Km4City as Smart City Semantic Model and Tools (submitted for Open Track)

Claudio Badii, Pierfrancesco Bellini, Ivan Bruno, Mariano DiClaudio, Giacomo Martelli, Paolo Nesi, Nadia Rauch
DISIT Lab, Distributed [Systems and internet | Data Intelligence and] Technologies Lab, Dep. of Information Engineering (DINFO), University of Florence, Italy,
Fax: 0039-055-2758570, tel: 0039-3355668674
http://www.disit.dinfo.unifi.it

Abstract. *our cities are not so smart as they could be* Many times on traveling we realize the needs of getting more specific information that what we can get from Goole, OpenStreet Map, Here, TomTom, etc. In most cases, the information is too complex and expensive for them to be profitable. This is true since the information is in the hands of public administrations, mobility operators, private services, parking companies, etc. **Km4city** provides a unique point of access for interoperable data of a city metropolitan area via web and mobile applications. Km4City covers aspects of mobility and transport, energy, banks, parking, commercial, culture, bike paths, garden areas, health, tourism, end much more. Florence area in Italy has, since July 2015, a demonstrator of this solution.

1 Pitch

How many times you are on traveling realizing the needs of getting something specific, and thus, take the decision on stopping and searching locally since global operators tool do not provide the right information. The information is not accessible for them, and it is too complex for them to cope with so variable and complex information, it is not commercially viable for them. This is true since the information is of public administration in files, or the travel planner on the bus on private services, or from the parking company, etc. **Km4City is a solution.** Florence area in Italy has, since July 2015, a demonstrator of this solution.

adfa, p. 1, 2011.

A very large number of public and private data sets are available from local governments and are huge resources for Europe and for the cities. Local governs data are more detailed, some public other are private and accessible only as authenticated services. Open and private valuable data at level of city, region or at national level, are too fragmented and not accessible for the final users, citizens or companies that use to exploit them for providing services. They could be a source of revenues and are becoming a big business for local specific services, too small and complex to be of interest for worldwide operators. In most cases, the above mentioned local private and open data are not semantically interoperable and a huge effort would be needed to create applications and services that provide them in aggregated manner. They can be used for providing integrated services in multiple domains to citizens. **Km4city** provides a unique point of access for interoperable data of a city metropolitan area via web and mobile applications (http://www.disit.org/km4city). Km4City covers aspects of mobility and transport, energy, banks, parking, commercial, culture, bike paths, garden areas, health, tourism, end much more. Km4City provides a set of scalable and efficient tools for data ingestion, management, aggregation, indexing and for producing in short time web and mobile applications have been realized and make accessible. A smarter navigation can be performed from http://servemap.disit.org and can be used for smart decision support system on the city based on System Thinking (http://smartds.disit.org).

2 Description

Despite the huge work performed by Public Administrations (PAs) on producing and publishing open data they are not typically semantically interoperable each other and neither with the many private data available in the city. Open data coming from PAs contains information about the city on multiple domains (such as data on the population, risk, votes, admin, energy, cultural heritage, etc.), location of point of interests, POIs, on the territory (including: museums, tourism attractions, commercial, restaurants, shops, hotels, charging points, waste points, bike paths, etc.), major GOV services (tax office, tourist office, hospitals, hosting houses, etc.), ambient data, weather status and forecast, changes in traffic rules for maintenance interventions, etc. Most of these data are produced by different entities, in different moments and time,

and not with the same vocabulary/formats, standards, velocity, structure, etc. In addition, private data coming from mobility and transport such as those created by Intelligent Transportation Systems, ITS, for bus management, and solutions for managing and controlling parking areas, car and bike sharing, car flow in general, delivering services, accesses on Restricted Traffic Zone, RTZ, people flows, etc. Other aspects are related to risk assessment and resilience of the city with respect to disasters and/or accidents may occur: water flooding, terrorists' attacks, weather, energy, etc. Both Open and Private Data are under control of their corresponding providers and thus their licenses have to be taken into account. They may include real time data (traffic flow, people flow, position of private and public vehicles, delivering services, events, etc.), railway and train status with respect to the arrival, parking status, ambient sensors (pollution, earthquake), and TVcameras streams for security and flow, status of the emergency services in the hospitals, etc.

PAs and city operators (energy, mobility, etc.) have large difficulties in elaborating and aggregating data to provide new services, even if they could have a strong relevance in improving the citizens' quality of life and services. Therefore, *our cities are not so smart as they could be* by exploiting a semantically interoperable knowledge base on the available data [1]. This condition is also present in highly active cities on open data publications and on smart city services. When data models are analyzed and then processed to become semantically interoperable, they can be used to create an integrated knowledge base that can be feed by corresponding data instances (with static, quasi-static and real time data). Therefore, variability, complexity, variety, and size of these data make the data process of ingestion, aggregation, to enable their exploitation a "Big Data" problem as addressed in [2]. One of the issues is the scarce (or non-existing) semantics interoperability among the data sets. Therefore, common models and tools are mandatory to reach the needed quality level for providing valid services, as numerically proof in [1].

Presently, there are some ontologies for modeling smart city aspects as derived from the EC project READY4SmartCities FP7 CSA, http://smartcity.linkeddata.es/ . On the other hand, only a few of them are available, open, and cover a large range of domains. These features are only endowed by **Km4City** ontology (**Knowledge Model for City**) [1]. Km4City is a comprehensive and open ontology for smart cities covering domains of: weather, cultural heritage, smart sensors, public structures, mobility, city parking, services, transportation, events, geographic locations, health, etc. It has in addition an interesting roadmap

in which other domains are going to be added with some national and European Commission projects working on it. On the other hand, the availability of a well-defined ontology with a large coverage is not enough to cope with the Smart City modeling and data integration, since the data have to be ingested and made semantic interoperable each other and with the model to enable the set up and delivering of "Smart Services" to citizens. A full set of tools are used for data gathering and processing, decision support, query API, scalability and workflow control.

The proposed presentation and demonstration are going to expose the integrated solutions around Km4City model which has been set up by the DISIT lab in Florence (http://www.disit.org/km4city) and adopted in some EC and national smart city projects Sii-Mobility (Smart City MIUR project), RESOLUTE H2020 (http://www.resolute-eu.org), Km4City service and tools in place in the Florence Area, with many industrial and research partners. Km4City models a large set of data kind and provides support for inference and reasoning, on time and space, on public and private data, on static and real time data. It includes a set of tools:

Km4City final user tools on WEB Http://www.km4city.org
Km4City final user tools on mobile: Android and iOS (links are reported in Appendix), search on Google Play and App Store for Km4City
Demonstrative open source mobile application exploiting ServiceMap API, presented at the Florence Open Data Day and accessible as open source via: http://www.disit.org/6595 .
Service map: http://servicemap.disit.org is a tool for PA administrators and developers to provide access at several kinds of geospatial queries in Florence and in the Tuscany region. The ServiceMap is also a tool for developers, which can be used to understand the usage of API to access at the Km4City services http://www.disit.org/6597 . The ServiceMap allows the visual creation of queries on the city, and may send to the connected user via email the SPARQL code of the visual queries performed, and in addition also a simple Query ID.
Km4City ontology model and documentation [1], document http://www.disit.org/5606, http://www.disit.org/km4city/schema/
Parallel and distributed architecture based on ETL, scheduler, HBase, Hadoop, for massive big data ingestion (static, quasi static, and real time data), reconciliation, data enrichment (for connecting Km4City URI to dbPedia, geonames, etc. [3]) and for making decision: [1]. Examples are accessible about the ETL transformation for data ingestion,

quality improvement, conversion in triples, reconciliation in SILK, [1], etc.

- **solution on SLIDE: http://www.disit.org/6669**
- Article of 2014: P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", International Journal of Visual Language and Computing, Elsevier, http://dx.doi.org/10.1016/j.jvlc.2014.10.023,
- **RIM**: RDF Index Manager: user manual, for versioning of graph databases RDF stores. http://www.disit.org/6755 [6]
- **DIM**: Data Ingestion Manager: http://www.disit.org/6732

3 WebAccess

The applications of Km4City are accessible via the web. The instructions are not needed start and use the end-users application, while the instruction for the administrator and developers are on km4city web page http://www.disit.org/km4city .

Citizens and tourists as Km4City (Firenze what where) publicly accessible as mobile applications on:

- Google Play: https://play.google.com/store/apps/details?id=org.disit.siiMobile
- Apple Store: https://itunes.apple.com/app/florence-km4city/id1028356115?mt=8
- Web: http://www.km4city.org

Public Administrators, Developers, mobility operators:

- ServiceMap: http://servicemap.disit.org (public)
- Smart decision support: http://smartds.disit.org (user paolo.nesi@unifi.it, password: “prova”);
- UPDATED solution on SLIDE: http://www.disit.org/6669

4 References

[1] P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", International Journal of Visual Language and Computing, Elsevier, http://dx.doi.org/10.1016/j.jvlc.2014.10.023

[2] Bellini P., Di Claudio M., Nesi P., Rauch N., "Tassonomy and Review of Big Data Solutions Navigation", Big Data Computing, 2013 by Chapman and Hall/CRC

[3] P. Nesi, G. Pantaleo and M. Tenti, "Ge(o)Lo(cator): Geographic Information Extraction from Unstructured Text Data and Web Documents", SMAP 2014, 9th International Workshop on Semantic and Social Media Adaptation and Personalization, November 6-7, 2014, Corfu/Kerkyra, Greece. technically co-sponsored by the IEEE Computational Intelligence Society and technically supported by the IEEE Semantic Web Task Force.

[4] P. Bellini, P. Nesi and N. Rauch, "Smart City data via LOD/LOG Service", LOD2014, Workshop Linked Open Data: where are we?, organized by W3C Italy, Rome, 2014.
[5] P. Bellini, I. Bruno, A. Cavaliere, D. Cenni, M. DiClaudio, G. Martelli, S. Menabeni, P. Nesi, G. Pantaleo, N. Rauch, "Km4City: Smart City ontology and tools for city knowledge exploitation", European Data Forum 2015, Luxembourg, 2015.
[6] P. Bellini, I. Bruno, P. Nesi, N. Rauch, "Graph Databases Lifecycle Methodology and Tool to Support Index/Store Versioning", International Conference on Distributed Multimedia System, Hyatt Regency, Vancouver, Canada, August 31 - September 2, 2015. http://www.disit.org/6749

5 Appendixes for requirements

5.1 Satisfactory of Minimal Requirements

Requirement 1: *The application has to be an end-user application, i.e. an application that provides a practical value to general Web users or, if this is not the case, at least to domain experts. It should show-case functionalities that the use of semantic web technologies can bring to an application.*

Km4City is substantially an end-users application for:

- **Citizens and tourists** as Km4City (Firenze what where) accessible as mobile applications on: Google Play, Apple Store, Web: http://www.km4city.org (see link below)
- **Public Administrators, Developers, mobility operators**:
  - ServiceMap: http://servicemap.disit.org
  - Smart decision support: http://smartds.disit.org
  - UPDATED solution on SLIDE: http://www.disit.org/6669

Requirement 2: *The information sources used: should be under diverse ownership or control; should be heterogeneous (syntactically, structurally, and semantically); and should contain substantial quantities of real world data (i.e. not toy examples).*

Km4City provides aggregated data via RDF store end point. **The data provided by Km4City are under diverse ownership and control and are produced in real time and statically by a number of partners:**

- **Municipality of Florence** provides static data (more than 200 data sets), real time data regarding WiFi sensors. Data sets are related to: services, locations, cultural heritage, tourism, mobility, energy, administrative, schools, hospitals and services, etc.
- **Province of Florence** provides data from MICC traffic manager system: (i) static data such as the positions of bus stops and traffic sensors, parking area, etc.; (ii) real time data such as: bus position and delay, traffic flow data sensors, parking status, etc.
- **Tuscany Region** provides static geodata of streets, numbers, services, zones, etc.
- **LAMMA** provides real time data as weather forecast, wind, etc.

- University of Florence provides aggregated data and enrichment data taken from dbPedia.

The data of Km4City are heterogeneous (syntactically, structurally, and semantically):

- **Syntactically**: several datasets are aggregated and reconcilied. See http://www.disit.org/6726 Data sets arrive as XML, CSV, KMZ, PDF, RDF triples, JSON, via files, HTML, web services, REST call, etc.
- **Structurally**: The information is **heterogeneous**: geographical, energy, services, mobility, weather forecast, events information and locations, wifi hot spots positions and measured powers with the Km4City Android Application in the hands of the final users.
- **Semantically**: The meaning of data sets is **heterogeneous** presenting structure and meanings that have been integrated into a common semantics by using the Km4City backoffice, including: services, geographic descriptions, mobility position and movements, weather forecast, events, sensors, wifi, parking, etc. The back office of Km4City is described in:
    - **UPDATED solution on SLIDE: http://www.disit.org/6669**
    - Article of 2014: P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", International Journal of Visual Language and Computing, Elsevier, http://dx.doi.org/10.1016/j.jvlc.2014.10.023,
    - RIM: RDF Index Manager: user manual, for versioning of graph databases RDF stores. http://www.disit.org/6755
    - DIM: Data Ingestion Manager: http://www.disit.org/6732

Requirement 3: *The meaning of data has to play a central role. (i) Meaning must be represented using Semantic Web technologies; (ii) Data must be manipulated/processed in interesting ways to derive useful information; and (iii) this semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all.*

In Km4City, the meaning of data is central since the inference produced into the RDF store is at the basis of the services provided via the WEB and Mobile applications.

- **(i) Semantic web technologies:** solution is fully based on Ontologies, RDF store and triples; queries are performed by the applications (Km4City, ServiceMap, smartDS, LOG, ..) to the RDF Store in SPARQL;
- (ii) processed in interesting ways: Triples are inferred at the time of indexing on Virtuoso RDF Store and at the SPARQL time; The enrichment of strings of the Km4City to find the LD/LOD in dbPedia is performed by using SPARQL queries and similarity; RDF store is endowed of reasoners for strings distance/index, georeasoning, time reasoning, graphic reasoning (inclusions and distances).

- (iii) the reasoning is central since the Applications of Km4City give this facility as a first feature to find information starting from geoposition, time instant, text, etc.

The obtained performance and reasoning capabilities, together, are not replicable on traditional databases.

5.2 Satisfactory of the Additional Features

- Km4City application provides an attractive and functional Web interface (for human users) see: http://www.km4city.org and for mobile applications see Google Play, Apple Store, etc.
- Km4City application is scalable (in terms of the amount of data used and in terms of distributed components working together). Km4City aggregate all data currently published on the Semantic Web and the process of integration and aggregation is parallel and scalable see http://www.disit.org/6669 for the architecture.
- A rigorous evaluation has been performed to assess the benefits of Km4City solution in terms of SPARQL performance in query and indexing. See http://www.disit.org/6750 and http://www.disit.org/6724
- Km4City is the unique solution that exploited in a so extensive manner the semantic technology for Smart City Services. For this reason has been invited to EDF2015 [5], Smart City 360, SMAU 2015, etc.
- Km4City functionality goes beyond pure information retrieval since the RDF stores of Km4City is exploited for Smart Decision Support System (http://smartds.disit.org), and reasoning on geospace.
- Contextual information is used by Km4City for ratings or rankings the services provided to the users. A personal assistant is going to the deployed in the next versions.
- Multimedia documents are used in Km4City, since some Digital Locations in Florence provide Images, Audio or Documents.
- Km4City integrates dynamic real-time data (e.g. workflows) in the front-end store, with static and quasi static data in the back office, see architecture (http://www.disit.org/6669). For this see DIM (data Ingestion Manager) http://www.disit.org/6732 and RIM (RDF Store Indexing Manager) http://www.disit.org/6755 . The RIM/DIM also supports versioning of RDF store.
- Km4City in providing the results is accurate as much possible exploiting ranking for similarities and distances according to the geo-context.
- Km4City supports English and Italian on WEB, Android App, Apple iOS App, for mobile, web, tablets, etc., and can be easily extended to other languages.