A Platform for Urban Analytics and Semantic Data Integration in City Planning

Achilleas Psyllidis(✉), Alessandro Bozzon, Stefano Bocconi, and Christiaan Titos Bolivar

Delft University of Technology, Delft, The Netherlands
{A.Psyllidis,A.Bozzon,S.Bocconi,C.TitosBolivar}@tudelft.nl

Abstract. This paper presents a novel web-based platform that supports the analysis, integration, and visualization of large-scale and heterogeneous urban data, with application to city planning and decision-making. Motivated by the non-scalable character of conventional urban analytics methods, as well as by the interoperability challenges present in contemporary data silos, the illustrated system – coined SocialGlass – leverages the combined potential of diverse urban data sources. These include sensor and social media streams (Twitter, Instagram, Foursquare), publicly available municipal records, and resources from knowledge repositories. Through data science, semantic integration, and crowdsourcing techniques the platform enables the mapping of demographic information, human movement patterns, place popularity, traffic conditions, as well as citizens' and visitors' opinions and preferences about specific venues in a city. The paper further demonstrates an implemented prototype of the platform and its deployment in real-world use cases for monitoring, analyzing, and assessing city-scale events.

Keywords: Urban analytics · Semantic integration · Crowdsourcing · Ontologies · SocialGlass · Urban computing · Smart cities

1 Introduction

The growing urban populations pose broad challenges to contemporary cities that span across domains as diverse as transportation, energy, and environment while also affect everyday human activities. Such challenges impact the quality of life in urban environments [1]. The rising complexity of the aforementioned domains is nowadays coupled with an increasing amount of data that reflect each urban system's performance as well as pertinent activities. Performance-related data mostly stem from distributed sensor resources across the city, which monitor each system's operations in real time. In addition, human-generated data produced through smart devices and social media platforms constitute an increasingly important source of information, especially with regard to human activities in cities. Yet, these data sources are currently insufficiently leveraged or used by planning authorities [2, 3].

In particular, city-related organizations create and operate on datasets based on each sector's specific purposes and problems at hand. This results into disparate data silos

G. Celani et al. (Eds.): CAAD Futures 2015, CCIS 527, pp. 21–36, 2015.
DOI: 10.1007/978-3-662-47386-3_2

linked to a single urban system, that are further characterized by diverse data models and schemas for storage purposes. Any correlation of information among different sectors is currently performed in a manual fashion, hence requiring great amounts of time and effort. Constrained by the interoperability barrier across the urban data silos, city planners accordingly tend to approach each urban system individually, while also employ a limited amount of the available data sources in their planning approaches [4]. Thereby, the emerging challenge lies in providing new methods and tools that support and extend the current urban planning processes, by further allowing for interoperable data sharing and reuse from diverse sources.

To address this challenge, the paper introduces SocialGlass, a novel web-based platform that supports the analysis, integration, and interactive visualization of large-scale and heterogeneous urban data, with application to urban planning and decision-making. SocialGlass systematically combines publicly available municipal records together with social media streams (Twitter, Instagram, and Foursquare in particular) and resources from knowledge repositories. To achieve this, it employs data science methods (e.g. anomaly detection analysis), semantic integration techniques (e.g. ontology-based data models), and crowdsourcing (e.g. sentiment analysis, reality sensing and verification). The presented framework pays special attention to the analysis and integration of content retrieved from social media. To further allow for the integration of heterogeneous urban data and for the extraction of semantically enriched content, an innovative domain ontology has been developed. Using well-studied techniques in knowledge representation [5, 6], it defines named relationships that link both the different urban systems together and the data generated within them, in a machine-processable way. Yet, as the system makes use of extensive human-generated data, crowdsourcing and reality verification are employed as part of the semantic enrichment process to tackle the relative bias, inherent to such pieces of information.

The aim of the presented platform is to provide a system and tools for efficient interpretation and understanding of urban dynamics, by embracing and harnessing the combined potential of heterogeneous city data sources. The contributions of this work are mainly three.

1. A novel web-based system for urban planners and other city stakeholders that blends state-of-the-art methods and tools for sensor and social media content analysis, user modeling, semantic data integration, crowdsourcing, and visual data exploration (see Sect. 4).
2. An example implementation of Semantic Web technologies, aimed at rich data description and integration. Owing to an innovative ontology for planning and management, the system enables mapping, and semantic enrichment of urban data from diverse city sectors (e.g. transport, energy, waste, water etc.), in a machine-processable way (see Sect. 4.2).
3. An empirical validation of the approach and the system. The paper describes how a fully functioning prototype[1] of the web-based platform has been used to support

[1] A live instance of which is available at the following link: http://social-glass.org with demonstrations for each use case mentioned in Sect. 5.

several real-world case studies. The latter exemplify the framework's conceptual properties and its potential value as a solution to urban analytics and management, with application to city-scale event monitoring and assessment (see Sect. 5).

The rest of the paper is structured as follows: In Sect. 2 the state-of-the-art and background context, as well as the related work are described. Section 3 presents the proposed method. Section 4 illustrates the system architecture. In Sect. 5 the paper presents the deployment of the web-based software system and its application to real-world case studies. Ultimately, Sect. 6 summarizes the conclusions and discusses future lines of research.

2 Background and Related Work

2.1 Urban Computing

The work illustrated in this paper adopts concepts and methodologies from the emerging field of urban computing. Fueled by the increasing proliferation of sensor and actuator mechanisms embedded in diverse systems across cities, as well as by the growing amount of data generated in these systems, urban computing aims to enhance and extend the current approaches to studying the dynamics of cities [1, 7, 8]. It specifically acts at the intersection of cities and urban data. Often also referred to as urban informatics [9], this interdisciplinary field exploits the potential of diverse data sources ranging from sensor networks to mobile devices and social media platforms. To address the contemporary challenges of complex urban environments, urban computing employs data science methods, information retrieval techniques, data integration, analysis, and visualization technologies. The latter are merged with tools and methods utilized in urban planning and social science with regard to urban contexts. To this end, the proposed framework aims to contribute to the field of urban computing, by providing a system and tools for better investigating and understanding the dynamics of cities through heterogeneous social urban data.

2.2 Related Work

Despite the fact that the field of urban computing is presently at a nascent stage, there is already a plethora of projects with purposes relevant to those presented in this paper. The projects described in the following paragraphs are related to this work either in terms of the methods and tools they use or in terms of the goals they set and, therefore, serve as appropriate cases for comparison.

Based on the rapid proliferation of Location-Based Social Networks (LBSNs) the Livehoods Project by Cranshaw et al. [10] introduces a clustering model in order to study the social dynamics of a city. Their approach is grounded on spectral clustering algorithms applied to a large-scale dataset of Foursquare check-ins, in an attempt to overcome the limitations of the traditional methods used in urban analyses, such as surveys and field observations. Based on people's movement and activity patterns revealed by the data stream, the authors achieved to map distinct neighborhoods – coined livehoods – within the wider urban fabric of the studied city. In most cases, the extracted livehoods showed a different organization compared to the one defined by municipal boundaries. Even though this project makes use of an interesting approach and method (i.e. spectral clustering algorithms) to studying urban dynamics, it is solely based on a data stream from one single source (i.e. Foursquare). Thereby, it is confined by the inherent limitations of this particular source (e.g. demographic representativeness, cultural bias, data veracity etc.), which is also a shortcoming of traditional urban analysis methods.

Similarly, Noulas et al. [11] model geographic areas and human activity within cities, by means of venue categories and based on a large-scale Foursquare dataset derived from geo-referenced Twitter messages. A spectral clustering algorithm is again being used for identifying citizen profiles and different activity areas in the city. As is the case with the Livehoods project, Noulas et al. considered a single data source to achieve the project goals, thus falling short on similar limitations. Further, the paper did not investigate user modeling in depth, meaning that it did not explicitly describe whether a particular activity is related to a resident, a local visitor or a foreign tourist. Besides, the work of Noulas et al. limits semantic annotations to the venue categories and social activities the geo-coordinates refer to. Similar shortcomings to the two aforementioned projects are observed in the work of Del Bimbo et al. and their LiveCities application [12], as well as in the CityBeat project by Xia et al. [13].

The projects described above demonstrate partial similarities to the methods used in our work. The purposes, however, are substantially different, as the web-based platform illustrated here aims to facilitate and enhance current urban planning and decision-making processes. In this regard, the LIVE Singapore! project by Kloeckl et al. [14] closely relates to this paper's proposal. It introduces an open platform for the collection, combination, and distribution of large-scale real-time urban data streams. An interesting feature of the aforementioned work lies in the correlation and integration of heterogeneous data sources for richer insights about city dynamics. Yet, at the stage presented in the publication, the authors mainly utilized data streams from sensors and call detail records (CDR). Human-generated data from social media networks were not investigated. In this paper, a novel and comprehensive framework for better understanding cities through heterogeneous social data is demonstrated. In addition, the developed web-based platform provides new methods and tools to create, integrate, analyze, and valorize urban data sources for the benefit of urban planning and decision-making.

3 Method

The demonstrated urban platform enables the (simultaneous) mapping of demographic information, human movement patterns, place popularity, traffic conditions, as well as citizens' and visitors' opinions and preferences with regard to specific venues in the city. To this extent, it leverages on content retrieved from social media (Twitter, Instagram, and Foursquare), mobile phone data, and spatial statistics and demographics retrieved from open municipal records. SocialGlass is also capable of incorporating dynamic sensor streams, either from already existing sensor resources or newly distributed sensor networks (DSNs). The implementation of these features relies on various methods and techniques applied to different modules, clustered in larger components, that respectively cater for data ingestion and analysis, semantic integration, as well as exploration and visualization (Fig. 1).

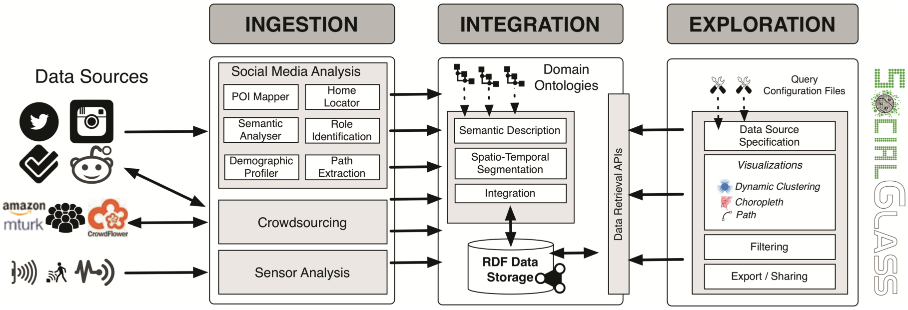

Fig. 1. Pipeline and system architecture of SocialGlass

The data ingestion and analysis component refers to the acquisition, cleansing, and processing of social and sensor data. In the proposed system, the active involvement of citizens is also possible in this process through a crowdsourcing sub-module that will further be described in the following section. To this end, data science methods, including information retrieval and social media analysis techniques, are employed for tackling issues pertinent to anomaly detection and data veracity. In addition, human computation techniques play a highly beneficial role when it comes to the efficient interpretation of personally, contextually, or culturally biased analyses of raw data [15]. This is particularly crucial for social web data, as their content is often unstructured in terms of machine readability and processing.

The semantic enrichment and integration component simplifies the interconnections amongst the different data providers. It specifically caters for interoperability issues across usage domains and their respective data sources, in an attempt to overcome the barriers set by the multiple data silos across city agencies. To achieve this, a knowledge representation model has been developed, based on a novel ontology that formally represents urban systems, the relations among them, and the corresponding data sources. The concepts of the aforementioned ontology are aligned with classes and properties from multiple external ontologies and controlled vocabularies [16]. This is for ensuring the interrelation among city sectors and further for facilitating the data integration process.

Finally, SocialGlass offers a web-based user interface for data exploration, interactive information visualization, comparison, and urban analytics. Through these components the different city stakeholders can gain insights about spatial and temporal parameters of the studied urban context. Data exploration takes place in a map-based web environment, on top of which city-related information is visualized in a layered fashion. The visualization types in use vary from dynamic point clusters to intensity heat maps, activity-related paths, and choropleths in combination with auxiliary spatio-temporal graphs.

4 System Architecture

To address the several scalability challenges posed by the diversity, quantity, and speed of social and sensor data sources, the platform adopts a modular architecture, optimized for loose coupling and scalability. Each module of the system focuses on and performs a single functionality. Thereby, a module can independently and redundantly be deployed, so as to enable dynamic allocation and distribution of computational and storage resources. The communication among the modules is achieved by means of message queues. The logic is that a module can post messages without knowing which module will consume them at the other end of the pipeline. As a result, the platform's back-end is capable of accommodating new data sources or analysis components with relatively low effort. This modular and loosely coupled architecture is applied to all system components described in the previous section, namely data ingestion and analysis, semantic enrichment and integration, and exploration and visualization. This section provides a detailed description of these components and their respective modules.

4.1 Data Ingestion and Analysis Component

The data ingestion and analysis component is further subdivided into three sub-systems, respectively devoted to social media analysis, crowdsourcing, and sensor analysis. Their functionalities are described in the following sections.

4.1.1 Social Media Analysis Sub-system

The Social Media Analysis (SMA) sub-system is devoted to the generation of information about a given urban environment, based on content created by users of popular social media platforms, such as Twitter, Instagram, and Foursquare. These platforms are particularly selected for their open-source APIs (Application Programming Interfaces). The modules constituting the SMA sub-system provide insights about citizens' activities, points of interest, as well as their spatial and temporal relationships. Each of these modules is briefly described below.

Point-of-Interest Mapper. This module maps geo-referenced micro-posts to urban Points of Interest (POIs). In the current implementation, this is achieved through the "check-in intent" Foursquare API[2], which selects the most popular venues amongst the ones that are in the proximity of a specific user's post. Because of the POI mapper, it is possible to enable the spatiotemporal analysis of an urban area according to a place's functionality (e.g. museum, educational institute, restaurant etc.). In the current implementation, POIs can also be specified independently. This is particularly interesting in the case of city-scale events that define their own POIs (such as in the examples described later in Sect. 5.2).

[2] https://developer.foursquare.com/docs/venues/search.

Semantic Analyzer. The content of the aforementioned micro-posts is processed through the semantic analyzer module, in order to extract entities (e.g. the name of a location) from its textual structure. These are further connected to entities in DBPedia[3]. In this way, it is possible to associate social media users, urban areas, and temporal intervals with the entities mentioned in the related posts, so as to create topics profiles. Entity extraction relies on DBPedia Spotlight, which features annotation functionalities in several languages.

Demographic Profiler. Given social media users observed in an urban area of interest, this module estimates their gender and age. This is achieved by means of a state-of-the-art face detection and analysis component (Face ++)[4], in combination with a dictionary-based gender recognition module (Genderize)[5].

Home Locator. This module estimates the home location of social media users owing to a recursive grid search, based on the geo-location of their micro-posts. The search finds the actual place where the user posts more often and uses that location as an approximation of the user's home location. Once identifying the coordinates of the estimated home location, the module makes use of a reverse geocoding service (Geonames)[6] to identify the user's city and country of origin.

User-Role Identifier. Particularly functional to urban analytics is the characterization of the types of people in a city. This module classifies people according to their home city – based on input from the home locator – in relation to the analyzed urban context. The system classifies people into residents (home location and analyzed urban context are the same), commuters (home location and analyzed urban context are different, but belong to the same country), and foreign tourists (home location and analyzed urban context are different and belong to different countries).

Path Extractor. This module constructs paths by examining all geo-referenced posts of a person in a fixed time period. Paths are formed through the coordinate concatenation of subsequent posts. Each point in the path is mapped to a known venue, based on input from the POI mapper module.

4.1.2 The Crowdsourcing Sub-system

To cater for issues of data sparseness, veracity, and sense making, the platform incorporates a crowdsourcing sub-system. This further allows for pro-active interaction with individuals and groups (drawn from social media or human computation platforms) for data creation, cleansing, and interpretation purposes. The crowdsourcing system can operate in two modes: (a) a Social Sensing mode, which has the ability to contact social media users in order to request services, such as on-demand data creation, cleansing,

[3] http://dbpedia.org/.
[4] http://www.faceplusplus.com/.
[5] http://genderize.io/.
[6] http://www.geonames.org/.

and linkage; and (b) a Human Computation mode to engage with anonymous crowd from human computation platforms, such as CrowdFlower and Amazon Mechanical Turk[7]. Examples of operations enabled by the crowdsourcing system may include: on-demand sensing of urban environment phenomena (e.g. rainfall, temperature variation etc.); disambiguation of textual and visual content with respect to sentiments; and verification of relatedness and appropriateness of images in relation to an observed event or a targeted group of people.

4.1.3 The Sensor Analysis Sub-system

Besides crowdsourcing, which relies on the participation of citizens who act as human sensors, SocialGlass also accommodates modules for analyzing data streams stemming from sensor devices distributed across the urban fabric. The latter may refer to already existing sensor resources and/or newly embedded DSNs. The combination of both human-generated data and information from sensor resources can cater for data sparseness and reality verification issues. Several experiments have already been conducted in this regard, while currently working on their integration in the system.

4.2 Semantic Enrichment and Integration Component

When combining heterogeneous data from complex urban environments it is common to encounter syntactic and semantic discrepancies, mainly due to spatial, temporal, and/or thematic diversities in the studied datasets. The use of different data models and schemas in the data repositories across city agencies further strengthens the aforementioned diversities. Therefore, discordant, incompatible with each other, and cumbersome data silos are created. The goal of the semantic enrichment and integration component is to specifically tackle this issue. By structuring and encoding municipal records, sensor and social media inputs from the ingestion component, it provides a coherent representation framework that simplifies their usage and unlocks their combined value. This is of particular importance to the city planning process, which can largely benefit from the simultaneous combination of heterogeneous urban information.

To achieve this goal, a novel ontology[8] (OSMoSys – OntoPolis Semantic Modeling System) was developed that describes the different urban systems (e.g. energy, waste, water, transport etc.), the respective data sources (e.g. spatial statistics, sensor data, social data etc.), the city technology enablers (e.g. interoperability types, connectivity, computational resources), and defines the relations among them. It also formally represents the different urban sectors/agencies and includes broader concepts that allow data mapping from different departments. 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 [17] and Smart City Concept Model [18], the Smart Cities Readiness Guide [19], and the Operational

[7] Respectively, http://www.crowdflower.com and https://www.mturk.com/.
[8] A running instance of OSMoSys is available at: http://www.hyperbody.nl/demo/osmosys.

Implementation Plans of the European Innovation Partnership on Smart Cities and Communities [20]. In addition, it reuses terms from the following vocabularies: *dc*; *dct*; *foaf*; *gml*; *schema*; *skos*; *vann*, and external ontologies: *DUL*; *CityGML*; *dbpedia-owl*; *OTN* (Ontology of Transportation Networks); *SSN* (Semantic Sensor Network ontology); *OWL-Time*.

To further provide city stakeholders with straightforward access to the above-described ontology, a web ontology browser (WOB) was developed [21]. The latter offers a web-based user interface (UI) for navigating through the complete taxonomic hierarchy of classes and properties (relations), without requiring computational skills in ontology development environments or relevant software. The UI layout of the WOB is organized in three panes, providing different navigation possibilities and views of the ontology (Fig. 2). 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. Further, 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 vocabularies. 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.

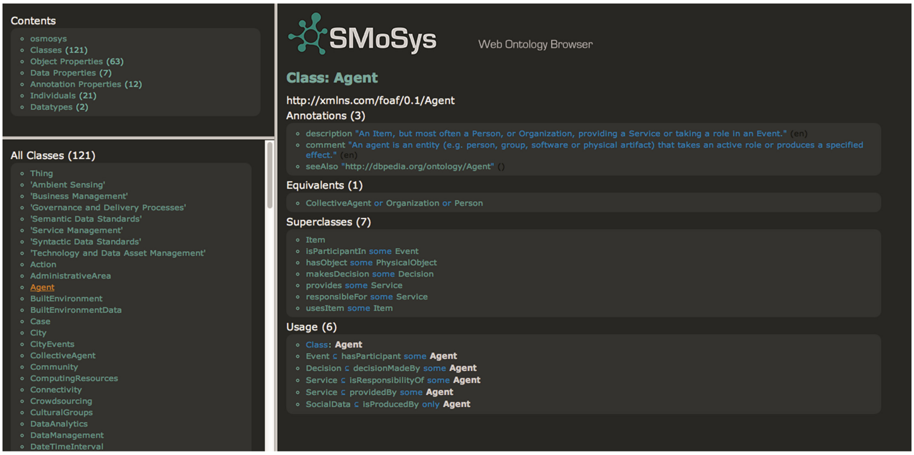

Fig. 2. Instance of the OSMoSys Web Ontology Browser user interface

The ontology was developed in OWL 2 (Web Ontology Language) and built with the Protégé 4.3.0[9] ontology development platform. The WOB is largely based on OWLDoc, as well as JavaScript and HTML5.

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

4.3 Data Exploration and Visualization Component

The data exploration and visualization component uses the APIs of the semantic enrichment and integration cluster to provide interfaces for urban analytics. Interactive data exploration tools allow users to ponder on the collected and linked data and to extract insights about the spatiotemporal analysis of a specific urban environment. To facilitate exploration, the platform's user interface enables the creation of an arbitrary number of layers that can be initiated, deactivated, and organized in a customizable order. Each layer visualizes a partition of the retrieved information and is characterized by (a) the data source, (b) the visualization type, and (c) the filtering method. Owing to the semantic enrichment and integration component, it is possible to specify custom queries, spanning across multiple data sources. In relation to the visualization type, the platform's current implementation offers the following set of map-based data visualizations:

- Dynamic Point Clusters;
- Choropleth maps;
- Activity Paths;
- Data graphs depicting the people's roles, age, gender, popularity of POIs and venue categories, temporal distribution of micro-posts, semantics, and spatial statistics.

Once retrieved, data can be filtered according to the time span, the POI category, the activity intensity, and person roles. Finally, the platform's users are also provided with functionalities for exporting a custom analysis result in various formats (CSV, JSON, RDF) and save each exploration status for sharing or further reuse purposes.

5 Empirical Validation

This section indicates how the above-described components and modules were implemented in a working prototype of the SocialGlass platform. It, further, provides real-world examples, in which the system was empirically tested.

5.1 Platform Deployment

The implementation of SocialGlass consists of a fully working web-based user interface (UI), grounded on the previously described architectural components. Figures 3, 4, 5 and 6 showcase some representative instances of data visualizations for urban analytics purposes in different cities. The illustrated examples demonstrate the platform's ability to effectively combine content retrieved from social media platforms (Twitter, Instagram, Foursquare) with mobile phone data, and municipal records about spatial statistics and demographics, within a specific urban context. Each stakeholder (e.g. urban planner) is enabled to make customized combinations, so as to approach and gain insights about a specific urban situation from different angles.

Through a pop-up menu the user first specifies the data source s/he is interested in (Fig. 3). The available options include social media data (Twitter and Instagram) and statistical records. Custom queries fusing multiple data sources together are also made possible through the semantic enrichment and integration component. As the platform

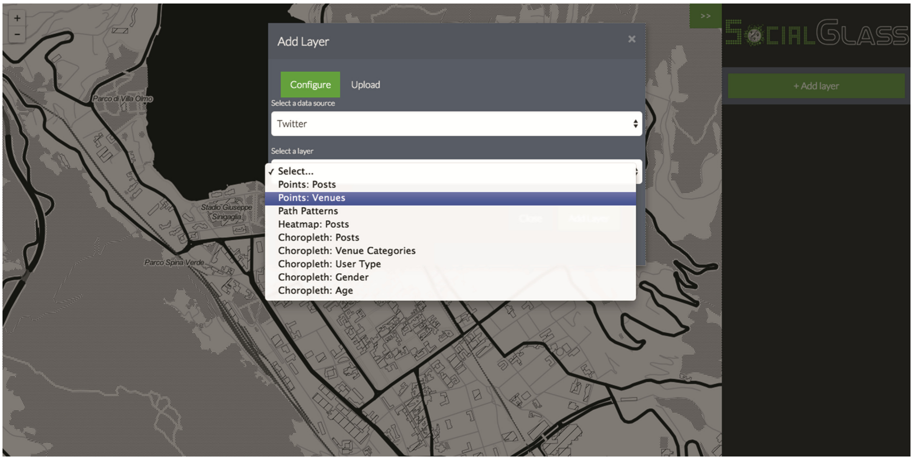

Fig. 3. Dropdown menu with various data visualization options

Fig. 4. Dynamic point clusters of Twitter micro-posts, in combination with path patterns of foreign tourists and their temporal variations in the city of Rotterdam

follows a layer-based visualization approach, users are further provided with various choices for visualizing and analyzing the selected data source, through a dropdown menu. The offered choices are listed below.

Dynamic Point Clusters. Available in the social media data sources, this visualization type displays geo-located objects (e.g. Twitter micro posts), POIs, and sensor records. Objects are dynamically clustered according to their spatial proximity to a certain location. Zooming modifies the cluster granularity, following the map's different Levels of Detail (LODs), thus giving an overview on the number of objects in a specific area (Fig. 4).

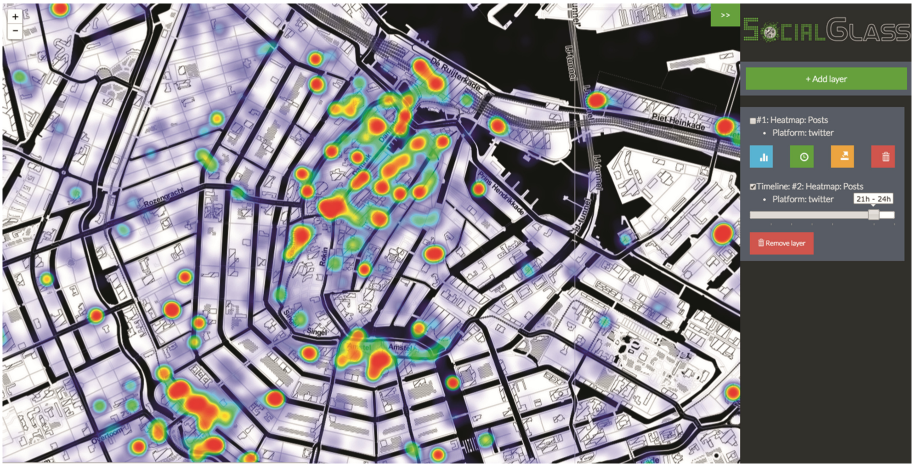

Fig. 5. Heat map of activity intensity during the Amsterdam Light Festival from 9 pm till 12 am

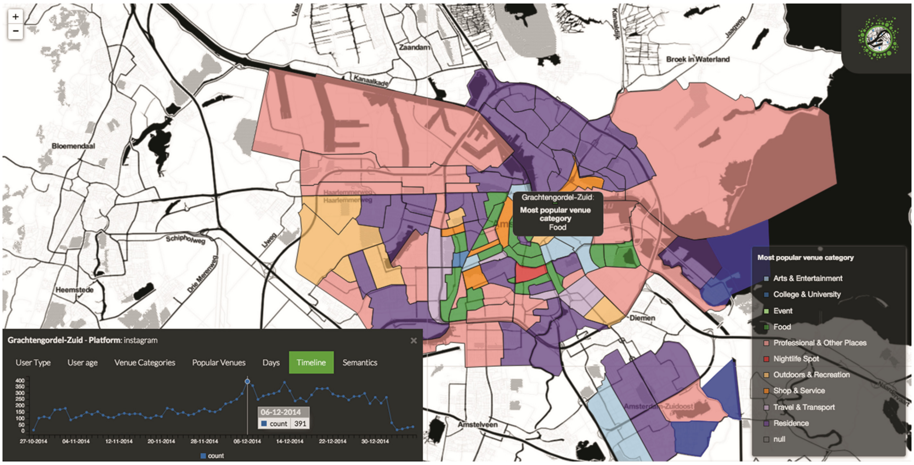

Fig. 6. Choropleth map visualizing the most popular venue categories and the corresponding activity intensity in each administrative district of Amsterdam, based on Instagram posts

Heat Maps. These graphical data representations of social media streams illustrate the frequency of objects, such as sensor measurements and/or micro posts, in a gradient, color-coded fashion. A timeline slider allows users to interactively analyze the real- or recent-time data streams in different time intervals (Fig. 5).

Choropleth Maps. This visualization option builds upon the urban space partitions of the underlying data source. For each partition it can show: (a) a color, representing a specific category (e.g. venue category, user roles, gender or age distribution, crime rates etc.); or (b) a color shade, in proportion to the measured variable. In other words, it depicts in real or recent time the social activity intensity in different administrative districts of a city (Fig. 6).

Path Patterns. A path pattern utilizes 3D arc projections connecting locations, typically POIs, which denote the paths taken by different people who traverse the city. The thickness and color of the arc represent the popularity of a specific path among the considered groups of people (Fig. 4).

By clicking on an area (choropleth maps – Fig. 6), point or a path (Fig. 4), a pop-up window appears in the lower left side of the UI. This further contains additional information about the selected area or path, such as: (a) the distribution of people according to their role (inhabitants, commuters, and foreign tourists), age, and gender; (b) the popularity of POIs and venue categories; (c) temporal distribution of micro posts over the days of the week; (d) semantic profiles of the people posting from the selected area; and (e) other static data from publicly available data sources (e.g. crime rate, income distribution etc.).

5.2 Real-World Use Cases

SocialGlass is designed for facilitating urban planners and city managers in understanding and making decisions about different urban phenomena. In this regard, city-scale events dynamically change ordinary activities in cities, by having a particular impact on a variety of urban aspects, such as mobility flows and occupancy levels in specific public spaces. This makes them interesting cases for empirically testing and validating the platform, in providing stakeholders with real-time (or near real-time) urban analytics pertinent to the events. In this context, SocialGlass has successfully been deployed for monitoring and assessing three city-scale events, in different cities and for diverse time periods. These namely refer to the (a) Milano Design Week 2014, (b) Amsterdam Light Festival 2015, and (c) Como Summer Holiday Season 2014.

Geographically limited to the city of Milan, the *Milano Design Week* (MDW) 2014 featured multiple events attended by half a million visitors in half a thousand venues that serve as temporary exhibition centers. The MDW case covers the largest urban area of the three presented examples, yet the deployment period is the shortest, as it refers to a ten-day monitoring in total, specifically from April 4, 2014 to April 14, 2014. The platform has been used for a twofold purpose: on one hand, to distill demographics and user profiles of attendees, so as to enable a venue recommendation system; on the other hand, to activate social media users for crowd sensing and data cleansing purposes.

The second deployment of the platform was for the purposes of the *Amsterdam Light Festival* (ALF) 2015. The latter refers to an art-related city-scale event, with a duration period from November 27, 2014 to January 18, 2015. The festival's organizers were interested in assessing the success of the event, while pulling demographic and mobility profiles of attendees to inform their planning decisions for the upcoming edition.

In both of the above-mentioned cases, instances of the platform were deployed focusing on the place and time period that each of the events took place. Social media data from Twitter (a total of ~15 K unique users, and 28,619,856 tweets) and Instagram (a total of ~80 K users, and 17,049,803 posts), pertinent to the events, were collected (before, during, and after their end) and were further integrated with municipal records and statistics gathered from the events' organizers. The system harvested information about the popularity of the festivals and the corresponding opinions of visitors. In this regard, special attention was paid to sentiment analysis (opinion miner module) and human computation components (image description). To further validate the results, the platform's urban analytics were compared with the observations collected from the events' official applications.

Finally, the *Como Summer Holiday Season* 2014 is a summer-long event organized by the municipality of Como in Italy. In this regard, it covers the longest deployment period of the three cases, but the surveyed urban area is the smallest in terms of scale. The goals of this use case were (a) to monitor for a long period of time the urban fabric of a mid-sized city, and (b) to validate the figures obtained from social media data against data produced by the main Italian mobile telecommunications provider. The work included the analysis and linking of more than 8,000,000 phone calls (incoming or outgoing), 6,000,000 of SMSs, ~42,000 tweets, ~4,500 twitter users and more than 22,000 unique hashtags. While exemplifying the potential of an approach based on heterogeneous data, the *Como Summer Holiday Season* served as a relevant testbed for the integration and analytical nature of the initiative.

6 Conclusions and Future Work

This paper introduced SocialGlass, a novel web-based system that supports the analysis, modeling, semantic integration, crowdsourcing, and interactive visual exploration of large-scale and heterogeneous urban data. The presented platform enhances the processes of urban planning and decision-making, by providing state-of-the-art tools and methods to better understand the dynamics of cities from various perspectives. The analysis and planning approach described in this paper is essentially data-driven. To that end, the paper showed how to efficiently incorporate dynamic data streams (real- or recent time) from sensors and social media platforms, in combination with (semi-)static datasets from the municipality. Contrary to conventional urban analytics, which are solely based on slowly updated and non-scalable census records and demographics, the presented platform provided a holistic framework for addressing diverse aspects of urban environments, by deploying a fully scalable city analytics framework.

The system's modular architecture allows new components and data sources to easily be incorporated in the back-end structure. Further, the developed ontology tackles the syntactic and semantic discrepancies among urban data of different nature and, thus, offers a new approach in bridging the interoperability barriers set by the existing data silos. Besides, the crowdsourcing components give citizens the opportunity to actively contribute to the system, while the map-based UI and its layered data visualization possibilities allow stakeholders to simultaneously explore various urban variables, according to their needs. Finally, the deployment of the platform in real-world cases showed its potential value as a solution for city-scale event monitoring and assessment, as well as for helping stakeholders in dynamically rearranging their plans and decisions.

As part of future work, the plan is to further develop the sensor analysis modules to better accommodate data streams from sensor devices across the urban fabric. In addition, mechanisms for deriving more detailed and accurate views of activities within cities from social media, as well as more advanced data search components are currently in progress. In this regard, the system seeks to break new ground in the emerging field of urban computing and, hopefully, aims at better support for planning.

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

1. Zheng, Y., Carpa, L., Wolfson, O., Yang, H.: Urban computing: concepts, methodologies and applications. ACM Trans. Intell. Syst. Technol. (ACM TIST) **5**(3), 1–55 (2014)
2. Ciuccarelli, P., Lupi, G., Simeone, L.: Visualizing the Data City: Social Media as a Source of Knowledge for Urban Planning and Management. Springer, Heidelberg (2014)
3. Vaccari, A., Liu, L., Biderman, A., Ratti, C., Pereira, F., Oliveirinha, J., Gerber, A.: A holistic framework for the study of urban traces and the profiling of urban processes and dynamics. In: 12th International IEEE Conference on Intelligent Transportation Systems, IEEE Press, New York, pp. 273–278 (2009)
4. Métral, C., Falquet, G., Cutting-Decelle, A.F.: Towards semantically enriched 3d city models: an ontology-based approach. In: Kolbe, T.H., Zhang, H., Zlatanova, S. (eds.) GeoWeb Conference Academic Track – Cityscapes, vol. XXXVIII-3-4/C3, pp. 40–45. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives), Vancouver, Canada (2009)
5. Domingue, J., Fensel, D., Hendler, J.A. (eds.): Handbook of Semantic Web Technologies. Springer, Heidelberg (2011)
6. Euzenat, J., Shvaiko, P.: Ontology Matching, 2nd edn. Springer, Heidelberg (2013)
7. Kindberg, T., Chalmers, M., Paulos, E.: Guest Editor's Introduction: Urban Computing. Pervasive Computing, vol. 6(3), pp. 18–20. IEEE Press, New York (2007)
8. Kostakos, V., O'Neill, E.: Cityware: urban computing to bridge online and real-world social networks. In: Foth, M. (ed.) Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, Information Science Reference, pp. 196–205. Hershey, New York (2009)
9. Foth, M., Jaz, H.-J., Satchell, C.: Urban informatics. In: Conference on Computer Supported Cooperative Work (CCSCW), ACM, pp. 1–8 (2011)
10. Cranshaw, J., Schwartz, R., Hong J.I., Sadeh, N.: The livehoods project: utilizing social media to understand the dynamics of a city. In: Sixth International AAAI Conference on Weblogs and Social Media (ICWSM), AAAI, pp. 58–65 (2012)
11. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In: Workshop on the Social Mobile Web at ICWSM 2011, AAAI, pp. 570–573 (2011)

12. Del Bimbo, A., Ferracani, A., Pezzatini, D., D'Amato, F., Sereni, M.: Livecities: revealing the pulse of cities by location-based social networks venues and users analysis. In: 23rd International Conference on World Wide Web (WWW), ACM, pp. 163–166 (2014)
13. Xia, C., Schwartz, R., Xie, K., Krebs, A., Langdon, A., Ting, J., Naaman, M.: Citybeat: real-time social media visualization of hyper-local city data. In: 23rd International Conference on World Wide Web (WWW), ACM, pp. 167–170 (2014)
14. Kloeckl, K., Senn, O., Ratti, C.: Enabling the real-time city: LIVE Singapore! J. Urban Technol. **19**(2), 89–112 (2012)
15. Bozzon, A., Fraternali, P., Galli, L., Karam, R.: Modeling crowdsourcing scenarios in socially-enabled human computation applications. J. Data Semant. **3**, 169–188 (2014)
16. Psyllidis, A., Biloria, N.: Ontopolis: a semantic participatory platform for performance assessment and augmentation of urban environments. In: 10th IEEE International Conference on Intelligent Environments 2014 (IE 2014), IEEE Press, New York, pp. 140–147 (2014)
17. British Standards Institution (BSI): Smart cities – vocabulary. BSI Standards, London (2014)
18. British Standards Institution (BSI): Smart City concept model – guide to establishing a model for data interoperability. BSI Standards, London (2014)
19. Smart Cities Council: Smart Cities Readiness Guide: The Planning Manual for Building Tomorrow's Cities Today. Smart Cities Council, Redmond (2014)
20. European innovation partnership on smart cities and communities (EIP-SCC): operational implementation plan. EIP-SCC, Brussels (2014)
21. Psyllidis, A.: OSMoSys: a web interface for graph-based rdf data visualization and ontology browsing. In: Cimiano, P., Frasincar, F., Houben, G.-J., Schwabe, D. (eds.) 15th International Conference on Web Engineering (ICWE 2015). LNCS. Springer, Heidelberg (2015, in press)

http://www.springer.com/978-3-662-47385-6

Computer-Aided Architectural Design: The Next City – New Technologies and the Future of the Built Environment
16th International Conference, CAAD Futures 2015, São Paulo, Brazil, July 8-10, 2015. Selected Papers
Celani, G.; Sperling, D.M.; Franco, J.M.S. (Eds.)
2015, XIV, 622 p. 380 illus., Softcover
ISBN: 978-3-662-47385-6