

# PCC2018-Bot: A Telegram bot for “Palermo Capitale della Cultura 2018” events powered by Linked Open Data and Schema.org annotations

A. Lo Bue<sup>1</sup>, A. Machi<sup>2</sup>, D. Taibi<sup>3</sup>

<sup>1</sup> ICAR-CNR, Palermo, Italy [icar.cnr.it](mailto:icar.cnr.it)  
[ing.antonino.lobue@gmail.com](mailto:ing.antonino.lobue@gmail.com)

<sup>2</sup> ICAR-CNR, Palermo, Italy  
[alberto.machi@icar.cnr.it](mailto:alberto.machi@icar.cnr.it)

<sup>3</sup> ITD-CNR, Palermo, Italy  
[davide.taibi@itd.cnr.it](mailto:davide.taibi@itd.cnr.it)

**Abstract.** This paper describes a practice for live social reuse of Schema.org annotations and Linked Open Data in the realm of events. The events of the “Palermo Capitale della Cultura 2018” initiative of the Ministry of Culture of Italy, were semantically enhanced by interlinking the available open data with related information inferred from the Linked Open Data cloud (namely DBpedia and Geonames). The resulting dataset - stored as an RDF graph within a triplestore was exposed via API in a knowledge graph by using the Schema.org vocabulary. The dataset was made also accessible to an event search assistant for tourists implemented as a bot for the instant messaging application Telegram. This effort shows how plain open metadata can be powered by Linked Data and semantic vocabularies like Schema.org, to become rich machine-understandable descriptions usable by automatic bots to provide improved question answering experiences for the social user.

**Keywords:** Semantic Web, Data Integration, Linked Open Data, Telegram bot.

## 1 Introduction

### 1.1 Linked Open data

The evolution of the Web is strictly connected to the way users interact with it. Nowadays, the potential users of web data are not only human beings but also software services and software agents. For this reason data should be published on the web using standards and technologies which can be understood and elaborated automatically.

At present, the most popular Web applications, such as Facebook and Youtube, offer Application Program Interfaces (APIs) that allow software agents to access the information they host. Semantic Web technologies provide an adequate technological substrate for supporting the representation of concepts and the relationships between them through ontologies, and the recent evolution of Linked Open Data is the natural way to publish, integrate, and link data semantically described. The information

available on the web uses different typologies and is published in heterogeneous formats. Linked Open Data (LOD) aims to provide a technological substrate for publishing structural data in a standardized format. The advantages of such an approach are tangible and it is increasingly common for data on the web to be published following the LOD principles [1]. While the linking of pages has marked the success of the Web, at the same time LOD aims to connect datasets and the concepts they host, providing information not only for humans but also for software agents.

SPARQL Query Language<sup>1</sup> is the standardized language to query and retrieve Semantic Web data stored as RDF<sup>2</sup> triples thus allowing and facilitating access to LOD resources. DBpedia<sup>3</sup>, can be seen as the semantic version of Wikipedia; it is the core of the Linked Open Data cloud, and provides a main access point for semantic enrichment.

## 1.2 Schema.org and bots

Schema.org was launched in 2011, as a result of a joint effort of the big players in the search engine field: Google, Microsoft, Yahoo!, and Yandex, with the aim of defining a shared vocabulary on common concepts of the real world.

Starting from the top-level concept, represented by the most generic type named *Thing*, main sub-concepts have been defined to represent concepts related to *CreativeWork*, *Organization*, *Person*, *Place*, *Product* and *Event*. Moreover, specific sub-types have been defined to represent concepts in different popular domains such as medicine<sup>4</sup>, or education [8].

As detected by the Web Data Commons<sup>5</sup> project, nowadays, the adoption of schema.org is more than 39% of all Web pages.

Recent studies report that over 2.5 billion people have installed an instant messaging app in its mobile phone, and already in 2015, interactions between people on the Web were mediated by instant messaging apps more than social network. More popular instant messaging platforms are: WhatsApp, Telegram, Viber [3].

Amongst them, Telegram is cross-platform and provides an appropriate API for building chat-bots to interact with a user or with a group of users. [4, 5].

## 2 Providing enriched contents on cultural events via bots

“Palermo Capitale della Cultura 2018” is one of the initiatives of the Ministry of Culture of Italy supporting coordination of cultural events in order to promote tourism in a chosen city. This paper describes a bot developed for instant messaging Telegram platform is presented to provide and share rich information about the initiative events. The bot enrich sparse open event descriptions with LOD data and provides the social

---

<sup>1</sup> <https://www.w3.org/TR/rdf-sparql-query/>

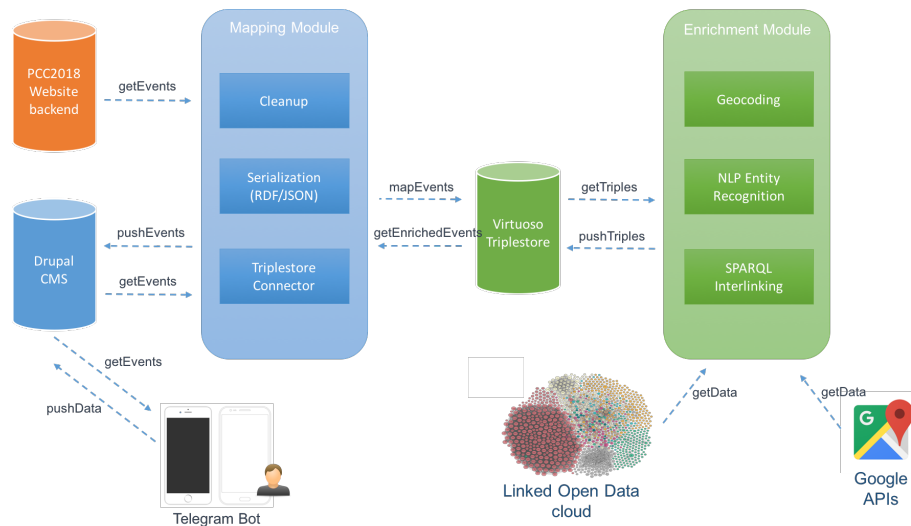
<sup>2</sup> <https://www.w3.org/RDF/>

<sup>3</sup> <http://wiki.dbpedia.org>

<sup>4</sup> <http://schema.org/docs/meddocs.html>

<sup>5</sup> <http://webdatacommons.org>

user an easy to use interface to browse them. The interface hides the complexity of queries required to semantically build useful descriptions according to user spatio-temporal context.



**Fig. 1. System overall architecture and data calls.**

Figure 1 shows the overall system architecture. The system serializes information about events in a knowledge graph containing not only the sparse open data published in the official website<sup>6</sup>, but also enriched entities extracted from the Linked Data silos that were interlinked as described in Section 2.1. This approach, based on knowledge graph and enriched entities, allows users to obtain additional information that were not included in the original website or in all the other services and mobile apps that are based on it.

Data extracted from the PCC2018 web site are firstly imported in a Drupal CMS in order to perform lexical cleaning of data and to improve improve efficiency of user query .

A mapping module translates CMS data into triples and stores them on a Virtuoso RDF triple-store. A interlinking module implemented via web-services enriches the resulting knowledge graph. A SPARQL endpoint answers semantic queries on the graph.

Semantically enriched event descriptions are then reimported and delivered to a Telegram bot through an API endpoint supporting field selection and range queries on temporal and spatial data.

<sup>6</sup> <http://palermocapitalecultura.it/>

## 2.1 Semantic Enrichment

Semantic Enrichment is a term used to describe the process of transforming plain data into structured data that contains machine-readable statements. This enrichment can happen using ontologies or taxonomies of controlled terms with semantics defined by the data owner or, in the context of Linked Open Data principles, reusing machine-understandable vocabularies with metadata values defined by external data providers as Europeana<sup>7</sup> or DBpedia<sup>8</sup>). Knowledge graphs published on the LOD cloud can be traversed to extract references or descriptions of related entities.

The main issue with semantic enrichment lies in the way to automate the process, in order to apply the enrichment to large volumes of data, instead of using manual domain expert annotation. Interlinking rules, distance measure algorithms as well as natural language processing techniques can support automated enrichment processes and generate well-formed semantic data that exploit the LOD cloud [7, 8].

In the context of this work, data enrichment was implemented using a mixed approach including programmatic tagging via external services and federated SPARQL queries to provide interlinking enrichment [6].

To semantically enrich event data, three specific enrichment techniques were used:

- Semantic named-entity recognition
- Geocoding enrichment
- Spatial interlinking

The interlinking module developed in Python, implementing appropriate connector interfaces between the external services and the triple-store.

For example, the text referring the locality where the event happens was geocoded using Google Geocode APIs<sup>9</sup>, then, expressed using Schema.org relations (Address, Administrative Areas, Latitude/Longitude coordinates) and finally reconverted in plain text address of the event place for sake of simplicity.

For the recognition of named entities the textual contents of the “title” and “description” fields of each events was sent as input text to DBpedia Spotlight APIs<sup>10</sup>. DBpedia Spotlight [9] is a tool for automatically annotating mentions of DBpedia resources in texts. by means. As output we received from the service, for each event, an array of related "DBpedia intities" expressed as *rdfs:seeAlso*<sup>11</sup> statements.

The geocoding and named-entity inferred triples were used as source data for the third type of interlinking, exploited via SPARQL federated queries<sup>12</sup> to merge facts about the same event extracted from different sources of the Linked Open Data cloud. In particular, the SPARQL query implemented, infers from DBpedia nearby entities (places, historical monuments, etc). A threshold to the Haversine geospatial distance from the Event place coordinates was used to define the effective region of interest around the event place.

---

<sup>7</sup> <https://pro.europeana.eu/page/linked-open-data>

<sup>8</sup> <https://wiki.dbpedia.org>

<sup>9</sup> <https://developers.google.com/maps/documentation/geocoding/intro>

<sup>10</sup> <http://spotlight.dbpedia.org>

<sup>11</sup> [https://www.w3.org/TR/rdf-schema/#ch\\_seealso](https://www.w3.org/TR/rdf-schema/#ch_seealso)

<sup>12</sup> <https://www.w3.org/TR/sparql11-federated-query/>

The semantically enriched events stored in the triple-store, were mapped back to the application main CMS, and served to the telegram BOT via the CMS output API endpoint.

Table1 shows the semantics and formats of event descriptors at various steps of the enrichment chain.

## 2.2 The Telegram bot

The Telegram Bot presented in this paper was designed to guide users in searching for events organized in the framework of the “Palermo Capitale della Cultura 2018” initiative. The interface facilitates the searching process by simplifying users’ interactions in order to provide access to information related to events matching their interest in a minimum number of clicks. In particular, customized keyboards were designed to help users in selecting straightforwardly the most commonly used search options.

The customized keyboard shown by the bot allows users to search along three specific dimensions: temporal, spatial and categorical (Fig 2a).

The spatial dimension allows users to search events in the nearby of their current position. Telegram Bot API supports the possibility to transmit to the bot the user coordinates, by using the *request\_location* parameter associated to the keyboard button. After selecting the options along the three dimensions, a list of events matching user’s preference is shown as selection list (Fig 2b).

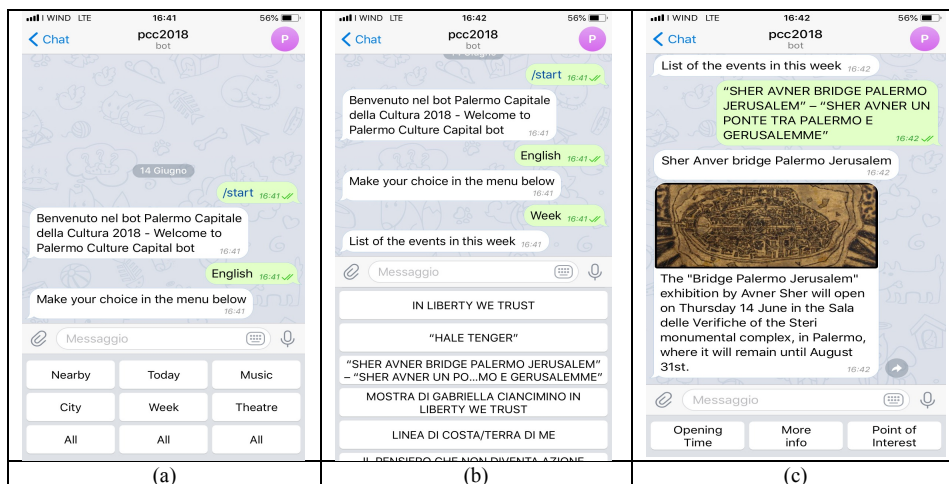


Figure 2: Bot Telegram User Interface

User can select from the list an event in order to access event description. In fact, for each event the bot answers with a message containing the title, an image related to the event and a brief description. Moreover, the bot can provide additional content related to opening hours, description details and a list of and nearby points of interest.

The enriched knowledge graph is used to provide this additional information. Specifically, the *rdf:seeAlso* property is used to provide detailed information to the enti-

ties related to the event, while the *geonames:nearby* property is used to provide information about the points of interest located in the neighborhood of the place in which the event takes place.

**Table 1: Semantic mapping and data formats of fields at various steps of the enrichment chain.**

<b>CMS input API</b>	<b>CMS to triplestore mapped graph</b>	<b>triplestore enriched graph</b>	<b>CMS output API</b>
event ( <i>object</i> )	schema:Event ( <i>rdf:resource</i> )	schema:Event ( <i>rdf:resource</i> )	EnrichedEvent( <i>object</i> )
placeName( <i>text</i> )	schema:location ( <i>rdf:resource</i> )	schema:Location ( <i>rdf:resource</i> )	placeName ( <i>text</i> )
--	--	schema:Address ( <i>rdf:resource</i> )	--
--	--	schema:geo ( <i>rdf:resource</i> )	plainAddress ( <i>text</i> )
--	--	schema:latitude ( <i>rdfs:Literal</i> )	longitude ( <i>float</i> )
--	--	schema:longitude ( <i>rdfs:Literal</i> )	latitude( <i>float</i> )
--	--	schema:streetAddress ( <i>rdfs:Literal</i> )	
--	--	schema:postalCode ( <i>rdfs:Literal</i> )	
--	--	schma:addressLocality ( <i>rdfs:Literal</i> )	
description( <i>html</i> )	schema:description <i>rdfs:Literal</i>	schema:description ( <i>rdfs:Literal</i> )	description( <i>html</i> )
imageUrl ( <i>url</i> )	schema:image ( <i>xml:Uri</i> )	schema:image ( <i>xml:Uri</i> )	imageUrl ( <i>url</i> )
category ( <i>text</i> )	temp:category ( <i>rdf:Literal</i> )	temp:category ( <i>rdfs:Literal</i> )	category( <i>text</i> )
organization( <i>text</i> )	schema:organizer ( <i>rdf:resource</i> )	schema:organizer ( <i>rdf:resource</i> )	organizationName( <i>text</i> )
--	--	rdfs:seeAlso ( <i>rdf:resource</i> )	more ( <i>url</i> )
--	--	geonames:nearby ( <i>rdf:resource</i> )	near to ( <i>url</i> )

### 3 Conclusions and future work:

Descriptions of more than about 200 events (plus replicas) in the framework of the initiative “*Palermo European Capitale della Cultura 2018*” were enriched with full geo-location information, completed with images related to location or named subject, and with references to Linked Data entities, extracted via Named-Entity recognition and via geospatial inference rules. A compact semantic search GUI was provided through a Telegram bot to allow users to easily search and share information.

This effort shows how plain metadata (and Open Data) can be powered by Linked Data and expressed through semantic vocabularies like Schema.org, to become rich-machine understandable data used unambiguously by automatic bots to provide improved question answering experiences for the social user.

Topics of current research are the enhancement of the enrichment inference engine and the upgrade of the bot to a conversational bot (chat-bot).

### References

1. Auer, S., The emerging web of linked data. Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications, (2011).
2. R. V. Guha, Dan Brickley, and Steve Macbeth. 2016. Schema.org: evolution of structured data on the web. Commun. ACM 59, 2 (January 2016), 44-51. DOI: <https://doi.org/10.1145/2844544>
3. Sutikno T, Handayani L, Stiawan D, Riyadi M A, Much I and Subroto I 2016 WhatsApp, Viber and Telegram: which is the best for instant messaging? Int. J. of Electrical and Computer Eng. (IJECE) 6 909-14 <http://doi.org/10.11591/ijece.v6i3.10271>
4. Pereira J. 2016. Leveraging chatbots to improve self-guided learning through conversational quizzes. In Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '16), Francisco José García-Peñalvo (Ed.). ACM, NY, USA, 911-918. DOI: <https://doi.org/10.1145/3012430.3012625>
5. Dietze S, Taibi D, Yu R, Barker P, and d'Aquin M. 2017. Analysing and Improving Embedded Markup of Learning Resources on the Web. In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 283-292. DOI: <https://doi.org/10.1145/3041021.3054160>
6. Lo Bue A., Machi A., Open Data Integration using SPARQL and SPIN - A Case Study for the Tourism Domain. AI\*IA 2015: Artificial Intelligence and Human-Oriented Computing, September 23-25 Ferrara (2015), Italy. LNCS 9336 pp 316-326
7. Lo Bue, A., Wecker, A. J., Kuflik, T., Machi, A., & Stock, O. (2015). Providing Personalized Cultural Heritage Information for the Smart Region-A Proposed Methodology. In UMAP Workshops.
8. Simou N., Chortaras A., Stamou G., Kollias S. (2017) Enriching and Publishing Cultural Heritage as Linked Open Data. In: Ioannides M., Magnenat-Thalmann N., Papagiannakis G. (eds) Mixed Reality and Gamification for Cultural Heritage. Springer, Cham
9. Pablo Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia spotlight: shedding light on the web of documents. In Proceedings of the 7th International Conference on Semantic Systems, Chiara Ghidini, Axel-Cyrille Ngonga Ngomo, Stefanie Lindstaedt, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA, 1-8.