6-2014

Semantic Modelling of Smart City Data

Stefan Bischof

Athanasios Karapantelakis

Cosmin-Septimiu Nechifor

Amit P. Sheth
Wright State University - Main Campus, amit.sheth@wright.edu

Alessandra Mileo

See next page for additional authors

Authors
Stefan Bischof, Athanasios Karapantelakis, Cosmin-Septimiu Nechifor, Amit P. Sheth, Alessandra Mileo, and Payam Barnaghi

Semantic Modelling of Smart City Data

Stefan Bischof*[1], Athanasios Karapantelakis†[2], Cosmin-Septimiu Nechifor‡[3], Amit Sheth§[4], Alessandra Mileo¶[5] and Payam Barnaghi‖[6]

[1]Siemens AG Österreich, Vienna, Austria
[2]Ericsson Research, Stockholm, Sweden
[3]Siemens SRL, Romania
[4]Kno.e.sis Center, Wright State University, Dayton, OH, USA
[5]Digital Enterprise Research Institute, National University of Ireland, Galway
[6]University of Surrey, Guildford, Surrey, United Kingdom

1 Introduction

Recent advancements in communication technologies for providing ubiquitous Internet access as well as advancements on reduction of cost and form-factor of mobile devices and sensors are seen as an enabler for the Internet of Things (IoT). The industry predicts an interconnected world of 50 billion devices by 2020[1]. The Web of Things (WoT) relies on the connectivity service of IoT to create services and applications exploiting the IoT data [1].

Cities present an opportunity for rendering WoT-enabled services. According to the World Health Organization, population in cities will double by the middle of this century[2], while cities deal with increasingly pressing issues such as environmental sustainability, economic growth and citizen mobility. In this paper, we propose a discussion around the need for common semantic descriptions for smart city data to facilitate future services in "smart cities". We present examples of data that can be collected from cities, discuss issues around this data and put forward some preliminary thoughts for creating a semantic description model to describe and help discover, index and query smart city data.

*bischof.stefan@siemens.com
†athanasios.karapantelakis@ericsson.com
‡septimiu.nechifor@siemens.com
§amit@knoesis.org
¶alessandra.mileo@deri.org
‖p.barnaghi@surrey.ac.uk

[1]http://share.cisco.com/internet-of-things.html
[2]http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/

2 Description of Smart City Data

Table 1: Information that could be potentially retrieved about cities.

Data Category	Owner (Data Publisher)	Data Description	Sampling
Transport	Traffic Authority	Maps of Cities (Roads, Street Names, POIs, subway and bus stations, etc.) [3]	Static
	Municipality	Public Transport Schedules [4]	Semi-Dynamic
	Traffic Authority	Transport Authority Updates (Roadwork, traffic status, etc.) [5]	Dynamic
Air Quality	Env. Agency	Particle concentration [6]	Dynamic
Traffic	Traffic Authority	Number of vehicles passing between two points, speed [7]	Dynamic
City Events	Cultural Groups	Entertainment (movie/theater plays)	Semi-Dynamic
Municipal Services	Municipality	Library Data [8]	Dynamic
		Waste Collection Data [9]	Dynamic
	Private Company	Parking Meters [10]	Dynamic
Citizen data	Private Individuals	Social Media Information: Tweets, Status updates and blog posts, popular places ("check-ins")	Semi-Dynamic
		Household Energy Consumption	Semi-Dynamic
Health data	Private and Public	Relevant information about potential or confirmed sources of health threats	Dynamic

[3] Openstreetmap offers geographical data under the Open Data Commons Open Database License (ODbL): `http://wiki.openstreetmap.org/wiki/API`

[4] The Danish authority "Rejseplanen" offers an API for timetables of public transport such as buses and trains: `http://labs.rejseplanen.dk/files/api/rest_documentation_latest.pdf`

[5] The Swedish transport authority Trafikverket issues updates on the status of the transport network as RSS feeds: `http://www.trafikverket.se/Aktuellt/Trafikverkets-RSS-floden/`

[6] The city of Brasov in Romania offers precipitation measurements as open data: `http://cdr.eionet.europa.eu/ro/eu/eiodata`

[7] The city of Aarhus in Denmark offers traffic data measurements as open data: `http://www.odaa.dk/dataset/realtids-trafikdata`

[8] Books on loan measurement from the city of Aarhus: `http://www.odaa.dk/dataset/pendlertal-fra-aarhus-kommunes-biblioteker`

[9] Waste collection data from households, provided from the city of Aarhus `http://www.odaa.dk/dataset/datasaet-om-sammensaetningen-af-affald`

[10] Parking meter data from the city of Aarhus `http://www.odaa.dk/dataset/parkering`

Table 1 shows an example of the type of data that can be collected from cities. This data is currently collected from cities and in companies participating in the EU FP7 Citypulse project[11]. The sampling column relates to the periodicity of the incoming data. Static indicates that the data is never updated and the dataset is used as reference (any updates are manual). Semi-dynamic sampling means that the data is updated periodically, whereas dynamic means continuous updating.

3 Challenges and Issues of Smart City Data

Smart city data sources offer a variety of data to be processed. The observation and measurement data is usually aggregated and filtered after collection. The data is then transferred often over several systems and transformed to data representation useful for interoperable publication. These published data sources are then made discoverable and become access-able via query and/or publish/subscribe facilities. Over these access interfaces the data is eventually integrated into higher-level services and applications [2]. The heterogeneity created on the different levels of processing this data gives rise to challenges in several dimensions.

Sensory devices devices measure different types of observations such as light, temperature, or sound. The different sensors and devices will provide data of different and even changing *quality*. The data is often continuous and over time data quality, data validity and device availability can change, thus resulting in highly *dynamic* data streams. Applications may select data sources by data quality or device reliability but also ignore data sources which are untrustworthy.

Since sensory devices will also record sensitive or private data the issues of *privacy* and *security* need to be considered and addressed over the whole processing pipeline. Encryption of sensitive data is the accepted best practice to protect data from unauthorized access during transmission and storage. Another mean to protect privacy of individuals is data aggregation.

Even devices of the same type will deliver data in *heterogeneous* formats or different units of measurements. The heterogeneity issues can partly be addressed by meta-data and semantics. Semantic annotations can also help mapping data between differing schema models on a higher-level. When semantically annotating data streams one has to carefully weigh *expressibility* vs. *complexity* as well as the sheer *volume* of the generated data which has to be processed. In typical scenarios the meta-data is larger than the actual measurement data. The data volume makes early aggregation and filtering necessary. Additionally the whole data processing pipeline has to be designed with scalability in mind.

Users and services will often want to get data from some specific (spatial) area and a certain period of time. In a large-scale distributed environment with highly dynamic resources such as sensors delivering a large amount of data, the usual steps of *discovering*, *indexing*, and efficiently *querying* data are complex

[11] `http://www.ict-citypulse.eu/`

tasks. Semantic Web technologies can help to some extent and are currently extended in this direction by the CityPulse project.

Eventually the data provided by semantically annotated streams, which is now integrated, aggregated, filtered, and combined via querying, usually need to be interpreted, combined with other data sources and analyzed. In this step the usual *data integration* issues arise in another incarnation: we have to integrate data with meta-data and also with different types of data from other sources such as static databases, Semantic Web knowledge bases or social web APIs. Again a semantic model can help to create an interoperable representation of data is provided by various heterogeneous resources.

4 Semantic Interoperability

As described in the previous section, smart city data are heterogeneous in nature (delivery format, point of origin, periodicity, etc.) and have different privacy, security and quality requirements. To realise the potential of a smart city, multiple of these data sources have to be combined. To address the issues of interoperability at sensor level, a W3C incubator group developed the Semantic Sensor Network (SSN) ontology [3].

SSN describes not only sensor device capabilities, but also organises the sensors into systems and describes processes that model sensor operations and can work across multiple domains. The goal is to correlate measurement data with capabilities of sensors (and sensor systems), however the descriptions about observation and measurement data are generic and cannot be used to annotate the data with domain knowledge - specific to applications. Therefore, SSN by itself cannot be used to describe smart city services (scenarios) in detail, as each service has its own quality requirements, relies on its own set of sensors, has different demands on data ownership (security, privacy concerns) etc.

Previous research has suggested building a linked-data approach for stream annotation [4]. According to this approach, external domain knowledge about the data can be provided on request - and can be specific per service rendered (e.g. quality description, sensor capabilities, etc.). The model proposed in [4] describes some basic, common attributes on the data stream but delegates details about the specific streams to other models (linked-data models).

5 Suggestions for Discussion

We are currently building a linked-data model for semantic annotation of data streams in smart city environments. The EU FP7 CityPulse project is working on linked-data descriptions for smart city data . The project also provides a set of smart city data access and processing scenarios. This can help to identify a set of common properties among smart city data (see table 1) that can be used for semantic modelling and description of multi-modal data in smart city applications and services. Going beyond the details of the model design, the common properties identified can help contextualise smart city data and simplify the connection between the descriptions in the model and data stream operations such as discovery, indexing and querying from applications, services or systems using these data. We therefore suggest the following topics of discussion around the design of a model for smart cities:

- Smart city data stream annotation: descriptions for data privacy, security, quality. Flexibility to support heterogeneity in observations, dynamicity of data streams - scalability.
- Data contextualization for optimised data stream discovery, indexing and querying. To start with, we may consider categorization of smart city data in hierarchical form, from general domain of observation (transportation, events, healthcare, municipal services, etc.) to observed physical phenomena (traffic, theatre plays, hospitals, water level, etc.) down to units of measurement (cars per minute, event timestamp, hospital capacity, percentage of water reserves, etc.). Outcome of the discussion will also influence the design of the model in the previous step. Refer to CityPulse (and other sources) for smart city scenarios to understand requirements from a smart-city service perspective (top-down approach).
- Use of linked-data for enriched processing of the annotated data. Consider extending similar approaches such as SECURE system, which uses background data for event detection [5].

Acknowledgments. This work is supported by the EU FP7 CityPulse Project under grant No.603095. `http://www.ict-citypulse.eu`

References

[1] S. Gustafson and A. Seth, "The web of things," *Computing Now*, vol. 7, no. 3, 2014. [Online]. Available: http://www.computer.org/portal/web/computingnow/archive/march2014

[2] P. Barnaghi, A. Sheth, and C. Henson, "From data to actionable knowledge: Big data challenges in the web of things," *Intelligent Systems, IEEE*, vol. 28, no. 6, pp. 6–11, Nov 2013.

[3] M. Compton, P. Barnaghi, L. Bermudez, R. Garcıa-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. D. Kelsey, D. L. Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, and K. Taylor, "The ssn ontology of the w3c semantic sensor network incubator group," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 17, no. 0, 2012.

[4] P. Barnaghi, W. Wang, L. Dong, and C. Wang, "A linked-data model for semantic sensor streams," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, Aug 2013, pp. 468–475.

[5] P. Desai, C. Henson, P. Anantharam, and A. Sheth, "Secure: Semantics empowered rescue environment (demonstration paper)," *4th International Workshop on Semantic Sensor Networks 2011 (SSN 2011)*, pp. 115–118, 2011.