Big Data Research 7 (2017) 1–15

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Producing Linked Data for Smart Cities: The Case of Catania ☆

Sergio Consoli [a,b,*], Valentina Presutti [a], Diego Reforgiato Recupero [a,c], Andrea G. Nuzzolese [a], Silvio Peroni [a,d], Misael Mongiovi' [a], Aldo Gangemi [a,e]

[a] National Research Council (CNR), Institute of Cognitive Sciences and Technologies, Semantic Technology Laboratory, Italy
[b] Philips Research, Data Science Department, 5656 AE Eindhoven, The Netherlands
[c] Department of Mathematics and Computer Science, University of Cagliari, Italy
[d] Department of Computer Science, University of Bologna, Italy
[e] Laboratoire d'Informatique de Paris-Nord (LIPN), University Paris 13, Sorbonne, UMR CNRS, Paris, France

ARTICLE INFO

Article history:
Received 24 August 2016
Accepted 7 October 2016
Available online 3 November 2016

Keywords:
Linked Open Data for smart cities
Semantics and eGovernment
Data modeling
Interoperability
RDF data processing

ABSTRACT

Semantic Web technologies and in particular Linked Open Data provide a means for sharing knowledge about cities as physical, social, and technical systems, so enabling the development of smart city applications. This paper presents a prototype based on the case of Catania with the aim of sharing the lessons learnt, which can be reused as reference practices in other cases with similar requirements. The importance of achieving syntactic as well as semantic interoperability – as a result of transforming heterogeneous sources into Linked Data – is discussed: semantic interoperability is solved at data level in order to ease further development on top. We present a comprehensive data model for smart cities that integrates several data sources, including, geo-referenced data, public transportation, urban fault reporting, road maintenance and municipal waste collection. We show some novel ontology design patterns for modeling public transportation, urban fault reporting and road maintenance. Domain practitioners and general members of the public have been asked to play with the prototype, and fill out a survey with questions and feedbacks. A computational experiment has been also conducted to evaluate the performance of our data model in terms of practical scalability over increasing data and efficiency under complex queries. All produced data, models, prototype and questionnaire results are publicly accessible online.

1. Introduction

Intelligent or *smart* cities are characterized by the combination and use of emergent physical infrastructures, information and communication technologies (ICT), and institutional settings for knowledge sharing and innovation. The aim of smart cities consists of increasing common problem-solving capabilities for the benefit of citizens and Public Administrations (PAs), advancing the information and knowledge capabilities of the community, and opening a new cycle of innovation and e-services [1]. By injecting advanced information technologies into the social system and by increasing its innovation capabilities, cities become more open, innovative, efficient, and manageable. In addition, the smart city paradigm has strong implications in the Public Administration management, in the way of doing politics, and in the relationship among politicians, public servants and citizens. Open Government's principles [2] like transparency, participation and collaboration are central keys for the integration of citizens within the smart city paradigm.

The development of a smart city involves a multitude of technologies and processes [3]. Big data management, the Internet-of-Things, networks of sensors and smart devices, embedded systems, the Internet of users and people, Cloud Computing, are all determining a deep revolution on transport, environment, business, and government by introducing new kinds of informational and cognitive processes, such as information collection and processing, real-time alerting and forecasting, collective and crowdsourced intelligence, and cooperative distributed problem-solving and learning [4]. However, the interaction among all actors and these heterogeneous solutions still remains a challenge. Transforming our cities into the smart cities of the future encompasses incorporating technologies and key digital advances, and links them with machine-to-machine solutions and real-time data analytics. Collecting data

☆ This article belongs to BDSC.

* Corresponding author at: Philips Research, Data Science Department, 5656 AE Eindhoven, The Netherlands.
E-mail addresses: sergio.consoli@philips.com (S. Consoli), valentina.presutti@istc.cnr.it (V. Presutti), diego.reforgiato@unica.it (D. Reforgiato Recupero), andrea.nuzzolese@istc.cnr.it (A.G. Nuzzolese), silvio.peroni@istc.cnr.it (S. Peroni), misael.mongiovi@istc.cnr.it (M. Mongiovi'), aldo.gangemi@istc.cnr.it (A. Gangemi).

http://dx.doi.org/10.1016/j.bdr.2016.10.001
2214-5796/

and transforming them into tangible insights is crucial for modern innovative smart cities.

In this context, the application of Semantic Web technologies on smart cities data has an extremely high potential and practical impact [5]. They facilitate data integration from multiple heterogeneous sources, enable the development of information filtering systems, and support knowledge discovery tasks. In particular, in the last years the Linked Open Data (LOD) initiative reached significant adoption and is considered the reference practice for sharing and publishing structured data on the Web [6,7]. LOD offers the possibility of using data across different domains for purposes like statistics, analysis, maps and publications. By linking this knowledge, interrelations and associations can be inferred, and new conclusions arise.

Since cities have large amounts of data, heterogeneous in nature and with different quality and security requirements, research on the opening process, data reengineering, linking, formalization and consumption is of primary interest in smart cities [8]. The heterogeneity problem has to be tackled at different levels. On the one hand, syntactic interoperability is needed to unify the format of knowledge sources enabling, e.g., distributed query [9]. Syntactic interoperability can be achieved by conforming to universal knowledge representation languages and by adopting standards practices. The widely adopted RDF, OWL and LOD allow us to achieve such a syntactic interoperability. On the other hand, semantic interoperability is also needed. Semantic interoperability can be achieved by adopting a uniform data representation and formalizing all concepts into a holistic data model (conceptual interoperability). RDF and OWL assist us in achieving the former goal. However, conceptual interoperability is domain specific and cannot be achieved only by the adoption of standard tools and practices. The large, heterogeneous data sources in smart cities make the problem even harder, as different semantic perspectives must be addressed in order to cope with knowledge source conceptualizations. To give an example, addresses of different data entities for our city data like “Hospitals”, “Churches”, “Post offices”, “Police”, “Schools”, etc. are aligned to the same conceptual entity “Address” (characterized by properties like “street”, “address number”, “zip code”, “city block”, ...). In this way it is possible to intercross data and exploit them more, providing application developers the opportunity to easily design their city services [10]. Semantic interoperability at domain level allows making sense of distributed data and enabling their automatic interpretation. The issue of resolving semantic interoperability among different data sources is moved from the application level to the data model level. Developers are then relieved from the burden of reconciling, uniforming, and linking data at a conceptual level, and are able to build their solutions in a more sustainable and efficient way. The published data sources are made discoverable and become accessible via queries and/or public facilities, and integrated into higher-level services.

In this paper we present a methodology used to collect, enrich, and publish LOD for the Municipality of Catania, a city in Southern Italy, in the context of the project PRISMA, “PlatfoRms Interoperable cloud for SMArt-Government” [11]. We present the collected city data, describe the process and issues to create a semantic data model, and discuss a developed prototype. We describe the employed procedures, ontology design patterns and tools used for ensuring semantic interoperability during the transformation process. Our model presents these heterogeneous data in a uniform and abstracted way, which is central for the provision of “smart” applications for the city. A user assessment carried out by domain practitioners and general members of the public is also reported. Users have been asked to use the prototype[1] and fill out a survey. An evaluation of the performance of our data model in terms of practical scalability over increasing data and efficiency under complex queries is depicted as well. Some use cases that demonstrate the utility of our model are also discussed.

2. Related work

The smart city paradigm appeared at the beginning of this century as a fundamental component of the global knowledge economy. It represents a model for organizing people-driven innovation ecosystems and city-based global innovation hubs [12,13]. Integration of data and applications in smart cities has been facilitated by the development of standards. Some of them have been developed to capture city messages and events, such as the Common Alerting Protocol,[2] the National Information Exchange Model,[3] and the Universal Core.[4] Other standards have been developed to describe the city organization, e.g. the Municipal Reference Model.[5] However, most city departments still deal with ad hoc, cumbersome code to map inputs and outputs from their legacy applications. Moreover the development of standard protocols does not solve all issues. Users and services often want to get data from some specific (spatial) area and a certain period of time. In a large-scale distributed environment such as a city having highly dynamic resources and delivering a large amount of data, the usual steps of discovering, indexing, and efficiently querying data are complex tasks.

Authors in [3] discuss the issues related to the smart city development over the past five years and highlight cutting edge research and the findings from technical development projects from leading authorities within the field to capture the transition to smart cities. It also anticipates the concept that Semantic Web technologies are invaluable for integrating data of a smart city.

Recently some effort has been spent to extend them in this direction [5]. City data provided by heterogeneous sources need to be appropriately interpreted, aggregated, filtered, annotated, and combined with other data sources in order to be queried or analyzed. Here typical data integration issues arise: data need to be integrated with meta-data and other data from different streams or resources such as static databases, Semantic Web knowledge bases and social web APIs. An appropriate semantic model can help to provide an interoperable representation of data [10]. Open Government’s principles [2] like transparency, participation and collaboration are also central keys for the integration of citizens within the smart city paradigm. Open Data and access to information are essential in the process of increasing positive interactions between citizens and the city administration [8]. One of the aims of using ICTs in smart cities is to enhance the communication and interaction among citizens and public administration, as shown by the LOD approach suggested in [14]. The proposed model describes some basic, common attributes on the characteristics of data for smart ICT systems. Their method delegates details about specific streams to linked-data models, which provide on demand and service-specific external domain knowledge. Linked Sensor Middleware [15] is an attempt to build a platform that bridges the real-world city data with the Semantic Web, thanks to wrappers for real-time data gathering and publishing, data annotation and visualization, and a SPARQL endpoint for LOD querying.

[1] Publicly available at http://wit.istc.cnr.it/prisma/webcontent/home.html.

[2] Oasis Common Alerting Protocol (CAP), v. 1.2, http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html.

[3] National Information Exchange Model (NIEM), v. 3.0, https://www.niem.gov.

[4] Universal Core (UCore) Common Data Model, http://ise.gov/universal-core-ucore.

[5] Municipal Reference Model, http://www.misa-asim.ca.