An Integrated Smart City Platform

Paolo Nesi[1], Laura Po[2], José R.R. Viqueira[3], and Raquel Trillo Lado[4]

[1] Information Engineering Department
University of Florence,
Via di Santa Marta, 3, 50139 Firenze, Italy
paolo.nesi@unifi.it
[2] "Enzo Ferrari" Engineering Department
University of Modena and Reggio Emilia
Via Vivarelli, 10 41125 Modena, Italy
laura.po@unimore.it
[3] Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS)
Universidade de Santiago de Compostela
Santiago de Compostela, Spain
jrr.viqueira@usc.es
[4] Depto. de Informática e Ingeniería de Sistemas (DIIS) e I3A
Universidad de Zaragoza
Zaragoza, Spain
raqueltl@unizar.es

Abstract. *Smart Cities* aim to create a higher quality of life for their citizens, improve business services and promote tourism experience. Fostering smart city innovation at local and regional level requires a set of mature technologies to discover, integrate and harmonize multiple data sources and the exposure of effective applications for end-users (citizens, administrators, tourists...). In this context, Semantic Web technologies and Linked Open Data principles provide a means for sharing knowledge about cities as physical, economical, social, and technical systems, enabling the development of smart city services. Despite the tremendous effort these communities have done so far, there exists a lack of comprehensive and effective platforms that handle the entire process of identification, ingestion, consumption and publication of data for Smart Cities.

In this paper, a complete open-source platform to boost the integration, semantic enrichment, publication and exploitation of public data to foster smart cities in local and national administrations is proposed. Starting from mature software solutions, we propose a platform to facilitate the harmonization of datasets (open and private, static and dynamic on real time) of the same domain generated by different authorities. The platform provides a unified dataset oriented to smart cities that can be exploited to offer services to the citizens in a uniform way, to easily release open data, and to monitor services status of the city in real time by means of a suite of web applications.

Keywords: Data Integration, Open Data, Linked Data, smart city platform, Smart Cities, ontology

1 Introduction

"A Smart City is an Information and Communication Technologies (ICT) enabled development which extensively uses information as a way to improve quality of life for its citizens and population at large" [1]. In this context, sensors, social media, web activities, tracking devices, etc. generate various and large amount of real-time data that Smart Cities need to deal with.

This paper proposes a platform for the European national administrations, citizens and companies for implementing data integration processes that short the time and costs in enabling the data gathering for smart city applications and services. Starting from local services available to citizens such as public transport information, commercial data, etc., it will build a unified dataset by using consolidated technologies and tools. This dataset that can be used and queried by several applications for exploring the data, searching for information, extracting statistical indicators or publishing open data. The proposed platform is based on four open-source mature tools that covers the entire process of the data value chain (see Figure 1). These tools have been used for several years in specific areas such as data integration (MOMIS [2,3]), data aggregation and reconciliation towards a Smart City Ontology (ETL tools and KM4City [4]), integration of environmental sensor data (SOS-SM [5–7]) with respect to the Semantic Sensor Network W3C ontology (SSN ontology) and population and update of ontologies (Infoboxer [8]).

This paper is devoted to describe how the proposed platform works, and how it can be used. The rest of the paper is structured as follows. In the next Section, related work is analyzed. After that, in Section 3, the proposed data value chain is depicted. Moreover, some of the web applications that interact with the unified RDF dataset are briefly described. Finally, in Section 4 conclusions and future work are depicted.

2 Related Work

The amount of data generated and collected by public administrations (city councils, regional administrations, etc.) has been exponentially increasing since the beginning of the digital era. Besides, since all nations of the Organization for Economic Co-operation and Development (OECD) signed a declaration establishing that all publicly-funded data should be made publicly available[5], Open Data Initiatives have emerged to ease the development of methodologies, technologies and standards in order to publish these data. One of the most successful Open Data Initiatives has been the Linked Open Data Initiative[6], which is focused on improving the access and integration of public data coming from different sources by using machine readable datasets.

In this context, numerous projects and different initiatives to exploit all these data in order to improve the management of cities and the quality of life of their

---

[5] http://goo.gl/XYhfEb
[6] http://linkeddata.org/

citizens have arisen: CitySDK[7], [9], [10], etc. These projects are focused on specific domains (e.g. energy, pollution, transport, tourism) or challenges (e.g. how to make predictions on a specific topic, integrating specific datasets, publishing data, ...). However, up to our knowledge, there is not a general platform that takes into account all the data value chain to offer smart city services by considering all the affected agents (administrations, citizens, companies, etc.). So, in this paper, we propose such a platform by integrating previous consolidated solutions focused on specific problems (data integration, ontology population, etc.).
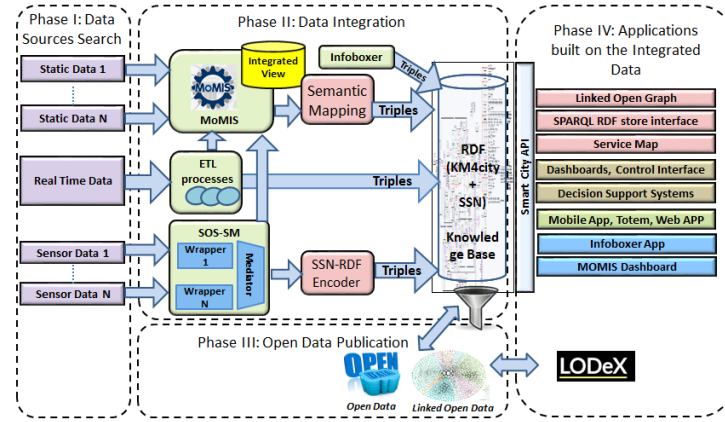
Fig. 1. The Integrated Smart City Platform and the 4 phases of the data value chain.

3   The Data Value Chain

The steps that guides users from the discovery of raw data and its ingestion to the analysis and exploitation of results realize the *data value chain*, as "in a Data Value Chain, information flow is described as a series of steps needed to generate value and useful insights from data"[11]. The proposed platform covers the entire process of the data value chain that is composed of four phases (see Figure 1).

Phase I (Data Sources Search) starts with the definition of a specific goal. Then, a deep and wide search of the relevant data sources (local, regional, national and international sources) is performed. Both public and private data sources are analyzed and selected w.r.t. the domain of the smart city project [12, 13]. The platform allows an iterative process, so new data sources can be incrementally considered to enrich the project.

---

[7] http://www.citysdk

Phase II (Data Integration) is devoted to the integration and mapping of the different selected data sources w.r.t. the Knowledge Model for City (KM4City) and the Semantic Sensor Network Ontology (SSN)[8]. Both KM4City and SSN are domain-specific ontologies; whereas KM4City focuses on providing a vocabulary to improve the effectiveness of smart city applications, SSN is an ontology developed by the W3C Semantic Sensor Networks Incubator Group (SSN-XG) to describe sensors and their observations. Integration and mapping are performed with different tools by considering the type of the selected data-sources. Thus, sensor data are processed by Sensor Observation Service Semantic Mediation (SOS-SM); real time data are filtered and transformed by means of Extraction Transformation and Load (ETL) tools (such as Pentaho[9]); and static heterogeneous data-sources are integrated by means of MOMIS[10], which can aggregate data coming from both structured or semi-structured data sources in a semi-automatic way to bring out new information from apparently unrelated existing data. The created integrated view is transformed into RDF triples. Moreover, this obtained RDF Knowledge Base can be extended or updated by using Infoboxer[11]. A tool oriented to non-technical users that helps to link the values introduced to existing entities in the data source and enforces semantic constraints on them.

Ingesting data from different kind of public and private sources necessarily requires to deal with aspects such as: variability, complexity, variety, geo-spatial aspects, integration and size of these data sources. So, data ingestion and aggregation processes must address the "Big Data" issues described in [4, 14, 15, 2]. This problem can be partially solved by using specific reconciliation processes to make these data interoperable with other ingested and harvested data. The velocity of data is related to the frequency of data update, and it allows distinguishing static from dynamic data.

Phase III (Open Data Publication) makes the resulting value-added information public and searchable on the Web as Linked Open Data. The owner of the data can choose or filter a portion of the RDF Knowledge Base and publish it as Open Data or Linked Open Data. In particular, the goal of this phase is to enable users to publish one or a set of datasets according to the 5-stars deployment scheme for the Linked Open Data proposed by Berners-Lee [16]. Thus, linking the dataset to external sources is unavoidable in order to create a proper 5-stars dataset.

Phase IV (Applications built on the Integrated Data) takes as input the integrated information and provides specialized applications such as tools for geographical querying, for exploring and navigating LOD sources, for analyzing statistical information, etc. In more detail, these applications exploit the RDF-Knowledge Base by making queries [17–20] and offer different services to different types of users (citizens, companies, tourist, staff of public administrations, mo-

---

[8] http://w3c.github.io/sdw/ssn/
[9] http://community.pentaho.com/projects/data-integration/
[10] http://www.datariver.it/data-integration/momis/
[11] http://sid.cps.unizar.es/Infoboxer

bile operators, etc.). For example: searching services around a certain GPS point such as looking for an area where restaurants are available, services to detect and predict critical conditions or discover cause-effect relationships, recommendation services tuned on the basis of statistical data, decision support services, dash-boards that allow to analyze Key Performance Indicators such as MOMIS-dashboard, etc. All the services can communicate with the system asking for data or providing data to the servers through the Smart City API [4].

4 Conclusions and Future Work

In this paper, we presented a platform to turn the potential of data for the economy and society into reality where public and private data sources can be exploited in order to improve the management of cities and the quality of life of their citizens. The proposed platform is a suite of four open-source mature solutions focused on: the integration of data sources (MOMIS), the publication of smart data related to cities by considering a unified view (KM4city), and the population and update of data sources by considering data provided by users (Infoboxer) and by sensor observation services (SOS-SM). Moreover, other mature tools, such as MOMIS dashboard, are also considered to enrich the platform.

The different components of this platform have been already tested and deployed in several contexts, successfully, but the integrated solution has not been tested yet. So, as future work, we would like to study how to improve the performance of the integrated solution and how to audit its deployments by considering standard metrics, such as the ones defined in the norm UNE 178301 "Smart Cities and Open Data".

References

1. Mamkaitis, A., Bezbradica, M., Helfert, M.: Urban enterprise: A review of smart city frameworks from an enterprise architecture perspective. In: IEEE International Smart Cities Conference (ISC2), Trento, Italy, September 12-15, IEEE (2016) 1–5
2. Bergamaschi, S., Beneventano, D., Corni, A., Kazazi, E., Orsini, M., Po, L., Sorrentino, S.: The open source release of the MOMIS data integration system. In: Proceedings of the Nineteenth Italian Symposium on Advanced Database Systems (SEBD), Maratea, Italy, June 26-29. (2011) 175–186
3. Bergamaschi, S., Po, L., Sorrentino, S.: Automatic annotation in data integration systems. In: On the Move to Meaningful Internet Systems 2007 (OTM) 2007, Vilamoura, Portugal, November 25-30, 2007, Proceedings, Part I. (2007) 27–28
4. Bellini, P., Benigni, M., Billero, R., Nesi, P., Rauch, N.: Km4city ontology building vs data harvesting and cleaning for smart-city services. J. Vis. Lang. Comput. **25**(6) (2014) 827–839
5. Regueiro, M.A., Viqueira, J.R.R., Taboada, J.A., Cotos, J.M.: Virtual integration of sensor observation data. Computers & Geosciences **81** (2015) 12–19
6. Regueiro, M.A., Viqueira, J.R.R., Stasch, C., Taboada, J.A.: Semantic mediation of observation datasets through sensor observation services. Future Generation Computer Systems **67** (2017) 47–56

7. Regueiro, M.A., Viqueira, J.R.R., Stasch, C., Taboada, J.A.: Sensor observation service semantic mediation: Generic wrappers for in-situ and remote devices. In: Proceedings of 35th International Conference on Conceptual Modeling (ER), Gifu, Japan, November 14-17. (2016) 269–276

8. Yus, R., Mulwad, V., Finin, T., Mena, E.: Infoboxer: Using statistical and semantic knowledge to help create wikipedia infoboxes. In: 13th International Semantic Web Conference (ISWC), Riva del Garda (Italy). Volume 1272., CEUR-WS (October 2014) 405–408

9. Khan, Z., Anjum, A., Soomro, K., Tahir, M.A.: Towards cloud based big data analytics for smart future cities. Journal of Cloud Computing **4**(1) (2015) 2

10. Albino, V., Berardi, U., Dangelico, R.M.: Smart cities: Definitions, dimensions, performance, and initiatives. Journal of Urban Technology **22**(1) (2015) 3–21

11. Curry, E.: The big data value chain: Definitions, concepts, and theoretical approaches. In Cavanillas, J.M., Curry, E., Wahlster, W., eds.: New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe. Springer (2016) 29–37

12. Beneventano, D., Bergamaschi, S., Gagliardelli, L., Po, L.: Open data for improving youth policies. In: Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD), Volume 2, Lisbon, Portugal, November 12-14. (2015) 118–129

13. Beneventano, D., Bergamaschi, S., Gagliardelli, L., Po, L.: Driving innovation in youth policies with open data. In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management IC3K, Lisbon, Portugal, November 12-14, Revised Selected Papers. (2015) 324–344

14. Nesi, P., Bellini, P., Bruno, I.: Graph databases lifecycle methodology and tool to support index/store versioning. In Guercio, A., ed.: The 21st International Conference on Distributed Multimedia Systems, Vancouver, Canada, August 31 - September 2., Knowledge Systems Institute (2015) 221–230

15. Nesi, P., Pantaleo, G., Tenti, M.: Ge(o)lo(cator): Geographic information extraction from unstructured text data and web documents. In: 9th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP, Corfu, Greece, November 6-7, IEEE (2014) 60–65

16. Berners-Lee, T.: Linked data. http://www. w3.org/DesignIssues/LinkedData.html (2006) Accessed: 2017-04-30.

17. Benedetti, F., Bergamaschi, S., Po, L.: Lodex: A tool for visual querying linked open data. In: Proceedings of the ISWC Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC), Bethlehem, PA, USA, October 11. (2015)

18. Benedetti, F., Bergamaschi, S., Po, L.: Online index extraction from linked open data sources. In: Proceedings of the 2nd International Workshop on Linked Data for Information Extraction (LD4IE 2014) co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 20, 2014. (2014) 9–20

19. Benedetti, F., Bergamaschi, S., Po, L.: Visual querying LOD sources with lodex. In: Proceedings of the 8th International Conference on Knowledge Capture, K-CAP 2015, Palisades, NY, USA, October 7-10, 2015. (2015) 12:1–12:8

20. Bellini, P., Nesi, P., Venturi, A.: Linked open graph: Browsing multiple SPARQL entry points to build your own LOD views. J. Vis. Lang. Comput. **25**(6) (2014) 703–716