3cixty@Expo Milano 2015: Enabling Visitors to Explore a Smart City

Giuseppe Rizzo, Raphaël Troncy, Oscar Corcho, Anthony Jameson, Julien Plu, Juan Carlos Ballesteros Hermida, Ahmad Assaf, Catalin Barbu, Adrian Spirescu, Kai-Dominik Kuhn, et al.

3cixty@Expo Milano 2015
Enabling Visitors to Explore a Smart City

Giuseppe Rizzo,[1] Raphäel Troncy,[1] Oscar Corcho,[2] Anthony Jameson,[3] Julien Plu,[1] Juan Carlos Ballesteros Hermida,[2] Ahmad Assaf,[1] Catalin Barbu,[3] Adrian Spirescu,[3] Kai-Dominik Kuhn,[3] Irene Celino,[4] Rachit Agarwal,[5] Cong Kinh Nguyen,[5] Animesh Pathak,[5] Christian Scanu,[6] Massimo Valla,[6] Timber Haaker,[7] Emiliano Sergio Verga,[4] Matteo Rossi,[8] José Luis Redondo García[1]

[1] EURECOM, France
[2] Universidad Politécnica de Madrid, Spain
[3] DFKI, Germany
[4] CEFRIEL, Italy
[5] Inria, France
[6] Telecom Italia, Italy
[7] InnoValor, The Netherlands
[8] Politecnico di Milano, Italy

Abstract. Planning a visit to Expo Milano 2015 or simply touring in Milan are activities that require a certain amount of *a priori* knowledge of the city. In this paper, we present the process of building such a comprehensive knowledge base, the 3cixty KB, that contains descriptions of events, places, transportation facilities and social activities, collected from numerous static, near- and real-time local and global data providers, including Expo Milano 2015 official services and several social media platforms. Entities in the 3cixty KB are deduplicated, interlinked and enriched using semantic technologies. The 3cixty KB is empowering the ExplorMI 360 multi-device application, which has been officially endorsed by the E015 Technical Management Board and has gained the patronage of Expo Milano 2015, thus offering a unique testing scenario for the 20 million expected visitors along the 6 months of the exhibit. As of September 7th, 2015 the 3cixty KB contains unique descriptions of $18,665$ events, 758 artists, $225,552$ places, $9,342$ transportation facilities, $95,570$ illustrating photos and $94,789$ reviews contributed by $81,944$ users. The 3cixty KB is accessible at `http://3cixty.eurecom.fr/sparql` while ExplorMI 360 at `https://www.3cixty.com`.

1 Pitch

Can we present the city of Milan in such a way that visitors can digitally plan in advance a visit, and explore what is happening once being in the city with the support of a digital guide? In the context of Expo Milano 2015, we have answered this question by collecting a vast amount of cultural and touristy data about the city of Milan from the E015 digital ecosystem,[9] and social media

---

[9] `http://www.e015.expo2015.org`

platforms (such as Facebook, and Google). The resulting knowledge base, so-called the 3cixty KB, acts as a curated data marketplace which is then used to feed the ExplorMI 360 multi-device application composed of a web application and mobile companion guides available for Android and iOS devices. The 3cixty KB contains descriptions of cultural events, places and their reviews and transportation means offered in a city. It also includes $285,773$ links to external resources on the Web providing attribution and enrichment. As optimization, we have decomposed the geographical extent into cells and materialized travel time distances among the different geographic points. This offers new opportunities for Expo Milano 2015 visitors to learn what will happen during the 6 months exhibit, how to get there, and more generally, how to take advantage of the city including visiting its sights or discovering some hidden places in a completely personalized experience through the ExplorMI multi-device application. The web application enables city visitors to explore the diverse entities of the 3cixty KB in several directions in parallel, that can be saved in a personalized wishlist. The mobile application, instead, offers a companion digital guide that is mostly used by visitors once they are in the city, and want to be reminded about their favorite places and events added in their wishlist. In addition, the 3cixty KB is also intended to be used by developers to build their own apps and to extend the city digital data ecosystem. The modular design of our workflow makes the 3cixty KB extensible to any other city in the world and we have started to extend the 3cixty KB to cover the cities of London (with a focus on the Queen Elizabeth Olympic Park) and Nice. *Data is the new oil* we repeatedly hear and are convinced of. But the typical city digital ecosystem is scattered across different data sources and with different accessing mechanisms [6]. The 3cixty KB aims to collect and harmonize descriptions of entities in one place, and to provide additional services such as reconciliation and deduplication for offering comprehensive and accurate descriptions, data analytics, data visualization and mobility optimization such as providing the time it takes to reach any place or event using public transportation from any point in the city. The 3cixty KB can empower a variety of applications and business models, as it is evident from the commercial interest from numerous companies.

2   Technical Description

The proposal of building a knowledge base for a city by combining open and private data, including data verification, reconciliation and validation, is not new. The KM4City (Knowledge Model for City) project has conducted an expert study over smart city datasets that has resulted in a new comprehensive ontology that aims to become a standard for smart cities [1]. The CitySDK project[10] offers a suite of uniform APIs to turn cities' data ready to be consumed by web developers. The project focuses on three dimensions: citizen participation, mobility and tourism. The STAR-CITY project focuses on traffic analytics for the city of Dublin, integrating numerous sensor data using semantic web technologies for accurately predicting road traffic conditions [5]. Distinguishing features

---

[10] http://www.citysdk.eu

of the 3cixty KB are: data-centric view of the platform, built on top of well-known ontologies, and exposed as a curated data marketplace to developers via semantic web technologies. We summarize the technical challenges we address for building the 3cixty KB and how it is used to offer services to end-users.

2.1 Knowledge Base Population

The population of the 3cixty KB follows the workflow: data source selection, data modeling, geographic grid data sampling, and data reconciliation [6].

Data Source Selection. A strict procedure has been followed for selecting the data sources to be included in the 3cixty KB, with the goal of maximizing the spatial and temporal coverage of data, its complementary in terms of semantics, its real-time nature and, more importantly, its usefulness for city visitors. An agile process has been established for continuously monitoring and updating the list of potentially relevant sources. Three types of data sources are used: E015 data services, social media platforms, and additional data about the Expo itself that has been generated by an editorial team (Table 1). The E015 digital ecosystem aims at sharing data relevant for Expo Milano 2015 visitors through services offered mostly by local providers. We combined such a wealth of data with content publicly available on social media platforms, the rationale being to increase the coverage of the points of interest and to complement the description of existing events with user-generated activities. Finally, our last data source is the list of events advertised by the various pavilions in the exhibit and the list of places where those events are taking place. An editorial team composed of five people have mapped on-site the geographical area and are continuously scraping the numerous agenda of events per pavilion.

Entity-Type	E015	Social Media Platform	Expo
Event	Expo in Città (19.51%), Teatri per Milano (0.52%), Fiera Milano (0.25%), Fondazione Arnaldo Pomodoro (0.02%), Leonardo Ambrosiana (0.01%)	Evensi (68.02), Eventful (5.64%), Lastfm (1.78%)	3cixty (4.24%)
Place	Dati Comune Milano* (2.18%), BikeMI (0.01%) Vaxita (0.29%), Expo in Città (0.24%), Fondazione Arnaldo Pomodoro (0.01%), Isnart (0.23%), Teatri per Milano (0.01%)	Yelp (32.17%), Google Places (30.69%), Facebook (19.58%), Foursquare (14.24%), Expedia (0.13%), Evensi (0.06%), Eventful (0.01%)	3cixty (0.16%)
Transport	Dati Comune Milano* (52.46%) Vaxita (7.05%), BikeMI (0.29%), Isnart (0.29%)	Foursquare (17.42%), Facebook (15.67%), Yelp (3.11%), Expedia (3.02%), Google Places (0.65%)	
Media	Expo in Città (3.4%), Vaxita (0.68%), Teatri per Milano (0.06%) Fondazione Arnaldo Pomodoro (0.01%), Leonardo Ambrosiana (0.01%)	Facebook (30.73%), Forsquare (24.09%), Evensi (13.51%), Google Places (12.77%), Yelp (11.42%), Lastfm (1.2%), Expedia (0.53%), Flickr (0.33%), Eventful (0.25%)	3cixty (1.09%)
Review		Foursquare (63.15%), Google Places (29.92%) Yelp (10.93%)	

Table 1: Breakdown figures per entity and data source type. The percentage indicates the contribution of the data source with respect to the total amount of unique entities (after entity reconciliation and resolution) per type in the 3cixty KB as of September 7th, 2015 (* is part of the public data offered by the Milan municipality)

Data Modeling. Our ontology design principle has focused on optimizing the coverage of the terminology in the context of city exploration. For each entity to model, we looked for existing ontologies by keyword searching in the LOV, Swoogle, Watson, and Smart City catalogues. We selected terms based on popularity measured through linked data usage and favoring schema.org when suitable. Two domain experts analyzed the ontologies that resulted from the search, and once consensus was reached, ontologies were taken and added to the 3cixty data model, which, therefore, consists of a network of existing ontologies. We have finally modeled a few additional properties published at `http://3cixty.eurecom.fr/ontology`. The documentation of the 3cixty data model with examples is available at `http://3cixty.eurecom.fr/documentation`.

Geographic Grid Data Sampling. The data fetching process uses, depending on the data source, a real-time and/or a batch procedure. The starting point is a grid covering a geographical area. In the case of the city of Milan, we re-used the Milan grid proposed in the 2014 BigData Challenge[11] and composed of 10,000 squared cells of 235x235 meters with a focal point in the center of the city. All objects in the 3cixty KB are either described by WGS84 points (e.g. places) or are related to other objects described as points (e.g. events that happen in places). In addition, every object belongs to a cell of the grid which enables us to optimize both the reconciliation process and the query evaluation time when computing travel distance queries such as listing all restaurants that are within 10 minutes walk from a particular hotel. E015 data is fetched through a proxy server that provides a uniform mechanism to access sources implemented using heterogeneous technologies. Dedicated collectors have been developed for each social media platform using their respective APIs. OpenRefine[12] is used for transforming data manually entered in tabular format. For any data source, descriptions of the real world objects instantiate the 3cixty model while keeping an attribution using the `dc:publisher` property. We generate timestamped dumps for each data source that are serialized in the Turtle syntax and updated weekly. To minimize the cost of the collection process, we only harvest the delta composed of the new instances available from each source. For handling real-time data (e.g. live positioning of the bus and metro, rooms availability in hotels), we will rely on publish-subscribe mechanism implemented in the general purpose complex event processing engine T-Rex [3]. We use a set of SPARQL construct queries to materialize frequent computations such as planned time-travel distances among places and events. Each dump is then loaded in a dedicated named graph for easing data management. Our triple store is the Virtuoso open source version.

Data Reconciliation. The different data sources overlap in terms of instances (e.g. the same hotel exists in multiple data sources). This concerns at least 0.67% of Events and 26.24% of Places. On two uniform randomly sampled datasets of 100 places and 100 events, the algorithm has an accuracy of 90.24% for places and 99.6% for events [6]. We perform pairwise reconciliation of the data sources,

---

[11] `https://dandelion.eu/datamine/open-big-data`
[12] `http://openrefine.org`

included self-reconciliation which is frequent within social media source using SILK. For the entity resolution, we adopt a conservative approach where we add only one element when multiple sameAs links were identified by the reconciliation procedure. Such sets are thus obtained as the distinct union of the outputs of the reconciliation process with all instances that do not hold any sameAs link. They are loaded in separate graphs, namely *events* and *places* graphs of reconciled entities, which are used by ExplorMI 360.

Both places and events are categorized according to the schema used by the data source providers[13] and the mapping established with two pivots schemes used to enable the parallel exploration of the data: the Foursquare taxonomy[14] that organizes places into a three-level hierarchy, and another taxonomy for events described in [4]. The category alignment has been supervised by two domain experts, using both the category label and description as well as the set of instances that belong to those categories. The result is represented using `skos:closeMatch` and `skos:broadMatch` axioms.

2.2   Knowledge Base Exploration and Usage

The 3cixty KB can be visualized through the ExplorMI 360 multi-device application or accessed via SPARQL and a dedicated API.

ExplorMI 360. ExplorMI 360 is intended to offer a 360° view of the city in different stages of the visit: a web application for preparing a visit to Milan and a mobile one as a digital companion guide for visitors. The web application implements a Parallel Exploration paradigm which extends Parallel Faceted Browsing [2] by enabling the display of multiple interrelated queries and their results on the screen at the same time. It also allows users to filter places or events in some novel ways such as *(a)* specifying a word or phrase that must appear somewhere in the texts associated with the item; *(b)* manually selecting items from a result list; or *(c)* specifying that a place or event must be within a certain distance of a specified location. It also allows pivoting using relationships among entities, such as discovering all hotels that are located within a certain distance of at least one park. These advanced features help to exploit fully the power of the 3cixty KB by enabling users to deal with multiple types of entities and multiple result lists at the same time. But this power is associated with a richness and complexity of the user interface that can be daunting initially to new users. For this reason, we have developed features that make it easier for new users to become accustomed gradually to the power of Parallel Exploration: *(a)* Zoom View: the unconventional inverted tree presentation of result lists on the screen is replaced with a more familiar set of result lists arranged in parallel columns; *(b)* Quick Start Apps enable users to create trees of interrelated queries simply by clicking on natural language questions within a menu. For example, if the user clicks on the question "What sorts of concerts are taking place during

---

[13] For example, Google Places use the set of categories described at `https://developers.google.com/places/supported_types`

[14] `https://developer.foursquare.com/categorytree`

the coming week and on the weekend?"[15], the same tree of results is created and shown as if the user had constructed it step by step himself. The tree is typically accompanied by several tooltips that explain to the user how he can explore the tree of results and also how he can change the details of the queries to obtain results that are more relevant to him (e.g., replacing "500 meters" by "200 meters"). Quick Start Apps are technically simple since they just associate a set of questions with a set of bookmarks that correspond to their answers. Hence it is possible even for non-technical users to create Quick Start Apps using a point-and-click interface. Finally, the web application includes a developer mode within which a developer can view the SPARQL queries generated by the user interface, so as to be able to reuse these queries (or variants of them) in other contexts.

The mobile application is available for Android and iOS. It guides users visiting and exploring the city by showing relevant and nearby information about events, points of interest, and transportation. Users can synchronize and modify their wishlist created beforehand in the web application, visualizing details about events and places and directions for reaching them. They can also receive notification alerts while walking around the city and getting closed to favorite items. The *mobility profile* feature allows users to authorize the application to track their movements and to generate a heatmap of their movements, suggesting them to discover new areas they have not yet visited. They can explore and discover new events and places, and add them into their wishlist. Visitors can also benefit from the *Thematic Tours* feature that helps plan visits at Expo Milano 2015 and optimises time. For example, by selecting a list of tags (`art and craft`, `music and show`, `family`) and the time available (e.g. night, day) users can get a list of suggested pavilions to visit or events to attend, taking into account the walking time between places and the average visit duration. The first user tests at the Expo indicate that the digital guide powered by the 3cixty KB offers a valuable and unique way to plan a visit and explore a city.

SPARQL and API: a Developers' Data Marketplace. 3cixty offers new opportunities to developers to build apps that can either access a SPARQL endpoint of use a dedicated API. The main differences are: cache support (API), personal data (API) versus freedom and expressivity in querying (SPARQL). The API offers automatic integration of public and personal data for an enhanced and personalized user experience. Some of these experiences are related to the social aspects, namely: augmentation/re-ranking of results based on items selected in the wishlist of friends, and augmentation/re-ranking of results based on items reviewed by users who are friends. Since the API provides user-personalized results, it is secured and can only be accessed via API key and/or user access token.

---

[15] `https://www.3cixty.com/webApp/#bookmark=GFBV9gyKaCtvfxV7wBO5IWvxNeYl4MCO-5`

References

1. Bellini, P., Benigni, M., Billero, R., Nesi, P., Rauch, N.: Km4City ontology building vs data harvesting and cleaning for smart-city services. Journal of Visual Languages & Computing 25(6), 827–839 (2014)
2. Buschbeck, S., Jameson, A., Spirescu, A., Schneeberger, T., Troncy, R., Khrouf, H., Suominen, O., Hyvönen, E.: Parallel Faceted Browsing. In: International Conference on Human Factors in Computing Systems (Interactivity Track) (CHI) (2013)
3. Cugola, G., Margara, A.: Complex Event Processing with T-REX. Journal of Systems and Software 85(8), 1709–1728 (2012)
4. Khrouf, H., Milicic, V., Troncy, R.: Mining events connections on the social web: Real-time instance matching and data analysis in EventMedia. Web Semantics: Science, Services and Agents on the World Wide Web 24(1), 3–10 (2014)
5. Lécué, F., Tallevi-Diotallevi, S., Hayes, J., Tucker, R., Bicer, V., Sbodio, M.L., Tommasi, P.: STAR-CITY: Semantic Traffic Analytics and Reasoning for CITY. In: 19th International Conference on Intelligent User Interfaces (IUI) (2014)
6. Rizzo, G., Corcho, O., Troncy, R., Plu, J., Ballesteros Hermida, J.C., Assaf, A.: The 3cixty Knowledge Base for Expo Milano 2015: Enabling Visitors to Explore the City. In: 8th International Conference on Knowledge Capture (K-CAP) (2015)

A   Evaluation Requirements

The application has to be an end-user application:
The 3cixty platform enables to develop end-user applications on top of a high quality data marketplace. ExplorMI360 is a multi-device application composed of a web site that enables city visitors to plan a trip by exploring and bookmarking what a city will offer at particular time and a mobile companion that guides the visitors showing what is nearby and alerting about all the items previously saved in a wishlist.
The information sources used:
The 3cixty KB is built using three types of data sources: the data services of the local providers participating to the E015 digital ecosystem, web content shared on social media platforms, and data about the Expo 2015 exhibit itself contributed by an editorial team which further improves the precision and coverage of places and events that are taking place. Data comes under *diverse ownership* and formal agreements have been signed with the providers of E015 digital ecosystem to access and reuse their information. Data is *very heterogenous*, with different structure and semantics, but the usage of common ontologies and graph-based RDF descriptions has largely eased the data integration process as well as the deduplication of the entities that are common across sources. Data describes *real world objects* sampled from *real world volume* in the domain of culture and tourism: places including sights and businesses, events and performers, transportation and user-generated content such as illustrative media and reviews.
Attractive and functional Web interface:
The web-based and mobile components of ExplorMI 360 provide two paradigms of interactions: Parallel Exploration enables to search and browse in a tree-style navigation in desktops and tablets devices; A mobile digital guide enables a geolocalized user to discover nearby places and events or to get notified about wishlisted items.

The meaning of data has to play a central role:

The overall design of the 3cixty KB has followed a data-centric approach: entities are described with common properties and are related to each other with relationships. This allows the user to generate complex queries mixing time, distance and social dimensions such as "What are the 4 stars hotels, well-reviewed by my friends, located within a few minutes by metro from this concert given at La Scala?". The TBox makes use of well-known ontologies. The ABox contains descriptions of numerous entities that are relevant in the domain of culture and tourism. All together, data analytics processes can be used to derive new insights about the city, such as generating urban summaries based on the concentration of businesses in various areas and the social activity, or better classifying businesses into fine-grained categories based on aggregated user-generated review content. The use of semantic technologies greatly simplifies all data integration steps, including the curation and the ingestion of new instances, and the discovery of duplicates which is fundamental given that we are integrating 20 data sources. It also enables the development of flexible visualization widgets, offering a standard mechanism for user interfaces to pull the data. We also offer personalization services using user profile data. User profile data is composed of basic user information, friend list and a mobility profile tracking the user movements. User-generated reviews aggregated over multiple sources for the places combined with one's friend list enables to rank higher the places favorited by one's friends, thus offering a social ranking in which the user presumably trusts better. Friends wishlists can finally be used to alert a user when he is nearing a particular place in the city.

Scalable application:

As of September 7th, 2015, the 3cixty KB contains around 10 million triples solely for the city of Milan. We are extending the 3cixty KB coverage to bigger geographic areas such as London and Nice. To offer a smooth user experience *(a)* some metadata, which are heavy to be computed on-the-fly such as travel distances, are materialized; *(b)* the API offers a caching mechanism and optimized queries.

Rigorous evaluations:

(a) Rigorous approach of data modeling: a team of two experts defined a list of inclusion/exclusion criteria and designed the TBox. *(b)* The data reconciliation, inspired by an already well-performing and published work [4], has been extended and the results have been statistically evaluated. *(c)* The user interfaces, which include some highly innovative elements, have been subjected to numerous phases of iterative evaluation.

Novelty:

Variety: events, places, transportation and user generated content are integrated using semantic technologies. *Veracity*: using local data sources, we optimized precision. *Volume*: 20 data sources totaling 10M triples. `Exploration`: The Parallel Exploration user interface represents a novel and powerful way of exploring semantic data, for both end-users and developers.

Functionality:

The 3cixty KB enables: *(a)* efficient semantic search, via facets, for data visualization and multilingual full text search on all labels. *(b)* geographic search, using a grid-based approach for indexing instances, thus optimizing the evaluation times when computing travel distance queries.

Contextual Information:

(a) The data reconciliation exploits contextual information about the type of content offered from the previously listed data sources. *(b)* The instances have a `topo:importance` property that indicates, with a range from 0 to 5 (Likert scale), their quality/importance. *(c)* The user interfaces exploit this to perform dynamic ranking in the visualization layer.

Multimedia documents are used in some way:

Images (and videos) are described and used to illustrate the two main entity types: event and place.

Use of dynamic data:

We crawl and reconcile near- and real-time data such as availability of hotel rooms.

Accurate results:

From the experimental settings, the results are encouraging both in terms of figures (quantitative analysis) and in terms of user experience (qualitative analysis) as indicated in the technical description.

Support for multiple languages and accessibility on a range of devices:

Textual metadata are offered in two languages (English and Italian) in the 3cixty KB. In addition, the web-based component is specifically designed to work well with Google Translate, in that texts that are part of the user interface (e.g. items listed in menus) are formulated in a relatively verbose way (e.g. "events that have the category Musical Concert") that tends to yield good automatic translations (as well has increased clarity for new users).