To trust, or not to trust: Highlighting the need for data provenance in mobile apps for smart cities

Mikel Emaldi
Deusto Institute of Technology
- DeustoTech
m.emaldi@deusto.es

Oscar Peña
Deusto Institute of Technology
- DeustoTech
oscar.pena@deusto.es

Jon Lázaro
Deusto Institute of Technology
- DeustoTech
jlazaro@deusto.es

Diego López-de-Ipiña
Deusto Institute of Technology
- DeustoTech
dipina@deusto.es

ABSTRACT

a

1. INTRODUCTION

According to the "Apps for Smart Cities Manifesto"[1], smart city applications could be sensible, connectable, accessible, ubiquitous, sociable, sharable and visible/augmented. It is not a coincidence that all of these features can be found in a standard smartphone: the popularity of these devices makes them the most suitable device to ensure access to the services provided by smart cities. As one of the main features of the smart cities is the participation of the citizens in their governance, it is not unusual that these citizens generate and share their own data through their smartphones. Reviewing the literature, we can find many examples of apps that deal with user generated data like Urbanopoly [2], Urbanmatch [3] or the popular apps related to the 311 service in cities like Calgary, Minneapolis, Baltimore or San Diego, all of them available in Google Play[2]. The IES Cities project goes one step beyond, providing an entire architecture to allow citizens to develop apps based on Linked Open Data [1] provided by government, through user-friendly JSON APIs. All of these works that manage user-generated data have the same worry about these data: are they reliable? How can we know if can a given user and, consequently, the data generated by him/her can be trusted? Recently, the W3C has created the PROV Data Model [6], for provenance interchange on the Web. This PROV Data Model describes the entities, activities and people involved in the creation of a piece of data, allowing the consumer to evaluate the reliability of the data based on the their

provenance information. Furthermore, PROV was deliberately kept extensible, allowing various extended concepts and custom attributes to be used. For example, the Uncertainty Provenance (UP) [4] set of attributes can be used to model the uncertainty of data, aggregated from heterogeneously divided trusted and untrusted sources, or with varying confidence. On this paper, we present how IES Cities platform integrates PROV Data Model and the related PROV-O ontology [5] and Uncertainty Provenance set of attributes, allowing the exchange of provenance information about user-generated data in the context of smart cities. The final aim is to enrich the knowledge gathered about a city not only with government-provided or networked sensors' provided data, but also with high quality and trustable data coming from the citizens themselves.

The remaining of the paper is organized as follows: in Section 2 the current state of the art on apps that deal with user data in the context of smart cities and on provenance data representation is presented. Section 3 outlines the main concepts about IES Cities project. Sections 4 and 5 describes the metrics to calculate the reliability of the data and its semantic representation, through a use cases, respectively. Finally, in Section 6 the conclusions and the future work are presented.

2. STATE OF THE ART

3. IES CITIES

4. PROVENANCE DATA BASED RELIABILITY

5. USE CASES

To illustrate the semantic representation of trust and provenance data through Provenance Ontology, a use case is presented: 311 Bilbao. 311 Bilbao, uses Linked Open Data to get an overview of reports of faults in public infrastructure. From the data owner's point of view, enrichment of their datasets by third parties, such as users of the 311 Bilbao application, revealed two problems: 1) the fact that data does not need to be approved before being published and that there is no mechanism to control the amount of data

---

[1] http://www.appsforsmartcities.com/?q=manifesto
[2] https://play.google.com

a citizen can add and 2) there is still the need for a way to differentiate the default trustworthiness of the different authors such as citizens and city councils. At the following triples, the representation of the provenance of a user generated report is shown:

```
1   @prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .
2   @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3   @prefix prov: <http://www.w3.org/ns/prov#> .
4   @prefix iesc: <http://www.iescities.org/ont#> .
5   @prefix :     <http://bilbao.iescities.org#> .
6
7   :report_23456
8       a prov:Entity;
9       prov:wasGeneratedBy    :reportActivity_23456;
10      prov:wasAttributedTo   :jdoe;
11      prov:invalidatedAtTime "2013-07-22T03:05:03Z";
12      prov:value             "The paper bin is broken";
13  .
14
15  :reportActivity_23456
16      a prov:Activity;
17      prov:startedAtTime     "2013-07-22T01:01:01Z";
18      prov:wasAssociatedWith :jdoe;
19      prov:endedAtTime       "2013-07-22T01:05:03Z";
20  .
21
22  :jdoe
23      a iesc:Citizen , prov:Person;
24      foaf:givenName  "John Doe";
25      foaf:mbox       <mailto:jdoe@example.org>;
26  .
27
28  :report_23457
29      a prov:Entity;
30      prov:wasAttributedTo :jane;
31      prov:value             "It is incorrect , another
32                              paper bin has replaced the
33                              old one, but 2 meters beyond";
34      prov:wasRevisionOf  :report_23456;
35  .
36
37  :jane
38      a iesc:Citizen , prov:Person;
39      foaf:givenName      "Jane";
40      foaf:mbox           <mailto:jane@example.org>
41      prov:actedOnBehalfOf :bilbao_city_council;
42  .
43
44  :bilbao_city_council
45      a iesc:Organization , prov:Organization;
46      foaf:name "Bilbao City Council";
47  .
```

On this piece of semantic information the `:report_23456` resource represents the report made by the user. This report is identified by its own and unique URL and provides information about the user that has made it and which activity that has generated this report (lines 14-27). The `:reportActivity_23456` shows details about the activity that generated the report, like when the user starts to reporting the issue and when has ended. At lines 22-26 the information about "John Doe", the user that reported the fault, can be seen. In the example given, another user, Jane (lines 37-42), has revised the report made by John (lines 28-35). As the `prov:actedOnBehalfOf` attribute represents, Jane is some kind of municipal worker of Bilbao City Council (lines 51-54). As Jane's report has more authority agains John's report, John's report is invalidated as property `prov:invalidatedAtTime` represents.

6. CONCLUSIONS AND FUTURE WORK

7. REFERENCES

[1] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.

[2] I. Celino, D. Cerizza, S. Contessa, M. Corubolo, D. DellAglio, E. D. Valle, and S. Fumeo. Urbanopoly–A social and location-based game with a purpose to crowdsource your urban data. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, page 910–913, 2012.

[3] I. Celino, S. Contessa, M. Corubolo, D. Dell'Aglio, E. Della Valle, S. Fumeo, T. Krüger, and T. Krüger. UrbanMatch-linking and improving smart cities data. In *LDOW*, 2012.

[4] T. De Nies, S. Coppens, E. Mannens, and R. Van de Walle. Modeling uncertain provenance and provenance of uncertainty in W3C PROV. In *22nd International World Wide Web Conference*, page 167–168, 2013.

[5] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, and J. Zhao. Prov-o: The prov ontology. *W3C Recommendation, http://www. w3. org/TR/prov-o/(accessed 30 Apr 2013)*, 2013.

[6] L. Moreau, P. Missier, K. Belhajjame, R. B'Far, J. Cheney, S. Coppens, S. Cresswell, Y. Gil, P. Groth, G. Klyne, et al. Prov-dm: The prov data model. *Candidate Recommendation*, 2012.