Ontology-Based Data Integration from Heterogeneous Urban Systems: A Knowledge Representation Framework for Smart Cities

Achilleas Psyllidis

Abstract

This paper presents a novel knowledge representation framework for smart city planning and management that enables the semantic integration of heterogeneous urban data from diverse sources. Currently, the combination of information across city agencies is cumbersome, as the increasingly available datasets are stored in disparate data silos, using different models and schemas for their description. To overcome this interoperability barrier, the presented framework employs a modular and scalable system architecture, comprising a comprehensive ontology capable of integrating data from various sectors within a city, a web ontology browser, and a web-based knowledge graph for online data discovery, mapping, and sharing across stakeholders. Linked Data, Semantic Web technologies, and ontology matching techniques are key to the framework's implementation. The paper ultimately showcases an application example, where the framework is used as a semantic enrichment mechanism in a platform for urban analytics, focusing particularly on human-generated data integration.

A. Psyllidis
Chair of Hyperbody – Digitally-driven Architecture, Department of Architectural Engineering & Technology,
Faculty of Architecture and the Built Environment,
Delft University of Technology (TU Delft), 2628 BL, Delft, The Netherlands

Email: A.Psyllidis@tudelft.nl

1 Introduction

At present, large-scale urban data are produced from diverse sources and increasingly become publicly available from governmental authorities. Each source reflects a specific aspect of the urban environment and the generated datasets vary in scale, speed, quality, format, and, most importantly, semantics. In principle, municipal data are mostly static or semi-static, in the form of demographic records, such as census statistics on population, age and gender distribution, average income, and land uses. Besides static records, planning stakeholders gradually utilize dynamic streams of information stemming from sensor resources, in relation to transportation flows, weather status, and environmental conditions. However, in the current context, city-related organizations create and operate on datasets based on each sector's particular purposes and problems at hand. Any correlation of information across different departments is presently performed in a manual fashion, hence requiring great amounts of time and effort.

This data interoperability barrier is further strengthened by the use of different data models and schemas across city agencies [13]. Equally, in urban planning and decision-making procedures each urban system is often approached separately, as a result of the complexity and diversity among the city data silos. This leads to an emerging need for frameworks that allow for interoperable urban data exchange and reuse, by systematically harnessing the combined potential of diverse data sources. Such an approach is particularly important within a smart city framework.

Motivated by this challenge, the paper introduces a novel knowledge representation (KR) framework for planning smart cities that enables the semantic integration of heterogeneous urban data stemming from diverse sources. The framework consists of three main components, namely (a) an ontology that formally describes the different city sectors, urban systems, the respective data sources, and defines the relations among them, (b) a web ontology browser that provides full access to the aforementioned ontology online, and (c) an interactive knowledge graph to facilitate the exchange of shared semantic definitions and correlations among multiple stakeholders, through a web-based graphical user interface (GUI).

The ontology for smart city planning and management presented in this paper is currently the first one in this domain that defines concepts and named relationships, linking both the different urban systems together and the data generated within them. By additionally implementing alignments with multiple external ontologies and controlled vocabularies, it can be used to semantically annotate data from heterogeneous city sectors in a machine-processable way. Moreover, the developed web ontology browser provides developers and other stakeholders with online access and navigation possibilities through the complete ontology hierarchy. By encompassing Linked Data [3] and Semantic Web principles [1, 7], it allows concept discovery, browsing over named relations, and links to aligned concepts and properties. Similarly, the interactive knowledge graph constitutes a novel visualization system for exploring the various entity clusters and relations captured by the ontology. It allows city stakeholders to perform term searching, concept clustering, and relations' discovery through an intuitive web-based platform.

The aim of the introduced KR framework – coined *OSMoSys* – is to assist in the mutual interaction among urban planners, city managers, decision-makers, and other city stakeholders by providing tools that are capable of operating across the diverse data silos. It also aims to simplify the processes of data mapping, sharing, and reuse across agencies. The contributions of this work are mainly three.

1. A novel ontology for smart cities that enables data integration from heterogeneous urban sectors and sources, irrespective of the different qualities, speeds, and scales. An additional innovative aspect of this ontology refers to the integration of concepts for human-generated data, coming from smart devices and social media platforms, being increasingly important sources of knowledge for the domains of urban analytics and decision-making.
2. A web ontology browser and an interactive knowledge graph that complement the processes of concept discovery and data mapping, by respectively providing an easy-to-use online ontology navigation platform with links to external data models and a web-based visualization system for intuitive concept exploration[1].
3. An application example of using the KR framework as a semantic enrichment mechanism in a recently developed system that supports the analysis and integration of heterogeneous urban data.

---

[1] The web-based knowledge graph and ontology browser are accessible via the following link: http://osmosys.hyperbody.nl

The rest of the paper is structured as follows. In Section 2 the related work is described. Section 3 introduces and presents the proposed method. Section 4 outlines the architecture of the KR framework and describes its three main constituent components. In Section 5 a use example of the knowledge framework is showcased. Ultimately, Section 6 summarizes the conclusions and discusses future lines of research.

2 Related Work

In the wider domain of computational urban planning there have recently been several attempts to tackle the complexity of contemporary city systems and the increasing amount of data they generate, through the use of ontology-based models [15]. However, in most of these cases, the developed ontologies either cater for an individual urban system or represent specific attributes of the urban environment. To this end, creating new ontology models, without establishing links among them, generally poses a risk as to merely adding up to the multiplicity of already existing, yet discordant, data models and schemas. The examples described in the following paragraphs showcase relevant approaches that serve as stepping-stones on the development of the broader framework proposed in this work.

To sufficiently describe urban space and facilitate interoperability among urban design applications, Montenegro et al. [14] developed an ontology for land use planning. Having as its starting point the Land Base Classification Standards (LBCS) by the American Planning Association, the ontology provides semantically annotated land use descriptions of spatial data. The main limitations of this study relate to the adherence of the developed ontology to a single model and, subsequently, to a particular aspect of the urban environment. However, the complexity of urban space requires links to multiple external ontologies or conceptual schemas in order to be comprehensively described and understood.

Métral et al. [13] describe the benefits of following an ontology-based approach for providing semantically enriched 3D city models. Besides presenting an ontology of the CityGML model [11], which incorporates both geometrical and topological concepts of urban objects, the paper discusses the Ontology of Urban Planning Processes (OUPP), with a focus on soft mobility. The work presents similar limitations to the previous one, yet it indicates an alignment with concepts from an external model, namely the Ontology of Transportation Networks (OTN). A further collection of ontologies, specifically created for projects in the domains of urban planning and development, is presented in [10]. In this case, the described knowledge models are accompanied by real-world case studies, covering an array of aspects that range from urban mobility to building regulations.

Closer to the intentions of this paper is the work of Bellini et al. [2]. Their paper describes the construction process of a knowledge base for smart-city services. An interesting element of this work is the introduction of an ontology for smart-city services that allows the integration of heterogeneous data from multiple sources, supported by links to external vocabularies. Nevertheless, it is mainly focused again on a single urban system (i.e. transport) and thus faces similar limitations to those of the aforementioned examples.

Unlike previous approaches, the KR framework for smart cities presented in this paper offers a unique framework that combines an innovative ontology, allowing data mapping from different urban systems, with web-based tools for intuitive concept sharing among city stakeholders.

3 Method

To address the challenges posed by cumbersome and disparate data silos across city agencies, the demonstrated KR framework follows an ontology-based approach to integrating urban data from different sectors. To this end, a comprehensive ontology for smart cities was developed that formally describes the different urban systems (e.g. energy, waste, water, transport, buildings etc.); the respective data sources (e.g. spatial statistics, sensor data, social data etc.); the smart city technology enablers (e.g. interoperability types, connectivity, computational resources etc.); and defines the relations among them. It also represents the various urban sectors/agencies and includes broader concepts that allow data mapping from different departments. The intention is to enable the exploitation of the reasoning potential inherent to ontologies, as well as the data discovery possibilities through expressive querying.

However, when combining data of different speeds (i.e. static, semi-static, streaming), scales, and qualities, stakeholders are generally confronted with syntactic, schematic, and semantic heterogeneities. These respectively refer to differences in data syntax, database schemas, and meaning interpretation. Notwithstanding, emphasis in this case is put on semantics. Providing shared semantic definitions is essential for both data integration and reuse into different contexts. Especially when it comes to sharing information across agencies that might refer to the same component, yet originally using a different term or data model to describe it. Therefore, entities of the developed ontology for smart cities are aligned with concepts and properties from multiple external ontologies and vocabularies, to ensure the interrelation among city sectors and further facilitate the data integration process. To achieve this, ontology matching techniques [9] are used. As the proposed semantic model aims to provide an overarching framework that encompasses the various urban systems, it is crucial for it to enable links among the already existing knowledge resources, which cater for more specific aspects of the urban environment. In the previous section it was shown that single sector-specific ontologies might well form such knowledge resources, yet insufficient in addressing cross-sector concepts. Mapping to a global framework is thus paramount, and the proposed one in this paper is capable of providing it.

The above-described domain ontology constitutes one of the three components the *OSMoSys* KR framework introduces, as part of its data integration approach. To further facilitate the collaboration among the different city stakeholders, it additionally incorporates a web ontology browser and a web-based knowledge graph for smart cities. The role of the two latter components is to provide online, intuitive, and interactive user interfaces for exploring the underpinning ontology. In this way, multiple stakeholders of different backgrounds, such as urban planners, decision-makers and other municipal authorities, possessing various levels of expertise, can equally access and discover concepts and relations represented in the semantic model. The web ontology browser incorporates the complete taxonomy of ontology concepts (classes, object and data properties, individuals etc.) and offers multiple exploration possibilities online to people who are not trained in knowledge representation formalisms, without requiring dedicated software platforms to be downloaded. By adopting Semantic Web and Linked Data principles, it provides direct links to other knowledge repositories that cater for more specific purposes, thus encouraging concept and data model reuse. Finally, the web-based knowledge graph offers an interactive GUI for visual exploration, navigation through linked open data, and concept discovery. It visually represents the semantic model as a dynamic graph of concepts and relations, using state-of-the-art web and data visualization technologies.

4 OSMoSys: The Knowledge Representation Framework

The *OSMoSys* KR framework for smart city planning and management adopts a modular architecture, so as to address the several scalability challenges posed by the diversity of urban data, as well as by the complexity of cities and respective agencies. These modules namely refer to (a) the Ontology for smart city planning and management, (b) the Web Ontology Browser, and (c) the web-based Knowledge Graph. Each of the aforementioned components has a specific functionality and serves a particular purpose. As such, they can be deployed independently, yet all three are simultaneously linked to each other. Owing to this modular architecture, the KR framework can be extended, accommodating new concepts, properties and data sources, and can further be shared and reused across sectors with relatively low effort. When a new entity is introduced to the system, all components are automatically updated. Thereby, the framework is capable of serving the changing needs of city sectors, based on the issue at hand, while adapting to the dynamic nature of urban systems. This section provides a detailed description of the modules that compose the *OSMoSys* framework.

4.1 Ontology for Smart City Planning and Management

4.1.1 Domain and Scope Definition

In order for the ontology to constitute an encompassing framework for the domains of smart city planning and management, its taxonomy is organized around eleven principal classes (namely, *Case*, *Decision*, *Document*, *Event*, *Method*, *Place*, *Sampling*, *Situation*, *Technologies*, *TemporalEntity*, *WebEntity*) that represent broader and cross-sector domain concepts. The latter are further specialized, by including subclasses that enable the representation of the various urban systems, their respective data sources, the different agent types, the diverse city services, and the technologies that facilitate their interconnection. In this regard, the complete ontology comprises 121 classes, 82 object, data, and annotation properties, as well as 23 individuals and data types, representing a total of 226 entities (based on various standards, relevant external ontologies, and controlled vocabularies, as described in Sect. 4.1.2).

The developed ontology defines a common semantic model describing data from disparate sectors within a city in a machine-processable way, with a focus on planning and management. To this end, it constitutes the first approach in this domain. Its scope is to specifically facilitate the mapping of heterogeneous data models, already in use in cities, to its overarching semantic framework. Nevertheless, it does not intend to replace any of the various existing data models, but rather to enable the interoperable information exchange among them and across city agencies.

4.1.2 Reuse of Existing Semantic Models

The ontology reuses terms and concepts from various European and American standards and roadmaps on smart cities, as well as from relevant external ontologies and controlled vocabularies. The former namely refer to the Publicly Available Specifications (PAS) on Smart Cities [4] and the Smart City Concept Model [5], the Smart Cities Readiness Guide [17], and the Operational Implementation Plans of the European Innovation Partnership on Smart Cities and Communities [8].

Besides, terms from the following vocabularies are reused: *dc*: and *dct*: a collection of metadata terms maintained by the Dublin Core Metadata Initiative; *foaf*: a dictionary of named properties and classes, integrating social, information, and representational networks; *gml*: an XML-based grammar and encoding standard of the Open Geospatial Consortium (OGC) for describing geographical features; *schema*: a set of schemas for HTML markup and structured data interoperability on the web for search optimization; *skos*: a representation standard for knowledge organization systems; *vann*: a controlled vocabulary for semantic annotations. Classes and properties of the proposed semantic model are aligned with the following external ontologies: *DUL* (Dolce+DnS Ultralite upper level ontology); *CityGML* (unified model for 3D city representations); *dbpedia-owl*; *OTN* (Ontology of Transportation Networks); *SSN* (Semantic Sensor Network ontology); *OWL-Time* (ontology of temporal concepts). Finally, the ontology fully complies with the following data modeling vocabularies: *owl* (Web Ontology Language); *owl2xml*; *rdf* (Resource Description Framework); *rdfs* (RDF Schema); *xsd* (XML Schema).

4.1.3 Ontology Development and Taxonomy Definition

The OWL2 (Web Ontology Language)[2] specification was used for developing the ontology for smart cities, and it was further built with the Protégé 4.3.0 ontology development platform[3]. Mapping the various standards mentioned in the previous paragraph into the demonstrated ontology, gives stakeholders the opportunity to take advantage of its inference and reasoning potential. In addition, it facilitates the reusability and extensibility of the standards themselves, especially through the web-based components of the *OSMoSys* model (the Web Ontology Browser and the Knowledge Graph) that complement the described ontology.

Moreover, the variety of vocabularies and external data models incorporated in the presented ontology indicates its intention to cover the complexity and diversity of city-related information, ranging from spatial and geometric features (e.g. *gml*, *CityGML*, *OTN*) to sensor, social, and information networks (e.g. *SSN*, *foaf*), and time intervals (*OWL-Time*). Yet, it substantially extends them by introducing multiple new concepts and relations, so as to also address the aforementioned smart city roadmaps and standards. A representative sample of these extensions is further described in the following paragraphs (for clarity purposes, concepts of reused vocabularies or data models are indicated by the accompanying namespace prefix).

One example case is that of the *schema:Place* principal class. The latter is further specialized, including concepts such as *schema:AdministrativeArea*, *dbpedia-owl:City*, *Item*, *foaf:Agent*, *DUL:InformationObject*, *dct:PhysicalObject*, *Service*, and *UrbanSystem*. This taxonomy development approach allows concepts, referring to both physical and virtual features of the city, to inherit (*rdfs:subClassOf*) the spatial properties (geo-reference) of the top-level class *schema:Place*. Subsequently, the data mapped to these concepts, irrespective of the city sector they refer to, will immediately inherit these properties as well. This is highly essential for the contemporary planning procedure that not only engages with the physical elements of cities and urban systems but increasingly also with the digital information embedded in them. The *UrbanSystem* class, introduced by the presented

[2] http://www.w3.org/TR/owl2-profiles/

[3] http://protege.stanford.edu

ontology, enlists a hierarchy of subclasses referring to the diverse urban systems (e.g. *BuiltEnvironment*, *Energy*, *Environment*, *Healthcare*, *Transport*, *Waste* etc.), while the *Service* class models the corresponding services provided by the different sectors. The respective data produced across the various city systems (*UrbanSystem*) are modeled as instances of the *UrbanSystemsData* class, which is a sub-concept (*rdfs:subClassOf*) of the *DUL:InformationObject*, and can further be used (*hasRoleIn*) in decision-making (*Decision* class). Hence, through the previously described taxonomy of classes and relations (object properties), the ontology enables the interoperability among various city sectors, its services, and its corresponding data sources.

With respect to the increasing amount of sensor devices in contemporary urban environments, the ontology incorporates the *ssn:Sensor* concept (under *dct:PhysicalObject*) of the Semantic Sensor Network (SSN) Ontology [6], yet it specializes and extends it, so as to better describe the present variety of sensor resources. Thus, the *SensorResource* concept is introduced that incorporates both the *ssn:Sensor* and the *SensorSystem* classes (e.g. a weather station comprising multiple sensors in one system, is not sufficiently described by the *ssn:Sensor* concept). Sensor resources in the demonstrated ontology are part of (*dct:isPartOf*) a wider sensor network (*SensorNetwork*) or a sensor web (*SensorWeb* under the *WebEntity* principal class).

One of the most innovative aspects of the ontology concerns its ability to integrate entities referring to human-generated data in cities. Stemming mainly from smart devices and social media platforms, they constitute increasingly important sources of information, especially for the domains of urban analytics and decision-making. For this, the ontology introduces multiple concepts to enable their modeling. For instance, as a specification of the *ssn:Sensor* class, the *HumanSensor* entity is added. The latter is defined as a person who reacts to stimuli and/or events and occasionally makes observations about the physical world. Therefore, it is made equivalent (*owl:equivalentClass*) to *foaf:Person*, for the system to realize that instances belonging to this entity solely refer to humans, instead of devices (e.g. *ssn:Sensor* or *SensorSystem*). As such, it is also related to the broader concepts of *foaf:Agent*, *foaf:Group*, *DUL:Community*, and *foaf:Organization*. Besides the *SocialData* and *SocialWebData* classes are introduced, under *DUL:InformationObject*, to respectively model urban data from census records or demographics, and streams from social media platforms (e.g. Twitter, Instagram, Foursquare etc.), which differ from sensor streams

(*SensorData*). These concepts are further related – through OWL object properties – to urban systems for further integration in the planning process. The ontology also includes a *Crowdsourcing* class, to describe a contemporary form of urban sensing based on human input.

Time intervals and diversities in data speed are also taken into account, modeled through classes such as *time:TemporalEntity*, *time:Interval*, *time:Instant*, *StaticSampling*, *Semi-StaticStampling*, and *DynamicSampling*.

4.1.4 Ontology Matching

The alignment of the developed ontology with the multiple controlled vocabularies and external semantic models is achieved through the use of ontology matching techniques. The latter enable the discovery of mutual relations and correspondences among entities belonging to different ontologies. In this case, correspondences are established through equivalence relations (*owl:equivalentClass*), links between individuals (*owl:sameAs*), and structural equivalence indications between entity IRIs (Internationalized Resource Identifiers) with the use of IRI refactoring. All the previous matching measures use basic similarity, meaning that they make comparisons based on a particular characteristic of the aligned entities (e.g. name, *dct:description*, IRI string etc.) [9].

4.1.5 Inconsistency Check and Evaluation (Reasoning)

The developed ontology has fully been evaluated for potential anomalies and inconsistencies, by performing tests with multiple open-source reasoners, namely FaCT++[4], HermiT 1.3.8[5], and Pellet[6]. The inferred ontology class hierarchy proved to be fully consistent in all aforementioned tests.

---

[4] http://owl.man.ac.uk/factplusplus/

[5] http://hermit-reasoner.com

[6] http://clarkparsia.com/pellet

4.2 Web Ontology Browser

To provide city stakeholders with straightforward access to the above-described ontology, *OSMoSys* further incorporates a Web Ontology Browser (WOB). The latter offers a web-based user interface (UI) for navigating through the complete semantic model hierarchy, without requiring computational skills in ontology development environments or relevant software. Unlike the majority of ontologies and data models, which remain accessible to a limited group of experts, the WOB allows open access to any interested party. Thereby, it aims to provide a tool for bridging together the disparate urban data silos, thus facilitating the collaboration among different city agencies.

The WOB for smart city planning and management is fully indexed, containing the complete list of entities (i.e. classes, object and data properties, annotations, individuals, and data types) of the developed ontology. Besides, any classes or properties mapped to entities of related ontologies or controlled vocabularies, include links to the external IRIs, in accordance with Linked Data and Semantic Web principles (based on *owl:sameAs* and *rdfs:seeAlso* statements).

The UI layout of the WOB is organized in three panes, providing different browsing possibilities and views of the ontology (Fig. 1). The upper-left pane lists the different ontology entities in groups of classes, object, data, and annotation properties, individuals, and data types. It also contains an option for returning back to the general overview. When any of the previous entity groups is selected, a complete index of the corresponding concepts in alphabetical order appears on the lower-left pane of the UI. Finally, the main pane accommodates the full semantics, descriptions, and annotations of each selected entity, as well as its relations to other classes, and links to external aligned concepts. Definitions and other annotations, object properties and taxonomic hierarchy information, as well as description logics formalisms are clearly organized in different blocks within the main pane, so that they are easily distinguishable. The user can interactively browse the different entities and explore relations, either through the side-pane indexes or by directly clicking on any term included in the main pane. In addition, when a new concept is introduced to the system, the WOB is automatically updated, hence facilitating the data mapping process.

Fig. 1. User interface of the Web Ontology Browser

4.3 Web-based Knowledge Graph

To further facilitate the data sharing and concept discovery processes across heterogeneous sectors, the *OSMoSys* framework also introduces a novel web-based Knowledge Graph (KG), to allow visual and interactive exploration of the underlying domain ontology, as well as linked urban data mapped into it (Fig. 2). Alongside the WOB module, it further provides a dynamic, graph-based representation of the cross-sector entities and their relations. As previously discussed, each of the *OSMoSys* components can be used independently, owing to its modular architecture, yet the combination of all three exploits its full potential. In particular, the KG mainly focuses on visual exploration and discovery of concepts and interrelations, whereas the WOB primarily facilitates the data mapping and integration process. Formal description, semantic annotation and introduction of new city-related entities take place in the backend ontology module, which in turn supplies and informs the other two components of *OSMoSys*.

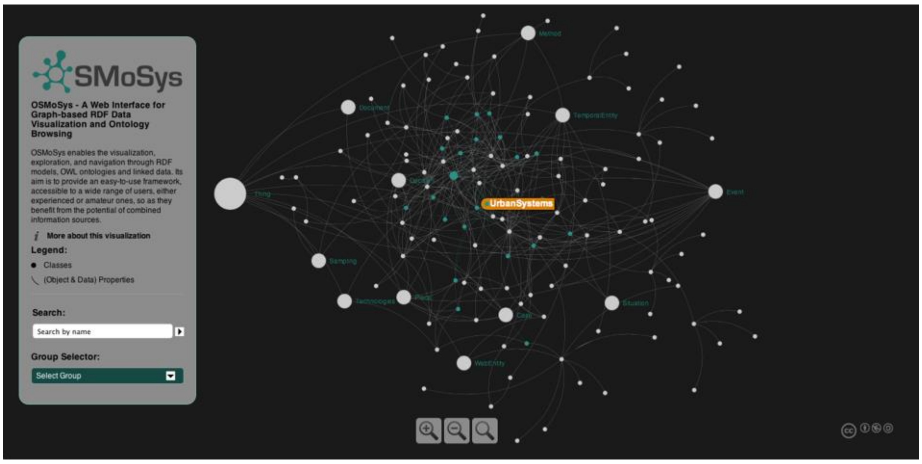

Fig. 2. User interface of the web-based Knowledge Graph. The dynamic graph layout displays ontology classes as nodes, and relations as edges. When hovering over a node, its label is highlighted. Different levels of detail are provided through semantic zooming

The KG receives as input urban data from diverse sources in RDF (Resource Description Framework) format, mapped into the above-described ontology, and displays them on a force-directed graph layout. An innovative aspect of the presented KG lies in its ability to visualize heterogeneous, yet semantically annotated, urban data as instances of ontology classes along with their relations in a single web-based platform. In this way, it aims to enable the interoperability among the various urban systems and, thereby, support planners and decision-makers in simultaneously correlating information from different sectors. Unlike the current manual and labor-intensive connections across data sources, the developed GUI provides stakeholders with easy navigation through the multiple domain concepts and allows them to effortlessly discover relations among entities and data.

The graph layout displays ontology classes as nodes, and relations (object, data, and annotation properties) as edges. Node size variations indicate whether an entity refers to a principal class or a subclass. The default top-level class, hence the largest node, is *owl:Thing*, since every individual, according to the OWL specification, is an instance of this particular class. In full display mode, besides *owl:Thing*, the user can easily recognize the eleven principal classes, around which the semantic model is organized (described in section 4.1). As a user zooms in, the graph displays different levels of detail, gradually making all subclass labels visible, in a process called semantic zooming. Hovering over a class node highlights its label along with its immediate relations to other classes, for visual clarity purposes. By further selecting a highlighted graph entity, an information pane pops up on the right, containing additional details about the class itself, as well as an index of all related entities. Moreover, the main graph display shows an isolated view of the selected concept and its connections. The user can easily return to the full network view, by choosing the corresponding option provided by the GUI. Navigation is also possible through the information pane index.

One of the key features of the KG is the search module. Thereby, the user is provided with the possibility to discover a concept, by performing keyword search, without necessarily knowing the exact class name in advance. All relevant results will appear as a list below the search field, on the left floating menu. In this way, the user can navigate through the various results and, correspondingly, highlight the selected concepts and their relations on the main display. Besides the search modules, users are also provided with a group selection tool. The latter is currently capable of grouping concepts of the semantic model, based on their function as top-level, principal, or sub-classes.

The web-based KG was created with the use of the Sigma.js JavaScript library[7]. The force-directed graph layout of the underlying ontology was built with Gephi[8], in combination with the Yifan Hu multilevel algorithm [12]. Pre-processing of the ontology graph was implemented using OntoGraf[9] and Graphviz[10]. RDF data are stored in and retrieved from a SparkleDB triplestore[11]. The KG is automatically updated, based on changes made in the underlying semantic model.

---

[7] http://sigmajs.org

[8] https://gephi.github.io

[9] http://protegewiki.stanford.edu/wiki/OntoGraf

[10] http://www.graphviz.org

[11] http://www.sparkledb.net

5 Use Case

OSMoSys has successfully been used as a back-end semantic enrichment and data integration mechanism in a recently developed web-based system[12] that supports the analysis, integration, and visualization of large-scale and heterogeneous urban data. An interesting case for urban decision-making refers to city-scale events, as they can have a strong impact on a variety of urban aspects, such as mobility flows and occupancy levels in specific public spaces in the city. Different stakeholders (e.g. event organizers, municipal authorities etc.) can benefit from real-time (or near real-time) urban analytics, in relation to the events, for making better-informed decisions. Yet, this is a challenging task, as it requires the simultaneous combination and evaluation of different data sources. In this context, *OSMoSys* enabled the semantic annotation and integration of heterogeneous data ingested in the aforementioned system, with application to the monitoring and assessment of a) the Amsterdam Light Festival (ALF) 2015, and b) Milano Design Week (MDW) 2014.

The duration of ALF was from November 27, 2014 to January 18, 2015. MDW concerned a shorter time period, namely from April 4, 2014 to April 14, 2014. Apart from multiple (semi-)static datasets (e.g. municipal demographic records, event maps with installations etc.), the system analyzed data streams from traffic sensors and social media platforms, namely Twitter and Instagram. For ALF, a total of 26.740.669 geo-referenced tweets and 15.959.566 Instagram posts were collected. In the MDW case, the system collected 1.879.187 geo-referenced tweets and 1.090.237 Instagram posts. These data streams referred to posts in the proximity of the installations or event area.

The ingested data sets and streams were mapped into the developed ontology as instances of classes to facilitate the processes of data exploration, analysis, querying, and visualization (Fig. 3). In both example cases, the collected data were mapped as instances of the class *dct:Event*, which connects them to a specific site (through the object property *atPlace*) and time period (*DUL:hasEventDate*). The property *DUL:hasParticipant* further links them with the event visitors (*foaf:Agent* and subsequently with *foaf:Person*).

---

[12] A live instance of the system is available online, with demonstrations for each of the described use cases, via the following link: http://bit.ly/1ACmp6s

Mapping to a specific city is achieved through *dbpedia-owl:City*, further linked with external GeoNames feature codes. To describe the nature of the ingested data (static, semi-static, real-time), the *Sampling* primary class is used and, respectively, its subclasses *osmosys:StaticSampling*, *osmosys:SemiStaticSampling*, and *osmosys:DynamicSampling*. Following the spatial and temporal annotations of the data inputs, the latter are further specified according to their source (*hasDataSource*: *Demographics*, *ssn:SensorInput*, *SocialWebData*). Classes and properties of the KR framework's ontology representing human-generated data sources were particularly important in describing the social media streams for both example cases. Especially, the classes *osmosys:SocialMediaPlatform* (containing instances for *Twitter* and *Instagram*), *osmosys:SocialWebData* (subclass of *DUL:InformationObject*), and *osmosys:HumanSensor* (subclass of *ssn:Sensor* and equivalent to *foaf:Person*). The latter assisted in clearly defining whether data streams referred to (*isProducedBy*) sensor devices (described by *ssn:SensingDevice*) or social media inputs produced by people (*HumanSensor)*. In this way, a stakeholder can perform queries about, for instance, the traffic flows on a specific day of the event, merging together measurements (*ssn:hasMeasurement*) produced by traffic sensors and geo-referenced tweets about the traffic conditions around the event site. In addition, data inputs were enriched with thematic annotations, with regard to the urban system they belong to. This is achieved via object properties, such as *refersTo*, *isAbout*, *isConstituentOf*, *isEmbeddedIn* and others, which relate data inputs to specific urban systems (e.g. *Transport*, *CityEvent* etc.).

Through the previously described mappings, RDF triples of the ingested data were generated, enriched with spatial, temporal, and thematic annotations. This allowed organizers and urban managers to perform specialized queries, based on the issue at hand. The web-based KG further enabled data exploration and visualization, facilitating urban analytics.

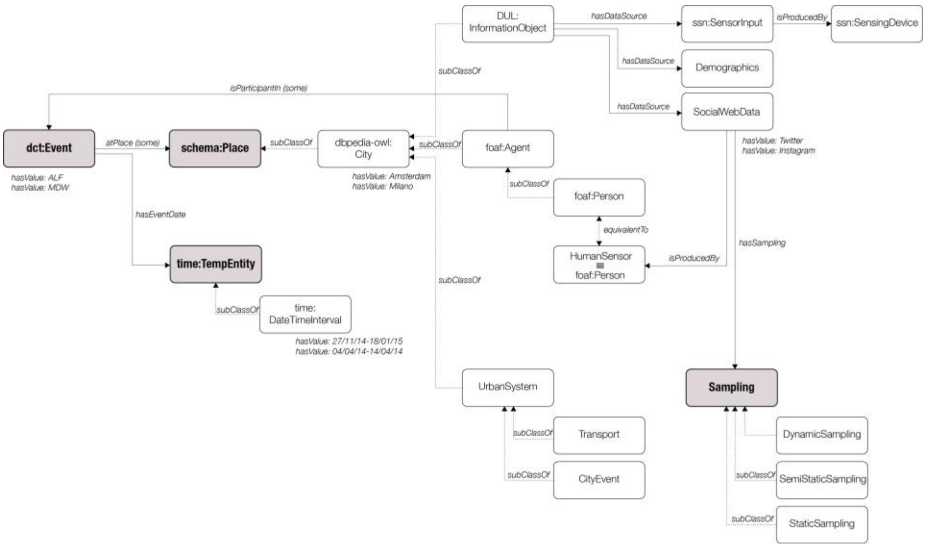

Fig. 3. Simplified view of the semantic data mapping into (parts of) the ontology hierarchy for the two city-scale events (ALF and MDW)

6 Conclusions and Future Work

This paper introduced *OSMoSys*, a novel knowledge representation framework for smart city planning and management that enables the semantic integration of heterogeneous urban data from diverse sources. It further showed how the framework's three main components cater for a modular system architecture that provides expansion and scalability potential, adapting to the changing character of urban systems. *OSMoSys* enables the mutual interaction among urban planners, city managers, decision-makers, and other city stakeholders by providing tools that are capable of operating across the disparate data silos.

In particular, the developed ontology tackles the semantic discrepancies among urban data of different formats and speeds, by annotating them with rich, machine-processable descriptions. The paper showed how data stemming from smart devices and social media platforms could also be mapped to concepts accommodated in the ontology hierarchy. In this way, these increasingly important sources of information can easily be integrated in urban analytics. Besides, the complementary web ontology browser and interactive knowledge graph give stakeholders the opportunity to navigate the ontology online and, further, perform data exploration and visualization.

Finally, the experimental implementation of *OSMoSys* in two real-world cases as a back-end for an urban analytics system, demonstrated its potential in semantically enriching and integrating historical (static and semi-static) and real-time urban data from heterogeneous sources.

As part of future work, the plan is to conduct additional tests on mapping open data (mainly municipal records and streams from mobile phones and social media) from different sectors and various cities to the framework, so as to further examine its validity and completeness. These tests will indicate the potential necessity for incorporating new concepts in the ontology, or new sub-ontologies. In addition, query support for linked open data requires further development. Covering a wider spectrum of urban concepts, the framework will constitute a valuable tool for urban planners and managers and help them exploit the increasing amount of available data sources.

Acknowledgements. This research is funded by the Greek State Scholarships Foundation I.K.Y. (by the resources of the Educational Program “Education and Lifelong Learning”, the European Social Fund (ESF), and the EU National Strategic Reference Framework (NSRF) of 2007-2013). It is further funded by a scholarship of the Alexander S. Onassis Foundation. It is also financially supported by the Foundation for Education and European Culture (IPEP) and the A. G. Leventis Foundation.

References

Antoniou G, van Harmelen F (2009) A Semantic Web Primer. The MIT Press: Cambridge, MA, USA

Bellini P, Nesi P, Rauch N (2014) Knowledge Base Construction Process for Smart-City Services. In: Bilof R (ed) 19th International Conference on Engineering of Complex Computer Systems (ICECCS 2014). IEEE, Los Alamitos, CA, USA, pp 186–189

Bizer C, Heath T, Berners-Lee T (2009) Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS) 5(3):1–22

British Standards Institution (BSI) (2014a) Smart Cities – Vocabulary. BSI Standards, London

British Standards Institution (BSI) (2014b) Smart City Concept Model – Guide to establishing a model for data interoperability. BSI Standards, London

Compton M, et al (2012) The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. Web Semantics: Science, Services and Agents on the World Wide Web 17:25–32

Domingue J, Fensel D, Hendler JA (eds) (2011) Handbook of Semantic Web Technologies. Springer, Berlin Heidelberg

European Innovation Partnership on Smart Cities and Communities (EIP-SCC) (2014) Operational Implementation Plan. EIP-SCC, Brussels

Euzenat J, Shvaiko P (2013) Ontology Matching (2nd ed). Springer, Berlin Heidelberg

Falquet G, Métral C, Teller J, Tweed C (eds) (2011) Ontologies in Urban Development Projects. Springer London, London

Gröger G, Kolbe TH, Nagel C, Häfele KH (eds) OGC City Geography Markup Language (CityGML) Encoding Standard. Document Reference Number OGC 12–019, version 2.0.0

Hu Y (2005) Efficient, High-Quality Force-Directed Graph Drawing. Mathematica Journal 10(1): 37–71

Métral C, Falquet G, Cutting-Decelle AF (2009) Towards Semantically Enriched 3D City Models: An Ontology-based Approach. In: Kolbe TH, Zhang H, Zlatanova S (eds) GeoWeb Conference Academic Track–Cityscapes. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives) vol XXXVIII-3-4/C3 Vancouver, Canada, pp 40–45

Montenegro N, Gomes JC, Urbano P, Duarte JP (2012) A Land Use Planning Ontology: LBCS. Future Internet 4(4):65–82. doi: 10.3390/fi4010065

Poveda-Villalón M, García-Castro R, Gómez-Pérez A (2014) Building an Ontology Catalogue for Smart Cities. In: Mahdavi A, Martens B, Scherer R (eds) eWork and eBusiness in Architecture, Engineering and Construction (ECPPM 2014). CRC Press, Vienna, pp 829–836

Smart Cities Council (2014) Smart Cities Readiness Guides: The planning manual for building tomorrow's cities today. Smart Cities Council, Redmond, WA, USA