Citizen-centric Linked Data Services for Smarter Cities

Dr. Diego López-de-Ipiña
MORElab Research Group, DeustoTech – Deusto Institute of Technology, Faculty of Engineering, University of Deusto
dipina@deusto.es

06/02/2014 @ Fondazione Bruno Kessler, Trento
Introduction

• Smart Cities improve the efficiency and quality of the services provided by governing entities and business and (are supposed to) increase citizens’ quality of life within a city
  – BUT, do they really address the user needs?

• “The city must become like the Internet, i.e. enabling creative development and easy deployment of applications which aim to empower the citizen” - THE APPS FOR SMART CITIES MANIFESTO
  – This view can be achieved by leveraging:
    • Available infrastructure such as Open Government Data and deployed sensor networks in cities
    • Citizens’ participation through apps in their smartphones
Why Smarter Inclusive Cities?

• Not enough with the traditional resource efficiency approach of Smart City initiatives
  • “City appeal” will be key to attract and retain citizens, companies and tourists
  • Only possible by user-driven and centric innovation:
    – The citizen should be heard, EMPOWERED!
      » Urban apps to enhance the experience and interactions of the citizen, by taking advantage of the city infrastructure
    – The information generated by cities and citizens must be linked and processed
      » How do we correlate, link and exploit such humongous data for all stakeholders’ benefit?
        • We should start talking about Big (Linked) Data
Towards Smarter Inclusive Cities

• Generation of user-centric solutions that exploit urban data in different domains and thus improve the use of existing urban infrastructure and resources
  – A sustainable and open ecosystem vs. “smart city in a box” solutions proposed by corporations such as IBM or CISCO

Apps for Smart city

= Intelligent places + Data + Mobile apps + Community + Services
  = City as a platform

Research Aim

• “To create a multi-device dataset and service ecosystem based on standard web technologies, that exploits the data shared by councils and their citizens, and provides to citizens, tourists and workers an enhanced, close to their needs, experience in a municipality”
Research Objectives

1. Transform and release municipalities’ open data into Open Linked Data
2. Generate structured council data from non-structured data coming from social networks and users
3. Integrate, normalize, link and publish data for its massive analysis and exploitation by councils, enterprises and citizens
4. Generate a scalable Cloud-based and NoSQL-based infrastructure that supports the urban apps and datasets ecosystem
5. Create a set of urban apps over the normalized data of a council, its citizens and companies
6. Disseminate and transfer the resulting service and dataset management platform to companies and administrations
IES Cities Project

• The IES Cities project promotes user-centric mobile micro-services that exploit open data and generate user-supplied data
  – Hypothesis: Users may help on improving, extending and enriching the open data in which micro-services are based

• Its platform aims to:
  – Enable user supplied data to complement, enrich and enhance existing datasets about a city
  – Facilitate the generation of citizen-centric apps that exploit urban data in different domains
IES Cities Stakeholders

• The main stakeholders of the resulting urban apps ecosystem by IES Cities envisaged smart city-enabling platform are:
  – Mainly the citizens as final users and app idea innovators
  – SMEs and public administration of different cities who satisfy the social and economic needs detected
LinkedData

• “A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF”

• Allows to **discover, connect, describe and reuse all types of data**
  – Enables to pass from a Web of Documents to a Web of Data
    • In September 2011, it had 31 billion RDF triples linked by 505 million links

• Thought to open and connect diverse vocabularies and semantic instances, to be used by the semantic community

• URL: [http://linkeddata.org/](http://linkeddata.org/)
Types of Linked Data

- Data published as LinkedData can follow the following **classification** according to Tim Bernes-Lee:
  - **1 star**: data available in the web (in any format), but with an open license
  - **2 stars**: data available is structured and readable by machines. For instance, Microsoft Excel instead of an image scanned from a table
  - **3 stars**: data available as in (2*) but they do not follow a proprietary format. For instance, CSV in Excel.
  - **4 stars**: data available in open manner using an W3C open standard (RDF & SPARQL) to identify things, so that people can link them
  - **5 stars**: data available following (4*) principles, including external links to data of other people
4 rules of Linked Data

1. Use URIs to identify things
2. Use HTTP URIs so that things can be referenced and unreferenced by people and user agents
3. Provide useful info (structured description and metadata) about the things/concepts referenced by the URI
4. Include links to other URIs to improve the discovery of related information in the Web
Mechanisms for Supporting Citizens & Developers

• **Provenance tracking mechanisms** to assess and qualify user-provided data, thus promoting valuable and trustable information and decrementing and eventually discarding lower quality data
  – W3C PROV Data Model for provenance exchange on Web

• **Human Computation** enables to leverage human intelligence to carry out tasks that otherwise would be difficult to accomplish by a machine
  – Gamification can also be used to incentivize citizen participation

• **JSON schema and query languages** to facilitate urban apps development
  – Structured and non-structured data in the form of RDF, CSV or even HTML pages can be easily mapped into JSON
1st Iteration Architecture
User-provided Data

- Smart Cities seek the participation of citizens:
  - 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
  - BUT, how can we know if a given user and, consequently, the data generated by him/her can be trusted?
    - W3C has created the PROV Data Model, for provenance interchange
Problems associated to User-provided Data

• The impact that citizens may have on improving, extending and enriching the data enabled services will be based upon
  – Quality of the provided data may vary from one citizen to another, not to mention the possibility of someone's interest in populating the system with fake data
  – Duplication, miss-classification, mismatching and data enrichment issues
Need to Measure Data Quality

• Need for evaluating the value and trust of the user contributed data requires a validation module:
  – Aids have to be provided before and after editing new entries and a two phase commit process for user provided data should be put in place.
  – A metric to measure data trustworthiness level, providing an extra confidence layer.
Provenance Semantic Representation

```xml
1 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
2 @prefix prov: <http://www.w3.org/ns/prov#> .
3 @prefix iesc: <http://studwww.ugent.be/~satvheck/IES/schemas/iescities.owl> .
4 @prefix up: <http://users.ugent.be/~tdenies/up/> .
5 @prefix : <http://bilbao.iescities.org#> .

8 entity (:report_23456, [prov: value="The paper bin is 9 broken"])
9 wasGeneratedBy (:report_23456, :reportActivity_23456)
10 wasAttributedTo (:report_23456, :jdoe)
11 wasInvalidatedBy (:report_23456, :invActivity_639, 2013-07-22T03:05:03)
12 activity (:reportActivity_23456, 2013-07-22T01:01:01, 16 2013-07-22T01:05:03)
14 wasAssociatedWith (:reportActivity_23456, :jdoe)
16 agent (:jdoe, [prov:type='prov:Person', foaf:name='John Doe', foaf:mbox='mailto:jdoe@example.org'])
18 entity (:report_23457, [prov:value="It is incorrect, another paper bin has replaced the old one, but 2 meters beyond "])
19 wasAttributedTo (:report_23457, :jane)
20 wasDerivedFrom (:report_23457, :report_23456, invActivity_639, -, -, [prov:type='prov:Revision'])
23 wasAssociatedWith (:invActivity_639, :jane)
25 agent (:jane, [prov:type='prov:Person', foaf:name='Jane', foaf:mbox='mailto:jane@bilbao.iescities.org'])
26 actedOnBehalfOf (:jane, :bilbao_city_concil)
28 agent (:bilbao_city_concil, [prov:type='prov:Organization', foaf:name='Bilbao City Council'])
```
Trust Metric

- Trust metric depends on **authority** (who creates it), **popularity** (number of references received), **recommendation** (votes), **reputation** of author, recency, geographical distance

\[
trust(report) = \frac{\sum_{p=[\text{auth, agree...}]}^{n} \alpha_p \cdot trust_p(report)}{n}
\]  

\[
trust_{\text{authority}} = \begin{cases} 0 & \text{if user } \neq \text{ authority} \\ 1 & \text{if user } = \text{ authority} \end{cases}
\]  

\[
trust_{\text{popularity}} = \frac{\text{visits}_\text{report}}{\text{visits}_\text{open reports}}
\]  

\[
trust_{\text{recommendation}} = \frac{\text{positive votes}_\text{report}}{\text{total votes}_\text{report}}
\]  

\[
trust_{\text{authority}} = \left(\max\left(1 - \frac{\text{currency}}{\text{volatility}}\right), 0\right)^{\text{sensitivity}}
\]

\[
trust_{\text{distance}} = \frac{1}{\text{geodistance}(\text{loc}_\text{report}, \text{loc}_\text{reported place})}
\]

- The end result is to generate annotations such as: 
  :report_23456 up:contentCondence '0.6'
Support for Developers: RDF to Relational

• There is a need to make the development of urban-apps based on Linked Data more straightforward → accessible to any web developer

• Enable SQL on SPARQL endpoints
  – Avoid complexity of SPARQL language
  – Reduce learning using prior SQL knowledge

• Transformation process:
  1. Define a mapping from Classes/Properties to Tables/Relations
  2. Convert \texttt{SELECT} to equivalent SPARQL query
  3. Execute query and convert results to SQL resultset.
RDF to Relational

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select (str(?b) as ?id) (str(?a) as ?hwa_timestamp) (str(?c) as ?hwa_externalTemperature) (str(?d) as ?hwa_internalTemperature) (str(?e) as ?hwa_luminosity) (str(?f) as ?hwa_status) from <http://agronutas> where {
  optional {?b <http://www.morelab.deusto.es/agronutasSimple.owl#timestamp> ?a .}
  optional {?b <http://www.morelab.deusto.es/agronutasSimple.owl#externalTemperature> ?c .}
  optional {?b <http://www.morelab.deusto.es/agronutasSimple.owl#internalTemperature> ?d .}
  optional {?b <http://www.morelab.deusto.es/agronutasSimple.owl#luminosity> ?e .}
  optional {?b <http://www.morelab.deusto.es/agronutasSimple.owl#status> ?f .}
  filter (
    xsd:int(?a) > xsd:int(23)
  ) .
}
```
IES Cities Player

- Web application wrapped in PhoneGap
- HTTP messages to ‘/service/’ REST interface
Platform Validation

• 311 Bilbao
  – Query reports about complaints concerning public infrastructure
  – Generate own reports → prosuming

• AirQual Madrid
  – Display near real-time information on air quality measurements across Madrid.
IES Cities Apps: 311 Bilbao
IES Cities Apps: AirQual Madrid

- non-RDF Open Data

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estacion</td>
<td>PM10</td>
<td>PM2</td>
<td>SO2</td>
</tr>
<tr>
<td>PLAZA ESPANA</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>ESCUELAS AGUIRRE</td>
<td>19</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>AVDA. RAMON Y CAJAL</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARTURO SORIA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VILLAVEDE</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>FAROLILLO</td>
<td>15</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CASA DE CAMPO</td>
<td>13</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>BARAJAS PUEBLO</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PLAZA DEL CARMEN</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>MORATALAZ</td>
<td>22</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CUATRO CAMINOS</td>
<td>15</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>BARRIO DEL PILAR</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Conclusions

• Capital to **include citizens in the Smart City innovation loop** and in the enrichment of the city knowledge with their data contribution
  – Only way to progress towards **Smarter (Inclusive) Cities**

• IES Cities aims to address this by offering:
  – Architecture enforcing **usability, interoperability, modifiability, scalability** and **portability** ...

• Added value for public bodies, developers and users
  – No need for republishing existing datasets
  – **REST** interfaces and **generic queries (SQL-based)** for intuitive development of IES Cities Services
  – Semantic technologies to support the **generation and validation** of Linked Open Data
Citizen-centric Linked Data Services for Smarter Cities

Dr. Diego López-de-Ipiña
MORElab Research Group, DeustoTech – Deusto Institute of Technology, Faculty of Engineering, University of Deusto
dipina@deusto.es
http://paginaspersonales.deusto.es/dipina
Citizen-centric Linked Data Services for Smarter Cities

Dr. Diego López-de-Ipiña
MORElab Research Group, DeustoTech – Deusto Institute of Technology,
Faculty of Engineering, University of Deusto
dipina@deusto.es
http://paginaspersonales.deusto.es/dipina
Citizen-centric Linked Data Services for Smarter Cities

Dr. Diego López-de-Ipiña
MORElab Research Group, DeustoTech – Deusto Institute of Technology,
Faculty of Engineering, University of Deusto
dipina@deusto.es

06/02/2014 @ Fondazione Bruno Kessler, Trento