

A Linked Data Lifecycle for Smart Cities in Spain

Almudena González, Boris Villazón-Terrazas, and José Manuel Gómez
{*agonzalez, bvillazon, jmgomez*}@isoco.com

iSOCO, Avda. del Partenon 10, Campo de las Naciones, Madrid, Spain

Abstract. Smart Cities combine diverse technologies to reduce their environmental impact and offer citizens a higher quality of life. In this paper we present an ongoing effort, within the context of Ciudad2020 project, for overcoming the challenge of homogenizing the citizen's access to services offered by heterogeneous, and independent entities within a Smart City scenario. We describe how we are applying the Linked Data Lifecycle, from specification to exploitation, within the vertical domains defined in such Spanish project.

1 Introduction

The increasing urbanization of the world, along with global problems of climate change, water scarcity, environmental degradation, economic restructuring and social exclusion require further consideration of the cities of the future. Moreover, we are seeing the rapid rise in the connection and usage of billions of low-end and affordable smart devices to the Internet.

Lately, the concept of Smart Cities has attracted considerable attention. However, there is a lack of formal models and consensual definitions. Cities are defined *smart* when their investments in the human and social capital, as well as in the communication infrastructures are aimed at fuelling a sustainable economic growth and a high quality of life [3].

Smart Cities combine diverse technologies to reduce the environmental impact and improve citizens lives. This is not, however, simply a technical challenge; organisational change in governments - and indeed society at large - is also necessary. Making a city smart is therefore a very multidisciplinary challenge, bringing together city officials, innovative suppliers, national and international policymakers, academics and civil society. Big industrial players as well as governments are focusing their research around Smart Cities, some examples are the European Union¹ and IBM². One particular challenge of the Smart Cities is to find the means to manage all the big data coming from the cities.

In this paper we present an ongoing effort, within the context of a Spanish Project, Ciudad2020³, for overcoming the challenge of homogenizing the citizen access to services offered by heterogeneous, and independent entities within a Smart City scenario. Next, Section 2 describes the related efforts for introducing Linked Data within Smart Cities. Then, Section 3 introduces the Linked Data Life Cycle within the vertical domains defined in the project. Finally, Section 4 presents some conclusions and future work.

2 Related Work

The number of Open Data portals is increasing, because of the demand of transparency and easy access to the data. In this context, there are also plenty of works related to

¹ <http://eu-smartcities.eu>

² <http://www.ibm.com/uk/smarterplanet>

³ <http://www.innprontaciudad2020.es/>

Smart Cities, though not all of them are related to Linked Data. In this section we present some of the approaches related to Smart Cities that follow the Linked Data paradigm.

Zaragoza Public Data Catalogue⁴ is an Open Data Portal that shows a Smart City as a city that allows mobility, knowledge and open access to the data. For this purpose, it includes twenty different datasets and mobile applications. It also provides a catalogue and a SPARQL End Point to the user.

Opendata Cáceres⁵ is an Open Data Portal that offers Linked Open Data datasets, allowing to the citizens and the companies access to the municipal data, facilitating reuse for developing applications.

United Kingdom Catalogue⁶ promotes Innovation as the key of and Smart City. This portal works with UK Public Sector information and data, encouraging the use and reuse of government datasets. It includes a directory of data available, applications and a SPARQL EndPoint.

Liviu-Gabriel Cretu [4] defines a Smart City as an event-oriented architecture, where digital devices allow interoperability between Internet of Services, Internet of Things and Internet of People. They explore the usability of the latest advances in SOA or Services-oriented Architecture and Semantic Technologies.

Lopez et al. [8] describe a Smart City as a complex system with heterogeneous data and present a Linked Data Platform for cataloguing, indexing and querying all the information.

Tallevi-Diotallevi et al. [9] aim to capture the pulse of the city of Dublin monitoring and decision-making with three aspects: Extending the SPARQL Language, processing heterogeneous data (streams and static) in real time and using a hybrid RDFS reasoner.

Balduini et al. [1] present a Streaming Linked Data framework to collect data streams, analyse and visualize the results using London Olympic Games and Milano Design Week as use-cases. This proposal is related to event analysis in the city, uses RDF for modelling and integrating data, SPARQL and sentiment analysis techniques for processing and analyse social data.

Table 1 summarizes the classification by domain, data and target audience of the Linked Data initiatives included in this survey.

The lack of Linked Data initiatives following a Multi-Domain, Multi-user and Multi-nature Data approach along with the needs of the Spanish Citizens and Public Administrations is what encouraged us to apply Linked Data Lifecycle within Ciudad2020 project along their vertical domains.

3 Generation and Exploitation of Linked Data within *Ciudad2020 Project*

Ciudad2020 project⁷ is focussed on the three fundamental axes of a Smart City, which are Energy, Transport, Environment, and City. Within this project, the Linked Data Portal⁸ was created for integrating the data coming from Smart Cities in the four axes, as shown in Fig. 1. This portal has several datasets whose contents include bike sharing systems, restaurants, museums, energy performance certificates and city tweets. For developers and for Public Administrations, it also provides a SPARQL endpoint and the possibility

⁴ <http://www.zaragoza.es/ciudad/risp/>

⁵ <http://opendata.caceres.es>

⁶ <http://statistics.data.gov.uk/flint-sparql>

⁷ <http://www.innprontaciudad2020.es/>

⁸ <http://ciudad2020.linkeddata.es/>

Reference	Domain	Used Data	Target Users
Zaragoza Public Data Catalogue ⁴	Mobility, Knowledge and Open Access	Open Data	Citizens and Developers
Opendata Cáceres ⁵	Culture, Transport, Environment, Society, Healthy, Energy	Linked Open Data	Citizens, Companies and Developers
United Kingdom Catalogue ⁶	Environment, Government, Mapping, Society, Health, Education, Bussiness and Justice	Open Data	Citizens, Developers, and Administration
Cretu, L.[4]	Event-driven Architecture Smart Cities	Semantic Web/Linked Data	Citizens
Lopez et al. [8]	Urban Monitoring	Static and Streaming Data	Citizens and Administration
Tallevi-Diotallevi et al. [9]	Transport, Environment and Energy	Streams and Static	Citizens
Balduini et al. [1]	City-scale Events	Streaming Linked Data, Social Data	Administration

Table 1. Linked Data initiatives related to Smart Cities

of querying streaming data, showing results in real time. In the following sections we are going to describe how we apply the Linked Data Life Cycle [7]; which consists of the following activities (1) specification, (2) modelling, (3) generation, (4) publication, and (5) exploitation; for each one of the Ciudad2020 vertical domains.

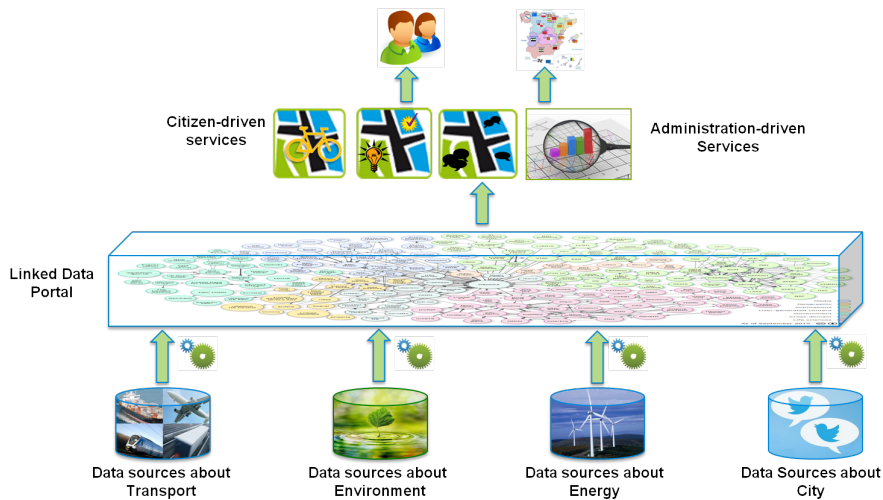


Fig. 1. High level overview of the Ciudad2020 Portal architecture.

3.1 Mobility and Transport

Specification. The Transport Portal⁹ combine Static Data Sources and Streaming Data Sources, using Linked Data as the homogenizer element for the process of combining this heterogeneous data:

- Static Data Sources. We use data of museums and libraries of the city of Leon and restaurants of Saragossa. This data come from Open Data Portals and Travel Guides, both available since February 2013:

⁹ <http://transporte.linkeddata.es/>

- *Junta de castilla y León Catalogue*¹⁰ is the Open Portal of the city of León.
 - *Zaragoza Public Data Catalogue*¹¹ is the Open Portal of the city of Saragossa.
 - *El viajero* [6] is a travel guide resulting from PRISA Group Data¹².
- Dynamic Data Sources. These data compress available bikes and slots in the different bike stations.
- *Citybike API*¹³ is the API for Bike Sharing Systems.

Data Conversion Services. We make different types of transformation depending on the nature of the data. In the case of Static Data Sources, we transform them by ETL processes (Extract, Transform and Load), generating data in RDF format.

On the other hand, regarding the Streaming Data Sources, we do not use an ETL transformation, since it would result in a hard and slow process. This Data Sources are shown as virtual RDF sources via streaming, by using the *morph-streams* technology, connecting the API REST of streaming services, web services and database producers based on complex events (CEP), establishing R2RML mappings [5].

Exploitation Use Case - Zaragoza Bizi. The *Zaragoza Bizi* Use Case¹⁴ combine static data and real time in the context of two cities: Saragossa and Leon. For instance, in Saragossa city there are 1300 bikes and more than 100 Km of cycle paths, serving citizens and tourists. This sums a total of 4,5 millions of uses and 13 millions of kilometres travelled in the last two years. At the same time, it avoided 2000 Tons of *CO*₂ emissions.

Within this use case, citizens and tourists can check the number of available bikes as well as the number of free slots. They can also check in the map the location of the stations (the exact latitude and longitude are provided), nearby points of interest (including restaurants and museums) by choosing a determinate distance, routes between different stations, sharing the resource via Twitter, and send a suggestion for updating a resource. Finally, for developers, they can also access the traditional RDF information associated to a resource or consult de SPARQL endpoint¹⁵.

3.2 Environmental Control

Specification We collect weather *data streams* from the weather API of OpenWeathermap¹⁶. It provides data from more than 40,000 weather stations. All weather data are obtained in JSON format.

Data Conversion Services In this case, the transformation process of the Streaming Data is the same we previously described in the Mobility and Transport section, we use the *morph-streams* technology [2].

Exploitation Use Case - Weathermap Meteo The *Weathermap Meteo* Use Case¹⁷ shows the current weather and it is available for 200,000 cities.

This Use-Case provides a *streaming querying* service. The users can get the current weather data for any location on the Earth in real time by consulting the SPARQLStream endpoint.

¹⁰ <http://www.datosabiertos.jcyl.es/>

¹¹ <http://www.zaragoza.es/ciudad/risp/>

¹² <http://www.prisa.com/es/>

¹³ <http://api.citybik.es/>

¹⁴ <http://transporte.linkeddata.es/browser.html>

¹⁵ <http://transporte.linkeddata.es/sparql.html>

¹⁶ <http://openweathermap.org/API>

¹⁷ <http://streams.linkeddata.es/register/weathermap>

3.3 Energy and Efficiency

Specification. The Energy Portal¹⁸ combines Static Data Sources from different cities using Linked Data:

- La Rioja city Energy Efficiency Certificates¹⁹: They provide an API to consult de database of certificates of buildings and projects located in the city, providing the rating, the consum, the emission and the address of the buildings.
- Navarra city Energy Efficiency Certificates²⁰: They provide a web search service to obtain the Energy information associated to new projects, new buildings and existing buildings. Filters by ranking and type of building are available.

Data Conversion Services. In this particular case, we transform the Static Sources by ETL processes (Extract, Transform and Load), generating data in RDF format.

Exploitation Use Case - *Navarra Energy Efficiency Certificates*

In this use case²¹ we show the statistics of the different Spanish provinces, which publish these certificates in Open Data portals.

We display a comparison of the provinces of La Rioja and Navarra, thereby the citizens can consult the energy performance certificates and check the number of available certificates with each rank, from *A* to *G*, where *A* means the most efficient buildings, and *G* means the least efficient buildings.

For instance, if the citizen choose the province of Navarra in the “Province” pull-down menu, the statistics of this city are shown. We can check that the most repeated rankings are *D* and *E*, with 1075 and 2368 buildings respectively.

3.4 City Data

Specification. The City Portal²² combine Static Data Sources and Streaming Data Sources, disambiguating and publishing those sources under the Linked Data paradigm.

Streaming Data Sources. Twitter API²³. We collect the tweets geolocated in the city of Saragossa and published during the first trimester of 2014.

Data Conversion Services We extract the named entities from Saragossa tweets. Once we get these named entities, we disambiguate them using the NERD tool ²⁴. We associate a Dbpedia²⁵ resource to each entity. We use the domain and semantic data ontology for disambiguating purposes. Finally, we generate RDF, we publish it and for the exploitation we use Flot Charts ²⁶ to display the graphics associated to the statistics. Regarding the publication activity, note that we cannot publish the tweet text, since the API Terms of Twitter specify that we may only return tweet IDs and user IDs.

¹⁸ <http://energia.linkeddata.es/>

¹⁹ <http://www.larioja.org/npRioja/default/defaultpage.jsp?idtab=772883>

²⁰ <https://administracionelectronica.navarra.es/webCertificacionesEnergeticas/BuscarCertificado.aspx>

²¹ <http://energia.linkeddata.es/browser.html>

²² <http://ciudad.linkeddata.es/>

²³ <http://www.twitter.com/>

²⁴ <http://nerd.eurecom.fr/analysis>

²⁵ <http://www.dbpedia.org/>

²⁶ <http://www.flotcharts.org/>

Exploitation Use Case - Zaragoza Tweets This use case²⁷ aims to show to the citizens a graphical visualization of statistics for the *named entities* in tweets geolocated in the city of Saragossa, during the first quarter of 2014, disambiguated and published in Linked Data. In this period, we observe that some of the most repeated entities named in the city are *Bands, Zaragoza city, Spain, Physical exercise*, etc. Each of these entities has an associated DBpedia resource. DBpedia defines a unique global identifier, including natural languages definitions and relations to another resources. For instance, the *Zaragoza* entity has the associated resource <http://dbpedia.org/resource/Zaragoza>.

4 Conclusion and Lessons Learned

In this paper we have presented (1) a small survey on research efforts related to applying Linked Data to Smart Cities, and (2) an ongoing work of applying Linked Data Lifecycle, i.e., generating, publishing and consuming Linked Data, within the vertical domains of the Ciudad2020 Spanish project.

The implementation of the use cases using RDF and Linked Data Principles has many benefits: (1) data interoperability (URI-based data integration), (2) flexibility (not fixed schema and no need to adapt SPARQL queries to a new schema, facilitating the incorporation of new datasets), (3) web-compatibility and (4) web-scalability (RDF unique identifiers). In contrast, the costs: dependence on the availability of data sources and license of data.

Regarding the future work we plan to (1) implement a Linked Data Platform within the project, following the W3C LDP Working Group²⁸ recommendations, (2) include and integrate new application domains, and (3) develop on top of the LDP a set of added value services, such as, recommender systems and analytics.

Acknowledgments This work is partially supported by the Ciudad2020 INNPRONTA project (IPT-20111006). We would like to thank Ontology Engineering Group-UPM.

References

1. Balduini, M., Della Valle, E., Dell'Aglio, D., Palpanas, T., Tsytsarau, M., Confalonieri, C., Social listening of City Scale Events using the Streaming Linked Data Framework, The Semantic Web-ISWC 2013. LNCS Volume 8219, 2013, pp 1-16.
2. J.P. Calbimonte, O. Corcho, A.J. Gray, Enabling Ontology-based Access to Streaming Data Sources, ISWC 2010.
3. A. Caragliu, C. del Bo, and P. Nijkamp. Smart Cities in Europe, 2009.
4. Cretu, L., Smart Cities Design using Event-driven Paradigm and Semantic Web, Informatica Economica, 16(4), 2012.
5. Das, S., Sundara, S., Cyganiak, R., R2RML: RDB to RDF Mapping Language, <http://www.w3.org/TR/r2rml/>, W3C Recommendation 27 September 2012.
6. Garijo, D., Villazón-Terrazas, B., Corcho, O., A provenance-aware linked data application for trip management and organization. In: 7th International Conference on Semantic Systems. 2011.
7. Hyland, B., Ateazing, G., Villazón-Terrazas, B., Best Practices for Publishing Linked Data, <http://www.w3.org/TR/ld-bp/>, W3C Working Group Note 09 January 2014.
8. Lopez, V., Kotoulas, S., Sbodio, M. L., Stephenson, M., Gkoulalas-Divanis, A., Aonghusa P. M., QuerioCity: A Linked Data Platform for Urban Information Management, ISWC, 2012.
9. Tallevi-Diotalle, S., Kotoulas, S., Foschini, L., Corradi, A., Real-Time Urban Monitoring in Dublin Using Semantic and Stream Technologies. The Semantic Web-ISWC 2013. LNCS Volume 8219, 2013, pp 178-194.

²⁷ <http://ciudad.linkeddata.es/browser.html>

²⁸ http://www.w3.org/2012/ldp/wiki/Main_Page