Integrating Distributed Data Bases in a Semantic Framework
The K-Metropolis Project

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

Keywords: e-Government, Semantic Web, Location based Services, Distributed Database Systems.

Abstract: This paper presents how data integration is obtained in K-Metropolis, i.e. a project inspired by the Connected Government framework and supported by the Regional Government of Sicily that aims at integrating data residing on personal, commercial and municipal databases of different organizations. In response to the user queries, K-Metropolis suggests the most suitable services by a decision support system illustrated in previous works using all the data available at urban level. E-payment of the recommended commercial services is supported. On request, e-government certificates are provided to the mobile users. When needed, the data and certificates are displayed on a Google Maps based interface. In a companion paper we show how K-Metropolis is able to display on a more powerful GIS not only geo-referenced data of the public services but also thematic drawings that further qualify these data.

1 INTRODUCTION

In the last decade, the adoption of semantic technologies for facilitating data interoperability has been a major endeavour for the most of the companies employing information systems interrelated to data of other companies, typically with the goal to support enterprise market needs or to track goods for food origin control, e.g. (Sheth, 2005) and (Salampasis et al., 2005).

Recently, also Public Administrations (PAs) are more and more interested in integrating all their archives to better plan and control the economic and social activities developing at either regional level or urban scale, e.g., (Zhai et al., 2008) and (Costanzo et al., 2012). In all these systems, individual citizens and resources are represented by Uniform Resources Identifiers (URIs) that are linked together by properties, thus giving rise to a subject-predicate-object graph, aiming at describing individuals or objects by their distinctive features, whereas knowledge needed for the data integration process is encoded in ontologies.

In addition to data integration, achievable through ontologies mappings to manage the data semantic heterogeneity, a significant advantage in resorting to semantic web technologies is the inherently flexible data model upon which the information system rests. The model does not need to be complete and fixed, rather it remains open to changes and refinements, also related to unforeseen ways of consuming data (e.g. the type of queries to be issued).

In fact, novel properties can be easily integrated in the subject-predicate-object graph, that is by its very nature open-ended. This aspect is especially important in highly dynamic scenarios, with changing regulations and policies and, in general, where both structured, semi-structured and unstructured data should coexist.

Traditionally, the challenge of data integration has been tackled with Data WareHousing (DWH) technologies; currently, there is a trend towards leveraging on semantic web technologies for more agile, incremental integration of data, thus overcoming the inflexibility of the traditional DWH model and the boundaries of the single organization.

First examples of DWHs of semantic data are now being discussed, e.g., (Nebot and Berlanga, 2012), to retain the advantage that a DWH affords for performing On-Line Analytical Processing (OLAP) analysis of semantic data and combining it with the inference power of the annotation semantics. On the other hand, data and services integration issues affect also Geographic Information Systems (GIS) technology, whose key role in decision making in the public sector is well

Giordano D., Torre A., Alessi S., Costanzo A. and Faro A..
Integrating Distributed Data Bases in a Semantic Framework - The K-Metropolis Project.
DOI: 10.5220/0004558203220328
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (ICEIS-2013), pages 322-328
ISBN: 978-989-8565-59-4

established. It has been recently remarked that although much research has been done on the integration between DWH, OLAP and GIS, the main shortcoming of the proposed solutions is that they are not open and extensible, whereas the design of the spatial dimensional schemas and of meta-models for the semantic integration among the metadata of such technologies is still problematic, e.g., (Nebot and Berlanga, 2012).

Effective solutions for the integration of all of the above technologies from a semantic web standpoint are therefore topical, since they will be at the core of the new breed of unified information systems needed by advanced e-government and "smart cities" applications that are evolving having in mind the *Connected Government* framework whose aim is "to shift from a model of providing government services via traditional modes to integrated electronic modes wherein the value to the citizens and businesses gets enhanced" (Saha, 2010).

This work presents how the mentioned data integration is obtained in K-Metropolis, i.e. a project supported by the Regional Government of Sicily, to collect data originating from personal, commercial and municipal databases of different organizations to offer suitable e-services to citizens and businesses. When needed, such data are displayed on a Google Maps based interface, whereas the use of a more powerful GIS is discussed in a companion paper (Giordano et al., 2013). Sect.2 illustrates the functional and implementation architecture of K-Metropolis, whereas sect.3 sketches the solution adopted for data integration and the user interface to provide the needed location based information services to desktop PCs and user mobiles.

2 THE K-METROPOLIS PROJECT

The K-Metropolis project focuses on data integration to support urban/metropolitan activities from two main points of view: i) logistics activities and personal mobility based on traffic ontologies, e.g. (Faro et al., 2003), and ii) management of the municipal data bases to collect taxes and to improve the services offered to the citizens. Concerning the first point, K-Metropolis plans to use sensing infrastructures, mainly based on computer vision techniques, e.g. (Faro et al., 2008, 2011a, 2011b) and (Crisafi et al., 2008), to support mobility and logistics information in real-time. Real-time provision is also a requirement taken into account to design data integration to obtain the certificates on citizens' demands sent from fixed and mobile devices.

The available proposals for distributed e-government applications typically consist in the integration of several data bases coded in proprietary format. On the contrary, in the K-Metropolis project, the data are organized in a sort of "Location-aware" Data Warehouse (LA-DWH) where semantic relations link citizens activities to services, e.g., finding and reserving "nearest to me" services such as parks, pharmacies and fuel stations, or on-line issuing of personal certifications.

By this choice, the main ingredients of the above sketched urban/metropolitan information system, i.e., the activities and service flows involving citizens and the relevant location (street, apartment, office, stores, parks, etc.), can be identified by stable URIs such as codes and addresses (e.g., fiscal codes, inventory codes, interest points addresses) and geo-coordinates. Since the K-Metropolis DWH is geo-referenced, the responses to the user queries can be enriched by graphics and maps.

Any real-time, territorial and semantic DWH could be structured by either a centralized or distributed architecture. In the centralized solution, the semantic data of the DWH would be obtained by a staging area that transforms the data coming from the remote SQL tables of the various sources into triples following ontologies and mappings expressed by RDF statements (Hayes, 2004).

However, such centralized solution needs a continuous transformation of the data from the original archives and their uploading into the urban DWH to maintain data coherence. Also, this architecture does not satisfy reliability and privacy constraints, and places a high computational load on the central system. For this reason, K-Metropolis adopted a distributed architecture consisting of a Cloud of computing nodes, as shown in fig.1, where the data are kept on the source nodes in both SQL and RDF formats.

In this architecture the procedures used to *Extract-Transform-Load* (ETL) the information from the distant data sources into the database of each relevant node can be either off-line or real-time procedures. The latter approach is mandatory to update the DBs of the distributed semantic DWH devoted to control dynamical urban systems, e.g., traffic flow optimization by modifying suitably the traffic lights cycles or first aid interventions by suggesting the best routes to the ambulances (Costanzo et al., 2013). In this architecture, the central server is used only to manage directories, accounts and error recovery, whereas the data marts

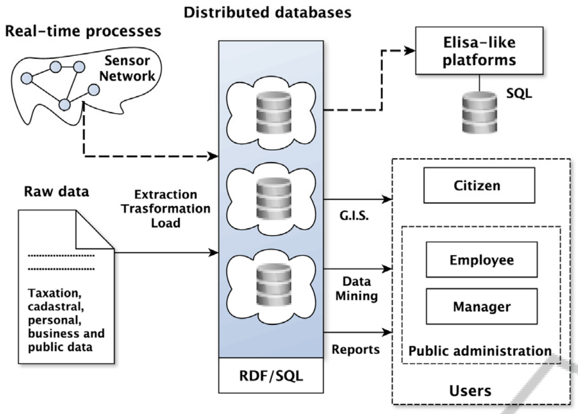

Figure 1: K-Metropolis: functional architecture.

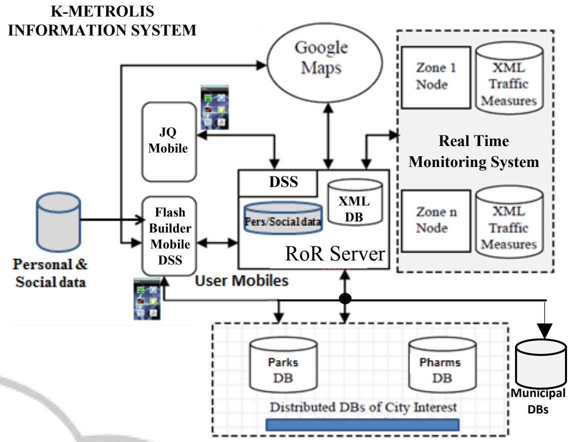

Figure 2: K-Metropolis implementation architecture.

are built by using distributed queries issued in a user friendly format by a fixed or mobile device. Such queries are converted into a language, i.e., SPARQL 1.1 (Prud'hommeaux and Seaborne, 2008), suitable to extract responses from the triple stores of the Cloud.

To guarantee a smooth transition from the current relational archives widely used by the Public Administration to the envisaged semantic triple stores, the proposed architecture supports also queries expressed by distributed SQL formulas so that it is possible to retrieve data from either SQL tables or RDF archives.

This choice will allow us to query existing SQL-based centralized DWH, e.g., as adopted in the *project ELISA* promoted by the Italian Ministry for the Regional Affairs described in http://www.tributi.eng.it/blog/progetto-elisa/, or to query RDF triple stores extracted from SQL tables in case the data owners have defined the controlled vocabulary and related RDF formats to give semantics to their data.

These ideas have inspired the implementation architecture of the K-Metropolis project illustrated in fig.2, where we point out the central XML DB containing the data taken by a distributed real time monitoring system dealing with urban transport infrastructures, e.g., traffic and car pollution, or city utilities such as electricity, water, gas. Also, XML and SQL data stores belonging to organizations offering public services (e.g., parks and pharms) and Municipal DBs are taken into account to better support walking and driving citizens.

In particular, fig.2 points out that K-Metropolis is centered on a Ruby on Rails (RoR) server since this software environment allows us to implement the information system per use cases using the Model-View-Controller paradigm (Hartl, 2011).

The server is provided with a Decision Support System (DSS) illustrated in previous works, e.g., (Costanzo et al., 2013), to take advantage from all the urban databases. If the server may enter into the nodes of the monitoring system, then the DSS may provide suggestions useful to control specific parts of the urban infrastructures, e.g. (Aoun, 2013); if the server may access only the general information stored on the company servers devoted to manage the single infrastructure or utility, then the DSS will provide suggestions to improve daily maintenance, optimizing the overall costs and planning future expansions, e.g., (Al-Hader et al., 2009).

In both the above cases, the K-Metropolis choice of using XML protocols to exchange data with the distant computing systems facilitates the required DB interoperability, as clarified in sect.2.

The relevant points of interest and possibly the best paths to reach them are sent to the mobile users by means of JQMobile scripts (David, 2011) so that they may be displayed on a Google Maps based user interface of any mobile.

Moreover, a software based on the Flash Builder framework (Corlan, 2009) has been developed to implement the main DSS functionalities on the most powerful mobiles, e.g., android or iphone devices. This allows these mobiles to play an autonomous powerful role in the ubiquitous information system. Indeed, in this case, the mobiles may carry out e-government operations without the intervention of the server by means of their local DSS which takes into account the personal data stored on the mobiles, instead of the ones resident on the server, and the required business data resident on the remote data bases accessed directly through the GPRS network.

3 DATA INTEGRATION

As known, the Ontology Web Language (OWL) is a language for making ontological statements to be used over the World Wide Web, and all its elements (classes, properties and individuals) are defined as RDF resources (Smith et al., 2004).

Unlike traditional systems that operate with a scoped (or closed) vision of the reality that implies that a not known statement is automatically false, OWL makes an *open world assumption* considering the unknown information as not explicitly false.

One of the main attractive feature of OWL is that it allows us to integrate disparate data stores at the only condition of describing the original data by OWL statements belonging to a shared representation of the concepts underlying these data.

Of course, ontologies have an effective utility depending on if they are agreed by the cooperating organizations and communities. For this reason, the activity carried out by the Linked Data group at the W3C and the Government Linked Data (GLD) that are publishing data sets and knowledge bases for supporting e-business and e-government activities involving different organizations working at global scale (Bizer et al., 2009) are very important. Such ontologies may be refined by adding concepts able to integrate the databases of companies and the ones of the public departments that operate at local scale.

Although we have not yet well established global and local ontologies defining e-government and e-business activities, the mentioned semantic approach remains the main methodology for the integration of public data. For this reason, in K-Metropolis we chosen to integrate the distributed data bases by transforming the original SQL data into XML statements to be easily mapped on their turn into the standard RDF triples, once they will be available, without changing the codification of the SQL data stored at the lower layers.

For example, in K-Metropolis the XML description of the parks of a city that can be accessed by either the RoR server and the Flash Builder based applications is as follows:

```
<parks>
    <park>
        <organization> … </organization>
        <name>  … </name>
        <id>  … </id>
        <address> … </address>
        <closing day> … </closing day>
        <vacancy> … </vacancy>
        <lng> … <lng>
        <lat> … <lat>
    </park>
</parks>
```

The only condition to allow the mobiles to retrieve all the parks is that the RoR server is provided with a XML directory service that contains the web of the addresses of the various organizations that offer parking services, i.e.:

```
<park directory>
    <organization>
    <URL>  … </URL>
    <id>  … </id>
    </organization>
</park directory >
```

Analogous XML descriptions have been defined to represent the data of the other urban/metropolitan Public Services. In this way, we may send to the mobile users all the points of interest (POIs) and provide the needed services independently on the organization to which they belong to.

Fig.3 shows the main functionalities offered currently to the citizens to help their mobility (i.e., to find nearest convenient “parks” and “fuel stations”), to assist the search of essential services (i.e., “pharms” and in a near future “hospitals”, and “first aid centres”) and to support typical e-government operations using the icon “offices”.

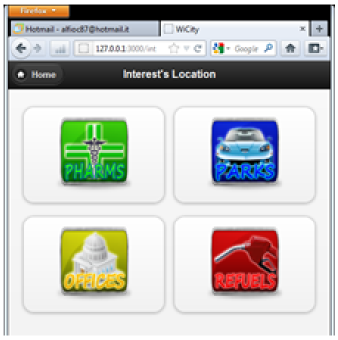

Figure 3: Some e-service and e-government functions offered to the mobile users.

Such functionalities make use of the data integration technology to offer effective services to the users, e.g., fig.4 shows how the list of the parks of different organizations is displayed on the user mobiles after the Flash Builder based application resident on the mobile has extracted the data relevant for the user query from different distant servers using the described XML approach without any server intervention. Similar information may be obtained for the fuel stations, the pharms and the other mentioned essential services.

Figure 4: Parking list meeting the user query. The parks belong to different organizations.

After having extracted the services required by the user, the DSS resident on the RoR server or on the mobile is able to suggest: a) the most suitable services, b) where they are located on Google Maps, as illustrated in fig.5, and c) the best path to reach them, as shown in (Costanzo et al., 2013).

Figure 5: Parking localization displayed on a Google Maps based representation.

Also, K-Metropolis allows the users to pay in advance all the mentioned services using PayPal (fig.6).

Figure 6: e-Payment using PayPal for parking the car in the park chosen by the user.

Let us note that not only e-commerce activities but also e-government requests coming from the citizen mobiles are supported by K-Metropolis, e.g., if we press the icon associated to the offices of the Public Administration in fig.3, the mobile user may ask certificates (fig.7.left) or may be supported in compiling an *affidavit* (fig.7.right). This has been obtained by transforming the municipal data from relational tables to XML representations.

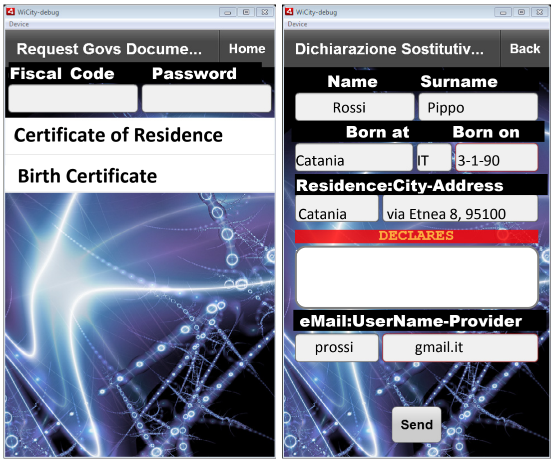

Figure 7: e-Certificates, on the left, and e-declaration, on the right, required from the mobile either directly or through the K-Metropolis server.

4 CONCLUSIONS

In the paper we have sketched the project named K-Metropolis aiming at integrating disparate urban data bases to support user mobility and to provide e-commerce and e-government services to desktop PCs and mobiles.

Although the proposed data integration has been carried out by filtering a list of XML records collected from distant sites using simple *select-where* like operations, this simple data integration method is powerful enough to cover many relevant use cases required by citizens and companies.

Since *join* operations between distributed data bases may be useful too, the mentioned companion paper, i.e., (Giordano et al., 2013), illustrates how K-Metropolis is able to accomplish this complex task, as well as it presents how the use of a powerful GIS may be used to display not only geo-referenced data of the points of interest but also maps that further qualify the land use for urban studies or to design suitable interventions to improve civil protection.

In particular, in this companion paper we show how using RDF data representations instead of the XML schemes may improve data integration and illustrate the SPARQL queries that are able to access distributed RDF triple stores to carry out both *select* and *join* operations.

Let us note that many approaches have been proposed to offer information services to mobile users. The dedicated navigators installed on the cars, e.g., Garmin and Tom-Tom, were the first examples of this technology. They may be easily used by the drivers, but provide only transport information that don't take into account very often the real time car traffic flows.

Although some protocols have been proposed to improve the real time functionalities of such navigators, e.g., VANET described in (Offor, 2012), the car navigators remain with a limited area of application. In particular, they cannot be used easily to support walking people mobility, neither can be used to carry out e-commerce and e-government tasks.

For this reason, different location based information systems were proposed in the last years. They are mainly resident on mobiles and appear to be a new generation of location based services (LBSs) that help better people mobility as well as facilitate e-commerce and e-government operations, as foreseen in (TRG, 2008).

However, all the proposed LBSs of this second generation are mainly proprietary systems, so they miss two basic requirements of the modern LBSs that are at the basis of the K-Metropolis project, i.e., the requirements that the urban data bases should be open and interoperable, as claimed in (Teller et al., 2010).

Therefore, our future work will be mainly the one to study carefully the available urban ontology, e.g., (Teller et al., 2007; Berdier and Roussey, 2007), to choose the ones that favour the implementation of an urban presentation layer based on an standard vocabulary that allows the above mentioned K-Metropolis applications, resident on either the server or the mobiles, to access all the public data available at citywide scale thus supporting the activity of the mobile users as completely and flexible as possible.

REFERENCES

Al-Hader, M., et al., SOA of smart city geospatial management. *Third UKSim European Symposium on Computer Modeling and Simulation*, EMS '09, 2009.

Aoun C., The Smart City Cornerstone: Urban Efficiency. Schneider Electric White Paper, 2013.

Berdier C., Roussey C., Urban Ontologies: the Towntology Prototype towards Case Studies. Springer, 2007.

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

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

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

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

Crisafi A., Giordano D., Spampinato C., 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, 2008.

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

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

Faro A., Giordano D., Spampinato C., 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, 2008.

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

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)*.

Giordano D., Torre A., Salemi C., Alessi S., Faro A., An Ontology based Approach to Integrate Data and Maps in the Government Enterprise Architecture: a Case Study. *Proc. 15th International Conference on Enterprise Information Systems*, ICEIS, Angers, INSTICC, 2013.

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

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

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

Offor, P.I., Vehicle Ad Hoc Network (VANET). 2012. Safety Benefits and Security Challenges. Available at SSRN: http://ssrn.com/abstract=2206077.

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

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

Salampasis, M., Tektonidis, D., Kalogianni, E., TraceALL: A Semantic Web Framework for Food Traceability Systems. *Jornal of Systems and Information Technology, Vol. 14(4)*, 2005.

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

Smith, M. K. , Welty, C.D., McGuinness D.L., OWL Web Ontology Language Guide, W3C Recommendation. http:// www. w3.org /TR/ 2004/REC-owl-guide-20040210, 2004.

Telematics Reserach Group (TRG), Preview portable navigation: the future is bright for connectivity. http://www. telematicsresearch.com/PDFs/TRG_Press_Jan_08.pdf,2008.

Teller J., John R. Lee J.R., Catherine Roussey C., (Eds.), Ontologies for Urban Development, Springer, 2007.

Teller J., Billen R., Cutting-Decelle A.F., 2010. Ontology based approaches for improving the interoperability between 3D urban models. Journal of Inf. Technology in Construction.

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