly reported while seeking cases of double surnames or duplicate names of citizens. Also, using sometimes Latin and sometimes Greek characters or the abbreviations of people's names are commonly reported. For example instead of the full name "Constantine" someone appears as "Con/ne" while elsewhere the shorter version "Costas" is used. Such problems pose a risk of malfunction of the existing system.

Another challenge is the search for a citizen's family tree in order to find his/her closest relatives other than the direct relatives in his/her immediate family. The current system gives authorized users the ability to view only register status events that concern them and choose the register status documents which they can print either individually or en masse. According to these prints, the Municipal Registry employee can take the necessary actions in the Population Registers, such as deleting deceased from population registers, removing newly-weds from family register, doing neonatal entry in the family register, etc. [1]. Thus, in order to find the family relationships between citizens, it is necessary to present specially designed questions to existing databases to find possible links between members of different family registers to all the country's municipalities. The search is based on the entries and deletions from family register which unfortunately does not record any family relationship except that of spouses and children. Thus, parallel tables must be maintained in order to follow changes in Population Registers. This, however, results in storing unnecessary information or duplicates. Conversely, as we shall see below, the process of finding immediate family members of any degree with the assistance of the ontology is treated with a single DL-query or a corresponding SparqL query on the knowledge base maintained by the ontology.

4.2. Triplification — Steps and Methodology

The most systematic and efficient processing of heterogeneous data may be achieved when such data is transformed in the same format, without losing information. Furthermore, it is desired to assign semantics to the different data and to connect them. This increases the interoperability of the system data. An interesting and particularly easy to use data format that serves the above is the formatting to RDF triples. The conversion of data not in RDF ("triplification") is a particular process in a heterogeneous data integration system. Of particular interest is the automated production of RDF files of different formats such as .XLS, .CSV and .JSON, which are some of the forms of open data in public administration, but also from tabular data or data stored in relational databases.

The proposed conversion process is based on data stored in databases such as MySQL or other supporting SQL [15]. The output relates to the form of RDF / XML or Turtle, in order to reduce the file size produced in parallel with the complexity. Our interest concerns the extra yield of semantics to the data through the production of RDF triples. The basic tools of the methodology are shown in Fig. 4.

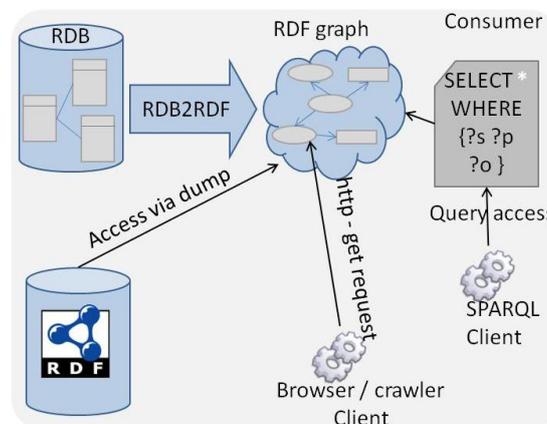


Fig. 4. The basics tools for RDB data triplification.

While the basic steps of the process are:

- a) Data analysis & Data Model
- b) Relational DataBase Creation
- c) RDF Graph Creation
 - 1) Publishing tool selection
 - 2) Definition and assignment of URI's
 - 3) Concept schemes definition
 - 4) Defining vocabulary
 - 5) Data linking
 - 6) Access in data using a simple browser

4.3. Detailed Steps

In this paragraph, we present the first steps in order to "build" a simulation of the database of the Ministry of Interior and the corresponding ontology and subsequently to transform relational data into semantically linked data.

4.3.1. Creating a relational database

In the environment of MySQL (in particular, the MySQL Workbench tool), we created a simulation of the database of the national population register with a model containing the key elements of the register, so that we could evaluate the process of conversion to RDF triples. The key elements were grouped into respective

tables which present the following:

- Citizen, for the essential elements of each citizen
- Municipality, for the details of each municipality of the country
- Family_register, for the elements of family-owned register of each citizen
- Family_register_alterations, to monitor changes in family register, and
- Marriage, for storing information relating to marriages of citizens.

In the graphic environment of the tool we created the appropriate relations that appear in Fig. 5 in order to correlate the data of the tables above. Then, we fed the database with "synthetic" data and extracted the base data into a dump file.

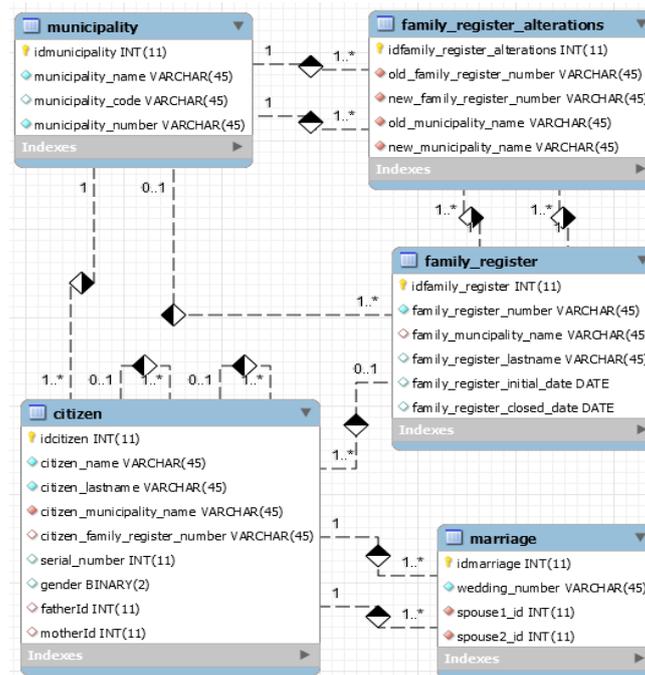


Fig. 5. The EER diagram of RDB municipal registry data.

4.3.2. Triplification using D2R server

To recover the data from our relational database and convert them to RDF triples, we used the package D2RQ-0.8.1, which works directly with MySQL data and provides among other tools the generate-mapping and D2R-server. The generate-mapping tool creates a D2RQ mapping file by analyzing the schema of an existing database. This mapping file, called the default mapping, maps each table into a new RDFS class that is based on the table name while it maps each column into a property based on the column name. This mapping file can be used as-is or can be customized. The command that we used to activate the tool was provided by the command line as:

"generate-mapping -o outputFileName.ttl -u root -p SqlServercode jdbc:mysql:///MySQLdatabaseName". Then, we activated the D2R-server with the command "d2r-server -p 8080 outputFileName.ttl" and with the help of a simple browser we went at "http://localhost:8080/snorql". The result of the process is shown in Fig. 6.

5. Design — Implementation of the Ontology of the Population Register

To address the difficulties mentioned above by approaching relational databases in the Protégé 4.3 environment in the context of this work, we created an ontology in order to demonstrate the retrieval of

information concerning family relationships between citizens through appropriate DL queries. Specifically, we created a basic ontology (as a simulation of an integrated ontology to be presented in following work) with three basic classes — similar to those of the relational database. We then provided our ontology with the necessary SWRL rules shown in Fig. 7 and the properties shown in Fig. 8. We see that a very simple rule of natural language that expresses "if X has Y as a parent and Y is a sister of Z, then Z is an aunt of X" has been translated in SWRL to declare when someone "Z is an aunt of X" with the statement "hasParent (? x,? y), hasSister (? y,? z) -> has Aunt (? x,? z)". Similar are the other statements of ontology. As shown at the tab of the properties of the ontology, the family relationships that interest us in this work but also the fact of the existence of family register for one person, have been implemented in the form of properties with suitable domains and sets of values.

SPARQL results:

s	p	o
db:citizen/2	vocab:citizen_citizen_municipality_name	db:municipality/1
db:citizen/1	vocab:citizen_citizen_municipality_name	db:municipality/3
db:citizen/4	vocab:citizen_citizen_municipality_name	db:municipality/3
db:citizen/5	vocab:citizen_citizen_municipality_name	db:municipality/3
db:citizen/7	vocab:citizen_citizen_municipality_name	db:municipality/3
db:citizen/8	vocab:citizen_citizen_municipality_name	db:municipality/3
db:citizen/3	vocab:citizen_citizen_municipality_name	db:municipality/2
db:citizen/6	vocab:citizen_citizen_municipality_name	db:municipality/2
db:citizen/3	vocab:citizen_fatherId	db:citizen/1
db:citizen/4	vocab:citizen_fatherId	db:citizen/1
db:citizen/5	vocab:citizen_fatherId	db:citizen/1
db:citizen/7	vocab:citizen_fatherId	db:citizen/4
db:citizen/8	vocab:citizen_fatherId	db:citizen/4
db:citizen/1	vocab:citizen_citizen_name	"Telis"
db:citizen/1	vocab:citizen_gender	"Man"
db:citizen/1	vocab:citizen_serial_number	1
db:citizen/1	rdfs:label	"citizen #1"
db:citizen/1	vocab:citizen_citizen_lastname	"Theoh"
db:citizen/1	rdf:type	vocab:citizen
db:citizen/2	vocab:citizen_citizen_name	"Olga"
db:citizen/2	vocab:citizen_gender	"Woman"

Fig. 6. The export of D2R server in out browser within Sparql query.

In order to test the correctness and consistency of our ontology, we fed it manually with synthetic data and then submitted the ontology to be evaluated by the activation of reasoners available. We also submitted specific questions to DLQuery Tab. As an example we refer the query in natural language "who has MariaTheo as an aunt?". In the environment of DL query, we implemented the statement "Citizen and hasAunt value MariaTheo" and its results are shown in Fig. 9. Clearly, the ontology is relevant.

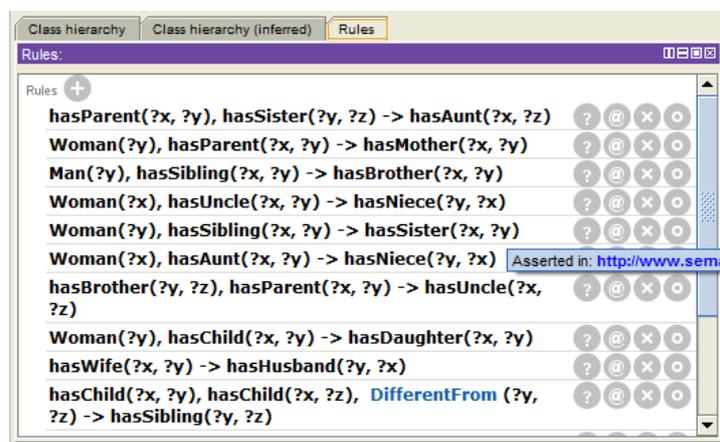


Fig. 7. Swrl rules in Protégé 4.3.

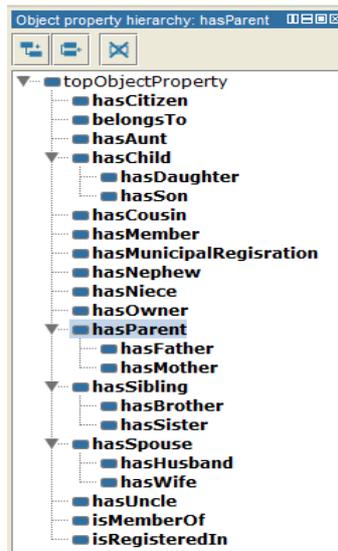


Fig. 8. Object properties of ontology for municipality registry.

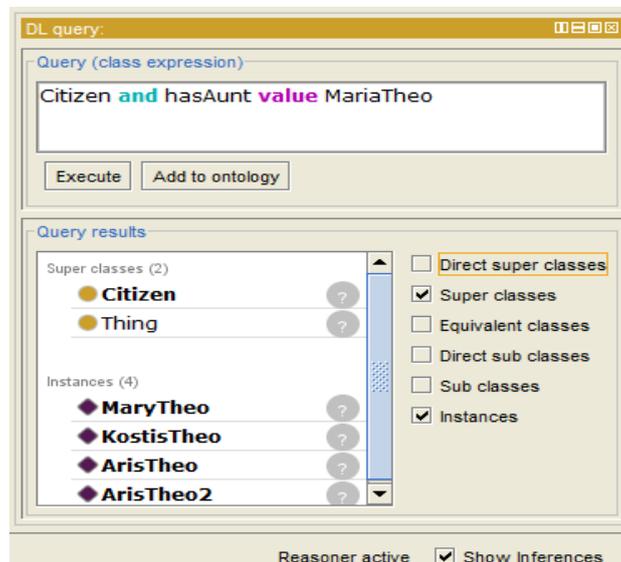


Fig. 9. A DL query in Protégé 4.3 and its results.

6. Conclusions and Future Work

The two main approaches to data management, namely relational databases and semantic web technologies, have been implemented simultaneously and shown to operate complementarily in managing public data.

The two technologies manage data differently. Relational databases manage the data simply by checking their content while returning data instances (snapshots of their content) to the user; but do not recognize or support their semantics. Another difference relates to the knowledge that can be generated from the stored data. Knowledge acquisition from a relational database or a knowledge base is often associated via the submission of relevant questions through appropriate interfaces. The answer to a question on the Semantic Web, is not just data control (as in databases), but it also requires the implementation of complex considerations that run a huge number of classes and instances. The classes and instances that are stored in an ontology are linked via stated rules and constraints that describe specific events and roles during transfer from natural language. For this reason, it is particularly important to combine the facts and concepts that are

declared in an ontology so conclusions about the correctness or otherwise of another event can be drawn or new unexpected events can be produced. This process is described as reasoning and is one of the main differences between databases and knowledge bases of the semantic web. Of particular interest is the sub-process of reasoning concerning the production of new events in the ontology, referred to as inferencing, which are added to existing events for future use.

But the fact that there is a large amount of data on the web stored in relational databases which may be obsolete, induces a high cost in redesigning a new system from scratch. It is therefore very important to produce semantic data using the already stored data in whatever format they are. To achieve the enrichment of the Semantic Web with real data, one should first map the data in relational schemas into the elements of each ontology.

Our future plans include the development of an application to automatically convert the output conditions of the RDF file D2R-server in conditions arising from mutually agreed vocabulary which will be described by a corresponding ontology and which will be associated with knowledge base which will be supplied automatically from existing relational database. This work is currently under way and will be presented elsewhere in the near future.

Acknowledgment

The publication of this paper has been partly supported by the University of Piraeus Research Center.

References

- [1] Ministry of Interior, Greece. (2013). *Circular of Ministry of Interior for the National Population Register*. From <http://www.ypes.gr>
- [2] Pasha, M., & Sattar, A. (2012). Building domain ontologies from relational database using mapping rules. *IJIES*, 5(1), 20-27.
- [3] Berardi, R., *et al.* (2013). StdTrip+K: Design rationale in the RDB-to-RDF process. In H. Decker, *et al* (Eds.), *DEXA 2013*, Part I: LNCS 8055 (pp. 303–310).
- [4] Būmans, G., & Čerāns, K. (2011). Advanced RDB-to-RDF/OWL mapping facilities in RDB2OWL. In J. Grabis, & M. Kirikova (Eds.), *BIR 2011*, LNBIP 90 (pp. 142–157).
- [5] Būmans, G., & Čerāns, K. (2011). RDB2OWL: A practical approach for transforming RDB data into RDF/OWL. *Proceedings of I-SEMANTICS 2010*, ACM.
- [6] Salas, P. E., *et al.* (2011). RDB2RDF plugin: Relational databases to RDF plugin for eclipse. *Proceedings of TOPI'11*.
- [7] Hert, M., *et al.* (2011). A comparison of RDB-to-RDF mapping languages. *Proceedings of I-SEMANTICS 2011, 7th Int. Conf. on Semantic Systems*.
- [8] Sahoo, S. S., *et al.* (2009). A survey of current approaches for mapping of relational databases to RDF. From http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport_01082009.pdf
- [9] Arenas, M., *et al.* (2012). A direct mapping of relational data to RDF, W3C. From <http://www.w3.org/TR/rdb-direct-mapping/>
- [10] W3C. (2012). Implementations. From <http://www.w3.org/2001/sw/rdb2rdf/wiki/Implementations>
- [11] The W3C Organization. From <http://www.w3c.org>
- [12] Lee, B. T. (2006). *Linked Data - Design Issues*. From <http://www.w3.org/DesignIssues/LinkedData.html>
- [13] Bizer, C., & Cyganiak, R. (2015). D2R Server: Accessing databases with SPARQL and as linked data. From <http://d2rq.org/d2r-server>
- [14] Cyganiak, R., *et al.* (2015). The D2RQ mapping language. Retrieved from <http://d2rq.org/d2rq-language#examples>

[15] Vergoulis, A., *et al* (2014). Data Governance in the era of the Web of Data: generate, manage, preserve, share and protect resources in the Web of Data. Technical Report.

Stamatis A. Theocharis was born in Athens. He is a graduate of the Department of Mathematics, University of Ioannina, Greece and a post-graduate of the Department of Informatics, University of Piraeus, Greece. He is currently a Ph.D. candidate in the domain of e-government and open data analysis with his supervisor Professor George Tsihrintzis. His current professional occupation concerns the object of the Ministry of Interior of Greece. Open data, semantic web and the technologies that support it and the e-government technologies are included in his research interests.

George A. Tsihrintzis is a member of the Council, the director of Graduate Studies in 'Advanced Computing and Informatics Systems' and a full professor in the University of Piraeus, Greece. He received the diploma of electrical engineer from the National Technical University of Athens, Greece (with honors) and the M.Sc. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, Massachusetts, USA.

His current research interests include pattern recognition and machine learning and their applications in multimedia interactive services, user modeling, knowledge-based software systems, human-computer interaction and information retrieval. He has authored or co-authored over 300 research publications in these areas, which include 5 monographs and 13 edited volumes.