

# An Interoperable Data Framework to Manipulate the Smart City Data using Semantic Technologies

Majdi Beseiso

Department of Computer Science  
Al-Balqa' Applied University  
Alsalt, Jordan

Abdulkareem Al-Alwani

Department of Computer Science &  
Engineering  
Yanbu University College  
Yanbu, Saudi Arabia

Abdullah Altameem

Department of Information Systems  
Al Imam Mohammad Ibn Saud  
Islamic University (IMSIU), Riyadh,  
Saudi Arabia

**Abstract**—During the last decade, enormous volumes of urban data have been produced by the Government agencies, the NGOs and the citizens. In such a scenario, we are presented with a diverse set of data which holds valuable information. This information can be extracted and analyzed and have a number of usages for the well-being of citizens. The major impediment to achieve this goal is the data itself, the available data are redundant, scattered and come with various legacy formats. Data interoperability, scalability and integration are paramount issues which could not be resolved unless the scattered data silos are accessible with a standard representation. In this paper, we propose a framework that resolves the data interoperability and associated challenges in the smart city environment. The framework takes the raw smart city data from several resources and stores them in a NoSQL database. The framework transforms the scattered data into machine-processable data. Besides, the database is linked with an API and simple dashboard for further analysis, which can be utilized to build big data applications based on urban data so that government agencies can get a summarized overview of resource distribution.

**Keywords**—Smart cities; Smart Data Integration; Big data; IOT; Software architecture

## I. INTRODUCTION

During the last two decades, several new nations emerged and played their profound role in changing human lifestyle. A smart city [1] is one of concepts that argue intellectual and social capital should be considered alongside with city's tangible assets at the time of measuring the urban performance. Moreover, a city is considered to be smart when human and social capital with modern information work in conjunction to fortify the economic development of a municipality. The economic development of a city leads to better life quality and improved citizen services. Several modern communication modalities are being incorporated to congregate crowd-feedback related to urban environment such as mobile devices, landlines, banking and social networks. In other words, to move forward towards smart city, ICT (Information and Communications Technology) and IoT (Internet of-Things) should be considered as key factors [2].

Easy and inexpensive availability of sensors made them favorable to be one of the important components in smart city infrastructure. Using sensors enhances getting real time feedback (i.e. city temperature, traffic flow and to measure the pollution in the air [3]. Moreover, with the conventional communication networks such as telephone line; a network of

customized sensors could be established with a centralized command and control room to observe specific city real time events. Combining these accessible data sets, spatially and temporarily leads to huge data sets must be analyzed.

In the few last decades, the scarcity of urban data was huge. Data scientists and statisticians were looking for ways to produce new urban data [4]. However, with the age of Internet of Things (IOT) the situation has been reversed. Diverse genres of urban data becomes everywhere in several format and presentation. The traditional data analysis tools became inefficient to manage the continuous stream of data. Providing actionable information requires new software architecture and data analytic tools. Several techniques have been developed, such as data warehousing and clustering and even modern cloud-based data management to organize this steadily growing urban data. Because of their rigid schema and legacy formats, it is considered an arduous activity to be integrated with all the data sets at one central repository to get the aggregated view [5].

With the emerging of the semantic web, it was introduced as a solution to resolve several data interoperability [6] and integration issues. As the Semantic Web provides a common framework that allows to transform the available unstructured and semi-structured data into a "web of data" in order to promote its widespread usage application, enterprise, and community boundaries [7]. This in turn is commonly used in several platforms to handle real-time data and historical data in smart city services. These dashboards provide intelligent decisions based on real-time and historical data. This work introduces a framework that resolves several data interoperability and associated challenges in the smart city environment. The paper is organized as follows. Section II provides literature review of similar frameworks. The model is discussed in Section III, followed by the methodology in Section IV. The experimental results are presented and discussed in Section V. Finally, Section VI concludes the paper and presents the future work.

## II. LITERATURE REVIEW

Currently, several Smart Cities initiatives around the world have become reality. In Spain, more than 20,000 sensors were installed for measuring air quality, monitoring parking spaces, distributing electricity, optimizing garbage collection and regulating light intensity... etc [8]. Moreover, in Rio de Janeiro, an Operations Center was established specifically to

analyze the data collected from sensors and actuators throughout the city [9]. The main target is to real time monitoring for the climatic conditions in order to predict natural disasters. Besides, reducing the response time in traffic accidents.

In [10], introduced a framework which combines the cloud computing and smart city; instead of storing the data gathered from different silos. Data should be maintained in cloud environment. This approach will be helpful when we have different devices for communication with different data exchange protocols [11]. The acquired data from different sensors usually followed in a certain format. A preprocessing step applied in the cloud to transform the data before storing to a specific format. Another framework introduced in [12] to resolve the personal data privacy and limited usage. It proposed a digital identity and trust control mechanism.

A distributed software infrastructures are introduced in [13] for general purpose services in power systems. The aim of the software architecture is to handle the interoperability across heterogeneous devices to manage a Smart Grid by creating a secure peer-to-peer network. A SemsorGrid4Env is a service-oriented architecture to design open large-scale semantic-based sensor network applications for environmental management[14]. It aims to enable rapid development of tiny applications and allows the integration for both real-time and historical data from several resources. This architecture is designed to environmental management and cannot be applied seamlessly to a city. At last, huge ICT companies like Apple, Google, Amazon, and Microsoft introduced Industrial ICT platforms for products and service. They provide also a place for external stakeholders to design a new customized platform according to the needs and requirements. Reusable common components and technologies form a basis for an industry platform, which is described by openness to external parties. These platforms aim to accelerate development time and to improve utilization of the digital technologies and ICT developments [15].

### III. TECHNIQUES TO MODEL DATA

One of the main issues in current smart city platforms is the rigid schema. Using a specific platform for electricity and others for gas and water. Besides using a tool for real time data for the same services, another tool is used for historical data and further data analysis. Moreover, in case two documents in XML format follow different schemas then auto merging of these documents is not possible which is essential in some cases. The concept of interoperability argues that data should be accessible in any computing environment. Employing the semantic data structure can decrease the interoperability problem. One of the current popular semantic data structure is the resource description framework (RDF). If datasets are available in RDF by sharing the common semantic universal resource identifiers (URI), the datasets could automatically be merged on the fly and the user can query through different sets of data [16].

#### A. Linked Open Data

Linked open data [17] can be defined as a semantic framework which establishes the links among diverse data sets. These data sets can be located in an organizational territory or can be geographically scattered. In addition to that, a linked data infrastructure could be established between the two databases located as in different data centers. Principally, linked data provides a guideline to publish data on the internet in a way that will be machine readable and human understandable. Linked data essentially employs HTTP and URI to expose and publish the data on the web. The primary format to organize the data as linked open data is RDF which is a W3C recommended semantic representations format.

#### B. Semantic Vocabularies

Linked Data is based on two technologies HTTP and the URIs. Mostly, web graphics and data are represented with the Uniform Resource Locators (URLs) however; the Uniform Resource Identifiers provide upper level or more generic representations of all sort of entities available on internet. Therefore, a HTTP enables the simplest universal scheme that could be utilized to extract the data about city landscapes and road structures in our case.

#### C. Triple Stores

The main difference between a triple store and a conventional database which manages the data as tables, is that a 'triple store' accumulates the RDF data and provides the inference over it [18]. The No SQL triple store, stores the data in key value pair format. It is often consider as a best storage solution where one have to deal with continuous stream of data such as social media contents management.

#### D. NoSQL

A No SQL database is used to store heterogeneous data from several resources. It was picked due to its scalability and performance in big data. Besides, all data are stored in terms of JSON format which could be used easily in web services. The results are typically returned in one or more machine-processable formats.

### IV. METHODOLOGY & PROPOSED FRAMEWORK

The aim of this work is to utilize the Linked data and other supporting semantic technologies to resolve the data interoperability and integration challenges. The framework introduces a semantic based programmable Application program interface (API). Using API simplifies the process of information gathering for developers or different users. The simplicity of the user interface enhances the process of gathering hidden information. The framework is based on four main functioning stacks. i) Data scraping layer, ii) Data adaptation layer. iii) Data management layer. iv) Application layer. As shown in Figure 1 below, the graphical view of the proposed framework manipulates the smart city data.

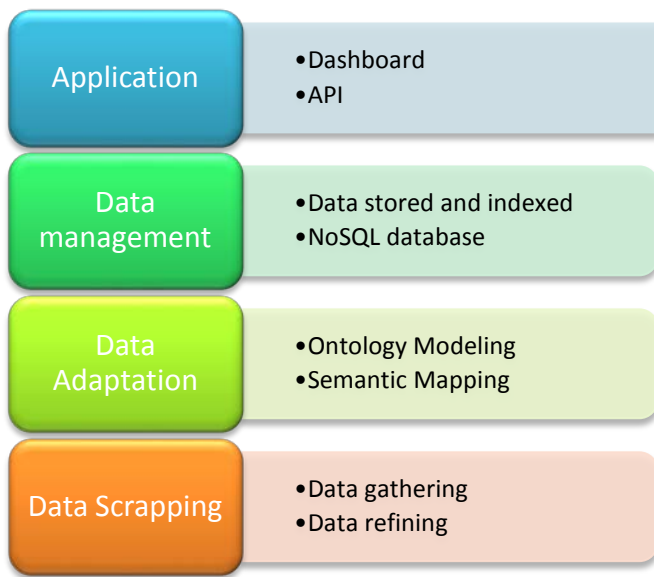


Fig. 1. Graphical View of Proposed Framework to Manipulate the Smart City Data

### A. Data Scrapping layer

The purpose of Scrapping layer is gathering data from several resources. After that data is cleaned using several algorithms based on the data format. This layer is based on two components as follows:-

#### 1) Data gathering

Data is scraped from numerous open data sources (i.e. census data bureau, NGO data and other open data that are available for public).The Saudi E-Government and census bureau data are providing huge data for this purpose. Currently, available data are mostly provided in spreadsheet, text or in database format. A data refinement process is necessary at this stage; data are invoked to check for data redundancy and incompleteness.

#### 2) Data refining

The underlying algorithm automatically checks common data cleansing issues (i.e. duplication and incompleteness of Meta information, and missing values). Besides checking further invokes, a number of procedures to clean the data are carried out. In this work, Open Refine is used to clean up the data, because it is one of the powerful tools for working with messy data and its simplicity of cleaning and transforming the data from one format into another format. In the beginning, Open Refine handles the raw data and provides various data manipulation tools to treat the data. Then data are stored the RDF files in triple store: a purpose built database for the storage and retrieval of triples through semantic queries. While a backup of data are stored in No SQL Database in a remote server for comparison and analysis purpose.

### B. Data Adaptation layer

This layer acts as a linked data generation factory. It holds ontology modeling and semantic mapping. It has two main tasks, ontology modeling followed by the semantic mapping which is used to semantify the data of smart city.

### 1) Ontology Modeling

Ontology modeling process can be composed of a number of subtasks such as designing and development. To analyze the data retrieved from numerous sources of smart city: a new ontology (SC-Ont) was developed. The SC-Ont ontology can be divided into three main modules. The core module provides the semantic vocabularies about the smart city environment. Followed by the provenance module which exhibits the information about the data producer and consumer. At last, the linked module, which anchors the semantic vocabularies in SC-Ont with other available ontologies, as it promotes the linkage of semantic vocabularies. This approach makes it more useful in linked open data environment. The more one has data in complex connection the more diverse information can be pulled out as shown in figure 2.

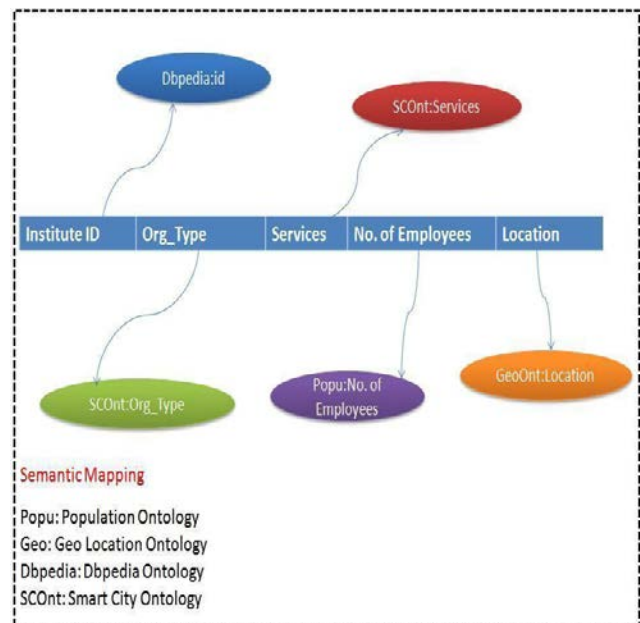


Fig. 2. The Semantic Mapping for Smart City Data

### 2) Semantic Mapping

It is done programmatically, basically the developer invokes the data stored in a database, ontology classes, then they interlink semantically. The result of this process is a new RDF file format. In RDF, data is arranged as subject predicate object triple. As the RDF data follows a common semantic modeling. Therefore automatic integration of various data sets would become a common job.

### C. Data management layer

This would be a usability layer to manage and handle the data for the developer. No SQL databases are used to store and handle heterogeneous data. In addition, Semantic web services are developed which would be helpful for the user to pull out the desired data from the central repositories. Web services are platform independent and do not require any installation.

### D. Application layer

An API and the developer tools are introduced in this layer. In addition, optimizing the RDF data according to linked data principal and exposing data as linked data API.

### 1) Linked Data Optimization

Linked Data Optimization (LDO) can be defined as a process that tunes the linked data. During the generation of linked data process. It has been noticed that most of the URI provider do not keep themselves updated or not usually well synchronized with the linked data cloud. Finding such kind of weak bindings and recommending new semantic vocabularies, the underlying algorithm checks the status of the data links. If it finds any data without link or data with a deadlink, it recommends an appropriate semantic vocabulary to provide keeping the semantic mapping alive.

### 2) API generation

To access the RDFized data programmatically, a simple API was introduced. The main purpose is to simplify the used access to the data by defining a set of functionalities. All the classes and methods are well documented. A developer can access any method to retrieve certain type of information and can also integrate with any semantic and syntactic system.

## V. RESULTS

The smart city environment usually incorporates a number of sensors, in which carbon emission sensor is one. Such kind of sensors continuously monitors the carbon dioxide emission in the urban air and report back to control center. The diseases caused by the air pollution could be controlled, keeping in mind the information of the polluted area. Such kind of information is very helpful in town planning and before making a decision to allot a land for a children's park. The conventional system used to collect and analyze such kind of information uses the sophisticated commercial tools to integrate the data. However, by using our developed solution, the data integration will become automatic.

To demonstrate the model, we compare our framework with District Information Modeling and Management for Energy Reduction (DIMMER) [19]. DIMMER is a distributed IoT software architecture to collect and correlate heterogeneous energy data into a distributed smart archive system for data analysis and management. As shown in table 1, both models aims to sharing the data among different stakeholders in Smart City scenarios. DIMMER consists of: i) Data-source Integration Layer; ii) District Services Layer; iii) Application Layer. It collects the data from heterogeneous IOT devices through device connectors while in the proposed framework, the data are collected from heterogeneous software resources on the web portals. Both models use services to access heterogeneous data sources and manipulate the data in each source, but in the proposed framework, the data format are unified into RDF format to make advantages for complex analysis from different data sets with heterogeneous data. That is the main reason for using No SQL database which handles heterogeneous data and performs faster than relational databases. In terms of customization, our proposed model could be applied in multiple servers which would not be an easy task in the DIMMER framework. At last, both models provide API and web interface for handling and retrieving data.

TABLE I. DIMMER AND PROPOSED FRAMEWORK FACTORS COMPARISON

Factor	Proposed framework	DIMMER
Source of data	Using web portals to fetch the data.	Manage and correlate heterogeneous data from several IOT.
Database	Includes a single point databases to store data in unified format.	Mange different DB and export data through web services.
Network	Data shared through web dashboard	Data shared through DIMMER network.
Multi-tenant	Could be applied	Could not be applied.
API	Available	Available

## VI. CONCLUSION

The proposed framework in this study has the adequate ability to resolve the data interoperability issues as the data will be available in RDF format. Furthermore, semantic framework showed potential in addressing the challenges pertaining to dynamic data integration and information retrieval in data science domain. This was made possible as common URI scheme was employed to represent the data. The proposed framework will also not only promote the usage of existing data but it will also allow any potential user to reuse any component of the system to improve upon existing smart city solutions as proposed in the potential applications.

### REFERENCES

- [1] S. P. Bingulac, "On the compatibility of adaptive controllers," in Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory, New York, 1994, pp. 8-16.
- [2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22-32, Feb 2014.
- [3] Bowerman, B et al. "The vision of a smart city." 2nd International Life Extension Technology Workshop, Paris Sep. 2000.
- [4] Su, Kehua, Jie Li, and Hongbo Fu. "Smart city and the applications." Electronics, Communications and Control (ICECC), 2011 International Conference on 9 Sep. 2011: 1028-1031.
- [5] Deren, LI, Shao Zhenfeng, and YANG Xiaomin. "Theory and Practice from Digital City to Smart City [J]." Geospatial Information 6 (2011): 002.
- [6] Santucci, Gérald. "From internet of data to internet of things." International Conference on Future Trends of the Internet 28 Jan. 2009.
- [7] Rizzo, Giuseppe, Federico Morando, and Juan Carlos De Martin. "Open Data: la piattaforma di datiaperti per il Linked Data." Informatica e diritto 20 (2011): 493-511.
- [8] L. Sanchez, L. Muñoz, J. A. Galache, P. Sotres, J. R. Santana, V. Gutierrez, et al., "SmartSantander: IoT experimentation over a smart city testbed," Computer Networks, vol. 61, pp. 217-238, 2014.
- [9] C. Gaffney and C. Robertson, "Smarter than Smart: Rio de Janeiro's Flawed Emergence as a Smart City," Journal of Urban Technology, pp. 1-18, 2016.
- [10] R. Lea and M. Blackstock, "City Hub: A Cloud-Based IoT Platform for Smart Cities," 2014 IEEE 6th Int. Conf. Cloud Comput. Technol. Sci., pp. 799-804, 2014.
- [11] Karnouskos, Stamatis, and ThiagoNass De Holanda. "Simulation of a smart grid city with software agents." Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on 25 Nov. 2009: 424-429.

- [12] Mitton, Nathalie et al. "Combining Cloud and sensors in a smart city environment." *EURASIP Journal on Wireless Communications and Networking* 2012.1 (2012): 1-10.
- [13] E. Patti, A. L. A. Syri, M. Jahn, P. Mancarella, A. Acquaviva, and E. Macii, "Distributed software infrastructure for general purpose services in smart grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1156–1163, March 2016.
- [14] "SemsorGrid4Env," Available: <http://www.semsorgrid4env.eu/>.
- [15] Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, 31(3), 417-433.
- [16] Deren, LI, Shao Zhenfeng, and YANG Xiaomin. "Theory and Practice from Digital City to Smart City [J]." *Geospatial Information* 6 (2011): 002.
- [17] Biron, Paul, Ashok Malhotra, and World Wide Web Consortium. "XML schema part 2: Datatypes." *World Wide Web Consortium Recommendation REC-xmlschema-2-20041028* (2004).
- [18] Sirin, Evren, and Bijan Parsia. "SPARQL-DL: SPARQL Query for OWL-DL." *OWLED* 6 Jun. 2007.
- [19] E. Patti and A. Acquaviva, "IoT platform for Smart Cities: Requirements and implementation case studies," 2016 IEEE 2nd Int. Forum Res. Technol. Soc. Ind. Leveraging a better tomorrow, pp. 1–6, 2016.