Citizen-centric Linked Data Apps for Smart Cities

Diego López-de-Ipiña[1], Sacha Vanhecke[2], Oscar Peña[1], Tom De Nies[2] and Erik Mannens[2]

[1]Deusto Institute of Technology, DeustoTech, University of Deusto, 48007 Bilbao, SPAIN
[2]Ghent University – iMinds – Multimedia Lab, Gent, Belgium
{dipina, oscar.pena}@deusto.es,
{sacha.vanhecke, tom.denies, erik.mannens}@ugent.be

Abstract. Open Government Data combined with dynamic data collected from either citywide sensor networks or apps running in citizens' smartphones offer ample potential for innovative urban services (apps) that realize Smarter Cities. This paper describes IES CITIES, a platform designed to facilitate the development of urban apps that exploit public data offered by councils and enriched by citizens. This solution addresses the needs of three main stakeholders in a city: a) citizens consuming useful data services in different domains but also contributing with complementary data to the city, b) companies leveraging the simple JSON-based RESTful API provided by IES CITIES to create novel urban apps, and c) the City Council, using the platform to publicize its urban datasets and track services assembled around them.

Keywords. Smart City, Linked Data, Apps, provenance, trust, JSON

1 Introduction

The world is undergoing the largest wave of urban growth in history. Experts predict the global urban population will double by 2050 meaning 70% of the total world population will be living in a major town or city. Consequently, there is an increasing need to assemble Smart Cities that effectively and efficiently manage the resources required by their increasing populations. A city may be considered smart when it improves the citizens' quality of life and the efficiency and quality of the services provided by governing entities and businesses.

The IES Cities platform is defined to promote user-centric mobile micro-services that exploit open data and generate user-supplied data. It contributes with an open citizen-centric Linked Data apps-enabling technological solution. Importantly, it focuses on enabling citizens to create, improve, extend and enrich the open data associated to a city in which micro-services, i.e. urban apps, are based. The main stakeholders of the resulting urban apps ecosystem are the citizens, SMEs and public administration within a city.

This platform takes advantage of the progress achieved lately in two key technological areas: a) *open government and urban sensor generated datasets*, b) *smartphones equipped with different sensors*, e.g. GPS, which can execute urban

apps, i.e. offering services for citizens in different domains (e.g. transport, security and so on).

Following the "Apps for Smart Cities Manifesto"[1] approach, IES CITIES aims to enable an ecosystem of urban apps that help citizens in their daily activities and actions within the city. In order to accomplish this, it addresses three main challenges: a) *extract and adapt heterogeneous structured and non-structured data* from council repositories, sensor networks, web sites and social networks, b) *validate, promote and integrate user-provided data* with open government data and c) *facilitate the development of urban apps by end developers*, thus fostering urban-related innovation.

2 Related work

Citizen execution of urban mobile apps may generate new data that enriches the datasets associated to a given city. However, the quality of the data generated has to be assessed and qualified, thus promoting valuable and trustable information and decrementing and eventually discarding lower quality data. The W3C has created the PROV Data Model [1] for provenance interchange on the web. PROV is a specification to express provenance records, which contain descriptions of the entities and activities involved in producing and delivering or otherwise influencing a given object. Provenance can be used for many purposes, such as understanding how data was collected so it can be meaningfully used, determining ownership and rights over an object, making judgements about information to determine whether to trust it, verifying that the process and steps used to obtain a result complies with given requirements, and reproducing how something was generated [2].

Human Computation [3] enables to leverage human intelligence to carry out tasks that otherwise would be difficult to accomplish by a machine. Different techniques are used to incentivise user participation, e.g., gaming with a purpose (GWAP) [4] to encourage people to help in human-computation through entertaining incentives. Mobile games with a purpose have already been used to annotate and enrich urban data, e.g. UrbanMatch [5] validates links between points of interest and their photos whilst Urbanopoly [6] adds and validates Venue-Feature-Value (VFV) facts. This work fosters the generation of provenance-based trusted citizen contributed Linked Data. Users' contributions are mediated by the IES CITIES-enabled apps that leverage the back-end provenance support of our platform, comprising a generic infrastructure for developers to easily create urban apps that consume and provide confidence ranked Linked Data about a given city.

Lately, some JSON query[2] languages and schemas[3] have emerged that intend to enable end-users to more easily specify and exploit the JSON data model. In our view, urban apps will be assembled from structured and non-structured data in the form of RDF, CSV or even applying web scrapping to HTML pages. Information in such models can be mapped into JSON, a *lingua franca* for web and mobile develop-

---

1 http://www.appsforsmartcities.com/?q=manifesto
2 http://www.jsoniq.org/
3 http://json-schema.org/latest/json-schema-core.html

ers that do not necessarily have to be exposed to the complexities of semantically modelled data. This work employs the TheDataTank (TDT)[4], a distributed open-source web framework that serves as a dataset adaptor to transform online data of different formats into a user-friendly Restful API.

3 The IES CITIES platform

The preliminary IES CITIES solution is composed of the following elements: a) a *mobile application*, namely the *IES CITIES player*, that allows users to search for and browse over available urban apps, based on their location and filters, and then execute these services and b) a *server* that acts as a mediator among urban apps front-ends played in citizens' smartphones and their business logic implementing back-ends accessing, exploiting and enriching publicly available datasets. Notably, IES CITIES allows users to contribute with and validate others' contributed data, through a mechanism for tracking information provenance and its associated trust and confidence.

The *modus operandi* of the platform is as follows. Firstly, the municipality registers with the IES Cities server its datasets descriptions, by means of a web form. It indicates where the dataset can be located and accessed (URI), what is the original format of the data (CSV, RDF, XML and so on), a description of the dataset expressed in JSON-Schema and, optionally, a mapping script between the original data source format and JSON, and *vice versa*. Secondly, a developer finds which datasets are available by browsing or searching in the IES Cities' dataset repository, and decides which ones best fit his application. Through a RESTful JSON-based API, she issues queries over the datasets abstracted as JSON data structures and retrieves results also expressed in JSON. Thirdly, once the application development has been completed, the developer registers it with the platform through a web form, providing among other details where the application is (URI), its type (Google Play, Local app repository, Web) and a description of its functionality, including snapshots. Finally, end-users, i.e. citizens, with the help of the IES CITIES player app, browse or search for available registered urban apps according to their location and interests.

3.1 Linked Data Design

Involving end-users in the creation of Linked Open Data implies that these data should be published according to some RDF specifications, adding some particular characteristics to the data that associate it to the IES Cities platform, and enforcing its reusability inside or outside the context of the IES Cities platform. Several widely used vocabularies were considered, such as Dublin Core (DC), which defines general metadata attributes such as title, creator, date and subject, and FOAF, that defines terms for describing persons, their activities and their relations to other people and objects. Reusing terms maximizes the probability that data can be consumed by applications that may be tuned to such well-known vocabularies, without requiring further pre-processing of the data or modification of the application [7].

[4] http://thedatatank.com/

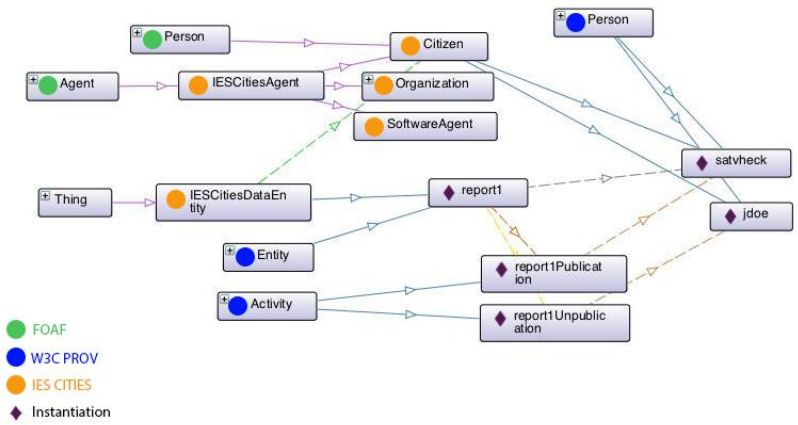

Fig. 1. IES Cities RDF schema

The IES Cities ontology[5] (see Fig. 1) defines two general classes. Firstly, the base for any data generated through the platform is the class `IESCitiesDataEntity`, which should at least include a DC term identifier property to easily query the data and if possible a description. Secondly, in order to incorporate the users of the platform in the model, a general `IESCitiesAgent` class is defined, that extends FOAF's Agent, and several subclasses such as `Citizen`, inheriting from FOAF's Person class, `Organization` and `SoftwareAgent`.

Enabling user contribution to open datasets requires incorporating metadata that enables tracking down the creator of a particular data entity, known as provenance data. A complete provenance ontology that fits the needs of the IES Cities project by separating the data from actual creation, modification and invalidation tasks is the PROV Ontology [8]. Starting Point classes, such as Entity, Activity and Agent, provide the basis for the rest of the PROV Ontology. Expanded terms provide additional terms that can be used to relate classes. These two categories are used in the IES Cities vocabulary to associate users as the publishers or invalidators of particular data.

Related to provenance are the credibility of data and the trustworthiness of their authors. IES Cities currently incorporates a simple mechanism to measure data credibility. Credibility is modelled in a 0 to 1 scale. A credibility score of 1 is assigned to each `IESCitiesDataEntity` at initialization, i.e. the system initially gives 100% credibility to the contents created. The resulting credibility level (CL) is recalculated every time a user votes up or down the published data, by following the simple formula: `CL = numPositiveVotes/numTotalVotes`. A more sophisticated credibility score calculation should take into account the user's reputation, based on earlier contributions. Additionally, contents could be rated according to a scale (e.g. 1 to 5) rather than applying the simple binary (correct vs. incorrect) approach now used.

---

5 http://studwww.ugent.be/~satvheck/IES/schemas/iescities.owl

3.2 Implementation details

Fig. 2 shows the IES CITIES platform architecture. The IES CITIES player mobile application is provided to access the services registered in the platform. Information about the services and users is persisted in a PostgreSQL database whilst open datasets are registered and accessible through two publication engines, namely CKAN[6] for structured RDF datasets and TheDataTank[7] for unstructured data. A Virtuoso RDF store is used to maintain linked datasets and store user-provided RDF data.

Open data is fetched through a JSON-formatted query sent to the server's RESTful "`/data/`" interface, together with the name of the requested datasets. This query consists of key/value pairs to specify required fields and optionally some parameters. The server-side's DataManager component resolves the location of the requested dataset from the publication engines and transforms the JSON query into the query language specific to the nature of the data's resolved endpoint. Currently, a query mapper for both SPARQL, the query language for RDF, and SPECTQL, the query language used by TheDataTank, is available. After executing the query, the resulting data is returned, in JSON format. The structure of the JSON submitted queries is:

```
{
"type":"data",
"requested”:{ "predicate1":"object1", "predi-
cate2":"object2" },
"optional" :{ "predicate3":"object3" },
"given"    :{ "predicate2":{ "type":"string",
                "value":"object2_value" }}
}
```

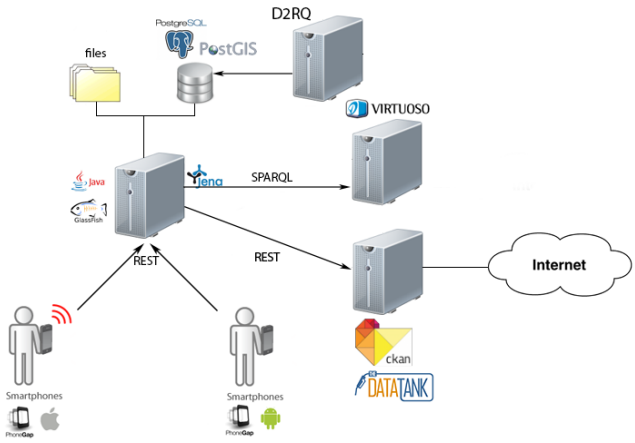

Fig. 2. Deployment of IES CITIES platform

6 http://ckan.org
7 http://thedatatank.com

A similar mechanism is also used for the generation and validation of data, using the SPARQL/Update extension and adding provenance meta-data compliant to the PROV-O vocabulary. This makes it, however, only available for linked datasets.

The IES Cities Player was developed using PhoneGap[8]. Fig. 3 shows a listing of services available on the user's location and ordered by average rating shown by the player app. Apart from location based look-up, users can also browse services using a set of keywords. When the user selects a service, a description is displayed, along with a list of reviews by other users and the average rating (Fig. 3). By clicking the Start button, the player launches the actual urban app.

4 Validation

The 311 Bilbao (Fig. 4) app has been implemented and deployed to validate the IES Cities platform. It uses Linked Open Data to get an overview of reports of faults in public infrastructure. It demonstrates how a developer can create a complex mobile app relying on semantic data, without technical knowledge of the SPARQL query language or underlying OWL ontologies. A query using the JSON query format proposed has to be assembled and issued to the "`/data/`" RESTful interface.

During initialization, the service queries for reported faults, and displays the result on a map (Fig. 4a). Using the filter functionality on the second tab (Fig. 4b), a user can choose to see reports of only a certain type. By clicking on the marker of a particular report, the ID and the underlying nature of the reported fault (Fig. 4a) is displayed. When the user decides to inspect the report an information page (Fig. 4c) is shown. On this page, users can see the full description and a photo of the report. Notably, they are also able to vote the credibility of the report up or down. Finally, they can create their own reports (Fig. 4d), specifying the faults nature, its location, a description and optionally add an image.

Fig. 3. Searching and browsing servicers.

8 http://phonegap.com/

From the data owner's point of view, enrichment of their datasets by third parties, such as users of the 311 Bilbao app, revealed two problems. The first one is the fact that data does not need to be approved before being published. Despite the fact that users are offered functionality to first review earlier submitted reports in the location where they are placed, some people may still add redundant reports. A future version of this application should enable users to further enhance and comment an already available report. In addition, proximity and content similar reports should be automatically grouped. Additionally, there is a need for a way to consider the reputation of the different authors, e.g. citizens and council staff. The owner of the dataset could use this to prioritize the processing of data from particular sources.

Fig. 4. 311 Bilbao IES CITIES service.

5 Conclusion and Further Work

The IES CITIES platform can be used by developers to facilitate the use of open data from arbitrary formats, relying on CKAN and the adaptor functionality of TheDataTank; all of it through a simple JSON query language. Moreover, the exemplary mobile app developed demonstrates how citizens' involvement in the management of a city can be increased by allowing them to actively participate in the creation of new data and the validation of open data provided by other users. These services also show the IES Cities platform's added value for public bodies, who can easily publish their open data in different non-proprietary formats, while making them accessible as common machine-readable format (JSON) through uniform REST interfaces.

Further work should address the following issues. More complex, derived provenance information should be incorporated, allowing tracing back the full revision chain of data modifications. Recent research has shown that the reconstruction of provenance, when it is (partially) missing, is feasible [9] and can be incorporated to complement our tracking approach. The current mechanism for defining credibility should also be extended. Instead of only setting a default score at creation time of the data, some level of 'trust' should be calculated as well, in an algorithmic manner,

based on the data's actual source and its changes [10], making it possible to prioritize the processing of data of more trusted parties, such as official public bodies, over data provided by citizens. The reputation of certain sources is to be calculated over time, and used to set an initial value for a source's credibility. However, in addition to reputation, provenance is an essential component towards trust assessment as well [11]. Applying automatic reasoning over the provenance of the data, seems a feasible method to implement these principles, as it only requires the addition of lightweight annotations to the provenance that is currently tracked [12]. These annotations allow for an on-the-fly selection of data, based on the preference and trustworthiness of its source trace.

References

1. W3C.: PROV Model Primer. Working Group. Note 30 April 2013, (2013), http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/
2. Groth P. and Gil Y. Editorial - Using Provenance in the Semantic Web. In: Web Semantics: Science, Services and Agents on the World Wide Web, vol. 9, no.2, Elsevier (2011)
3. Bozzon B., Galli L., Fraternali P., Karam R.: Modeling CrowdSourcing scenarios in Socially-Enabled Human Computation Applications. To Appear in Journal on Dada Semantics, Springer (2013)
4. von-Ahn, L.: Games with a Purpose. IEEE Computer, vol. 39, no. 6, pp. 92-94, (2006)
5. Celino I., Contessa S., Corubolo M., Dell'Aglio D., Della Valle E., Fumeo S. and Krüger T.: Linking Smart Cities Datasets with Human Computation – the case of UrbanMatch. In: Proceeding of the 11th International Semantic Web Conference, Part II, Springer LNCS 7650, pp. 34–49 (2012)
6. Celino I., Cerizza D., Contessa S., Corubolo M., Dell'Aglio D., Della Valle E. and Fumeo S.: Urbanopoly – a Social and Location-based Game with a Purpose to Crowdsource your Urban Data. In: Proceedings of the the 4th IEEE International Conference on Social Computing, Workshop on Social Media for Human Computation, pp. 910-913, DOI: 10.1109/SocialCom-PASSAT.2012.138, (2012)
7. Heath T. and Bizer C. Linked Data: Evolving the Web into a Global Data Space, Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, first edition, (2011)
8. Lebo T., Sahoo S., McGuinness D., and eds. Prov-o: The prov ontology, http://www.w3.org/TR/2013/REC-prov-o-20130430/. Last accessed: 2013-05-10, (2013)
9. Magliacane S. Reconstructing provenance. In: The Semantic Web–ISWC 2012, pp. 399-406. Springer Berlin Heidelberg (2012)
10. Halpin H. Provenance: The missing component of the semantic web for privacy and trust. In: Trust and Privacy on the Social and Semantic Web (SPOT2009), workshop of ESWC, (2009)
11. Ceolin, D., Groth P.T., van Hage W.R., Nottamkandath A., and Fokkink W. Trust Evaluation through User Reputation and Provenance Analysis". In URSW, pp. 15-26 (2012)
12. De Nies, Tom, Sam Coppens, Erik Mannens, and Rik Van de Walle.: Modeling uncertain provenance and provenance of uncertainty in W3C PROV. In: Proceedings of the 22nd international conference on World Wide Web companion, pp. 167-168 (2013)