FarolApp: Live Linked Data on Light Pollution

Nandana Mihindukulasooriya, Esteban Gonzalez, Fernando Serena, Carlos
Badenes, and Oscar Corcho⋆

Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
{nmihindu,egonzalez,fserena,cbadenes,ocorcho}@fi.upm.es

Abstract. FarolApp is a mobile web application that aims to increase
the awareness of light pollution by generating illustrative maps for cities
and by encouraging citizens and public administrations to provide street
light information in an ubiquitous and interactive way using online street
views. In addition to the maps, FarolApp builds on existing sources to
generate and provide up-to-date data by crowdsourced user annotations.
Generated data is available as dereferenceable Linked Data resources in
several RDF formats and via a queryable SPARQL endpoint. We pro-
pose Live Linked Data, a new approach to publish data about city infras-
tructures trying to keep them synchronized leveraging the collaboration
of citizens. The demo presented in this paper illustrates how FarolApp
maintains continuously evolving Linked Data that reflect the current sta-
tus of city street light infrastructures and use that data to generate light
pollution maps.

Keywords: linked data, evolution, crowdsourcing, light pollution

1   Introduction

Light pollution is one of the most unknown and rapidly increasing environmental
problems nowadays[1]. Artificial lighting is the main cause of the excessive level
of light pollution, leading to several problems affecting animal species, human
health and energy consumption. In cities, the street lighting infrastructures are
the biggest known polluters.

A starting point to reduce light pollution is to increase public awareness
about this type of pollution by providing the relevant data in a human consum-
able manner. Public institutions may implement wellness policies tailored to the
current situation and provide open data that may help getting a deeper under-
standing of this problem. However, to be usable, open data does not only have to
be accessible, but also be offered in a comprehensible way. Linked Data allows to
explicitly describe the meaning of data by using a vocabulary and linking them
to other data sources.

Cities are living organisms in constant evolution. *Live Linked Data*, i.e. con-
tinuously maintained and updated data, offers a more realistic representation of

a city rather than traditional static RDF dumps. However, in order to do this, someone has to be aware of the changes that occur in the city and reflect them in the data. Citizen science projects could bring some light into this situation.

FarolApp (http://farolapp.linkeddata.es/) is a Linked Data based application whose goal is to increase the awareness of light pollution among citizens and public administrations, presenting pollution maps and allowing citizens to contribute to that information. Citizens can participate both on-site or remotely using Google Street View. In this demo, we present the approach followed by FarolApp and its architecture. A video[1] also is available on the project website.

2   Approach

Linked Data Generation FarolApp is seeded with existing heterogeneous datasets. On the one hand, if the existing data is already in RDF, its vocabulary is mapped to the FarolApp Vocabulary[2] to create a new dataset by using SPARQL construct queries. On the other hand, if the existing data is not in RDF (e.g, CSV, JSON), it is transformed to RDF using external tools such as OpenRefine. Initially, FarolApp integrates data from ten different cities and the seed data is available in datahub[3].

Furthermore, during the generation process the data is linked to other relevant datasets. For instance, each street light is associated to the city and the country where it is located based on its GPS coordinates and linked to the corresponding entities in DBpedia. Provenance information are also provided to attribute the data with their original data sources.

Linked Data Publication The main interface of FarolApp is a web UI in which the street light data is overlaid on a map. Furthermore, the app provides pollution maps which are generated based on application data as well as an API where information about street lights can be extracted. Also, this information is published in Social Networks such as Twitter. All these representations use the Linked Data created and maintained by FarolApp as their data source.

In addition, FarolApp provides raw Linked Data as dereferenceable resources (supporting several RDF formats such as RDF/XML, Turtle, JSON-LD using HTTP content negotiation), RDF dumps, and via a SPARQL endpoint, enabling other third-party applications to reuse its up-to-date information. For instance, interlinking street light data with data about energy consumption, traffic accidents, or street crimes may reveal interesting findings.

Linked Data Evolution With the aim of keeping the data inline with the current state of the street lights infrastructure, it is necessary to get live information from any kind of sensors that are able to communicate their observations. Humans are the most reliable sensors in this scenario, mainly because there is no

---

[1] http://liveldp.github.io/demo
[2] http://goo.gl/5n8zEs
[3] https://datahub.io/dataset/farolapp-dataset

common data source, neither a sensor network (e.g. an IoT platform) to date suitable to build on for: i) identifying street lights whose data is not yet openly available, ii) updating the most interesting attributes of all of them. Therefore, this approach leverages crowdsourcing techniques [2] in order to extend, update and ultimately evolve the initial datasets.

Citizens can contribute by giving multiple value annotations for attributes of street lights. The range and type of values depend on the nature of the attribute. In some cases, values expected to be discrete to simplify annotations. For instance, the height attribute, which is normally a real-value, is converted to a categorical value such as low, medium and high.

FarolApp follows a consensus-based approach to conclude a satisfactory level of agreement from all annotations. Two main basic rules support consensus detection: i) a minimum number of annotations is required, ii) the standard deviation must be lower than an attribute-specific threshold. When both rules are satisfied, FarolApp transforms the agreed value into Linked Data format and enriches the dataset making it evolve.

3 High-level architecture

The Farolapp architecture (fig. 1) follows a *Staged Event-Driven Architecture* (SEDA) [3] that decomposes the flow into a set of components connected by an event-bus, offering some important advantages over traditional architectures: i) *Dynamic processing flows* created by the modules connected to queues, ii) *Isolated and distributed environments* where the modules can be implemented in different languages, iii) a *leaky bucket approach* handled by circular queues.
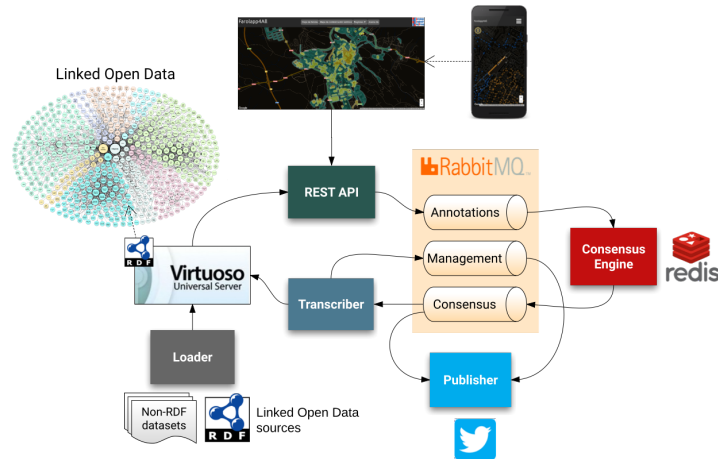

Fig. 1. The FarolApp architecture.

Web/Mobile UIs allow users to navigate and annotate the street lights information in an intuitive way using Google Maps. The REST API provides JSON-based descriptions about individual or clusters of street lights. The Loader bootstraps the application with the existing street light data. The Consensus Engine detects and notifies about consensus from annotations of street lights' attributes. The Publisher is in charge of publishing relevant events from FarolApp in social media platforms such as Twitter. The Transcriber persists the updates generated from the consensus engine based on the user annotations.

In addition, FarolApp makes use of a Virtuoso server as a triplestore for providing a SPARQL endpoint, RabbitMQ as a messaging broker, and Redis as an in-memory cache for annotations.

4   Related Work

Inspired by the DBpedia Live[4] approach, which transforms information from Wikipedia into Linked Data, FarolApp collects data from citizens, transforming them into Live Linked Data. Other approaches, such as Zooniverse[5] (a citizen science platform for crowdsourcing scientific research) or OpenStreetMap[6] (OSM, a collaborative platform to create a free editable map of the world) are not directly available as semantically enriched Linked Data. LOSM[4], which allows to query OSM data using SPARQL queries, is one step in that direction.

5   Conclusion and future work

FarolApp leverages Live Linked Data and serves as a proof of concept of such approach to provide up-to-date street light information and light pollution maps.

As future works we plan to integrate with IoT platforms from smart cities initiatives (e.g. photometers), keep track and offer historical data to study its evolution, and use OSM both as a data source and publisher. Besides, the UI will follow a more educational approach and provide multilingual support.

References

1. Falchi, F., Cinzano, P., Elvidge, C.D., Keith, D.M., Haim, A.: Limiting the impact of light pollution on human health, environment and stellar visibility. Journal of environmental management **92**(10) (2011) 2714–2722
2. Estellés-Arolas, E., González-Ladrón-de Guevara, F.: Towards an integrated crowdsourcing definition. Journal of Information Science **38**(2) (2012) 189–200
3. Welsh, M.: The staged event-driven architecture for highly-concurrent server applications. University of California, Berkeley (2000)
4. Anelli, V.W., Di Noia, T., Galeone, P., Nocera, F., Rosati, J., Tomeo, P., Di Sciascio, E.: LOSM: a SPARQL endpoint to query Open Street Map. In: Proceedings of the ISWC 2015 Posters & Demonstrations Track, Bethlehem, USA (October 2015)

---

[4] http://live.dbpedia.org/

[5] https://www.zooniverse.org/

[6] https://www.openstreetmap.org/