Methodology for the publication of Linked Open Data from small and medium size DMOs

Ander García[a], Maria Teresa Linaza[a], Javier Franco[a] and Miriam Juaristi[a]

[a]Department of eTourism and Cultural Heritage
Vicomtech-IK4, Spain
{agarcia, mtlinaza, jfranco, mjuaristi}@vicomtech.org

Abstract

Destination Management Organizations (DMOs) create and collect valuable data in the tourism sector, which usually is kept in isolated data repositories. Linked Open Data (LOD), which combines the Open Data movement and the Linked Data initiative, adds semantic annotations and links to external data, allowing DMOs not only setting their data free from isolated repositories, but also placing them in multiple contexts by pairing them with different LOD sets worldwide. However, the lack of clear methodologies, publication examples and tools focused on DMOs has been reflected on a scarcity of available examples of tourism LOD published by DMOs. This paper presents a methodology for the publication of tourism LOD for small and medium size DMOs, which has been implemented using Open Source tools. An example of publishing a multilingual dataset of Points of Interest (POIs) is provided as well as a mobile application based on the data.

Keywords: Linked Open Data; DMO; tourism data.

1 Introduction

Destination Management Organizations (DMO) are among the largest creators and collectors of data in the tourism sector. Such information is usually kept in closed legacy systems and isolated repositories (data silos), where it is difficult to find, access, combine and reuse. However, there are different approaches to turn tourism information that has been created and used only for internal purposes into valuable and informative datasets.

The Open Data (OD) philosophy is based on the general idea that data should be open, which means that "anyone is free to use, reuse, and redistribute it, subject only, at most, to the requirements to attribute and/or share-alike" (opendefinition.org [July 25, 2014]). The publication of updated and reliable OD by DMOs opens new possibilities, such as the reduction of the development costs of applications (mobile guides, websites) on top of the data, or the provision of innovative added-value services which combine multiple data sources. As a further step, the semantic annotation and linking of OD to other data under the Linked Open Data (LOD) initiative will allow DMOs to improve their publication initiatives. Not only will they be able to promote the access to the data, but also they will pair themselves with different LOD sets worldwide.

Generally speaking, a limited number of anecdotal datasets are used to provide examples of tourism LOD, as the publication process has not happened in a standardised way in most of the cases.

Despite the current availability of a fragmented set of tools to support Linked Data efforts, the lack of integrated Linked (Open) Data transformation and publication frameworks is still a barrier for DMOs, which demand solutions that could operate under their scale, heterogeneity and data quality standards.

To address some of these critical issues, this paper proposes a methodology for the publication of LOD based on Open Source tools mainly oriented to small and medium size DMOs. The methodology has been validated during the LOD publication of several Points of Interest (POIs) from a local DMO, which have been reused in the implementation of a mobile application.

This paper has been organised as follows. Section 2 reviews the state of the art of LOD and its application to the tourism domain. Section 3 describes the proposed methodology and its implementation based on Open Source tools. Section 4 presents an example of the publication of tourism LOD, and a mobile prototype reusing the data. Finally, Section 5 summarises the conclusions and proposes some future work.

2 State of the art

2.1 Linked Open Data (LOD)

On the one hand, the availability of Open Data (OD) has grown significantly, with pressure being placed on all kinds of public organizations to release their raw data. Although OD is often crucial for the development of public policy, it can also be valuable for other domains such as tourism. This research will follow the definition of OD provided by Janssen, Charalabidis and Zuiderwijk (2012) who defined Open Data as "non-privacy-restricted and non-confidential data which is produced with public money and is made available without any restrictions on its usage or distribution. Data can be provided by public and private organizations, as the essence is that the data is funded by public money".

On the other hand, Linked Data (LD) refers to "data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external datasets, and it can in turn be linked to from external datasets" (Bizer, Heath and Berners-Lee, 2009). LD identifies resources with unique Uniform Resource Identifiers (URIs) references, which can be dereferenced over the HTTP protocol into data that describes the resource.

The combination of the OD movement and the LD initiative creates a powerful option for data organizations and knowledge distribution. Linked Open Data (LOD) is the method of publishing structured data (e.g., RDF, XML) that can be interlinked among different websites enabling data integration and querying (Bizer, Heath and Berners-Lee, 2009). LOD is freely available to access, download, and use.

Tim Berners-Lee (www.w3.org/DesignIssues/LinkedData.html [July 20, 2014]) envisioned a five-star model that has been widely accepted as a framework for evaluating OD and LOD projects: data distributed on an open license (1-star); distributed as machine-readable, structured data (2-star); distributed in a non-proprietary format (3-star); data identified with unique URIs references (4-star); and

data linked to other people's data (5-star). This project will target the 5-star publication model.

2.2 Publishing Linked Open Data

Berners-Lee (www.w3.org/DesignIssues/LinkedData.html [July 20, 2014]) outlined back in 2006 a set of rules for publishing data on the Web in a way that all published data could become part of a single global data space. This data is expressed in the semantic language RDF (Resource Description Framework), a World Wide Web Consortium (W3C) data model specification based on making statements about resources in the form of subject-predicate-object expressions, called triples. RDF is based on ontologies and vocabularies to model the data. Ontologies are "the specification of conceptualizations, used to help programs and humans share knowledge" (Gruber, 1993).

Publishing a dataset as Linked Data on the Web involves the following three basic steps (Bizer, Heath and Berners-Lee, 2009): (i) assign URIs to the entities described by the dataset and dereference these URIs over the HTTP protocol into RDF representations; (ii) set RDF links to other data sources on the Web, so that clients can navigate the Web of Data as a whole by following RDF links; and (iii) provide metadata about published data, so that clients can access the quality of published data and choose between different means of access.

A crucial step in the publication process is the conversion of information datasets, stored in relational databases or structured documents to RDF datasets. This process, known as triplification, relies on deciding the way of representing database schemas or document concepts to an RDF vocabulary to be used as the basis for generating the RDF triples (Salas *et al.*, 2011).

Futhermore, De Faria *et al.* (2011) have identified three steps within the basic Linked Data publication process. The first step includes the determination and creation of vocabularies and ontologies which model the data to be published. In order to reduce the effort involved in the consumption and integration of the published linked datasets, it is crucial to maximise the reuse and extension of existing vocabularies and ontologies before proposing customised ones (Salas *et al.*, 2011). Secondly, after the original data has been converted to RDF, the entities in the dataset are linked either to entities in external datasets published by the same organization or to LOD published by a different organization, such as DBPedia (Bizer *et al.*, 2009).

Thirdly, the dataset is made available on the Web in three forms: (i) through dereferenceable URIs (where an RDF document is returned when an URI is dereferenced through an URI HTTP request); (ii) through a SPARQL endpoint (a SPARQL query interface, which can be considered as being to RDF what SQL is to a relational database, for the RDF dataset which is exposed on the Web); and (iii) through RDF data dumps (files containing the RDF triples of the dataset). The published datasets should be made discoverable on the Web by providing the appropriate high-level dataset and search engine descriptors.

Based on the previous general steps, researchers have proposed different methodologies to publish LOD for a variety of domains. Although most of the examples are focused on Public Administrations and propose methodologies to

publish Linked Open Government Data (LOGD), some of the examples are focused on generic data sources (Auer *et al.*, 2011; Kaschesky and Selmi, 2013) and others cover diverse data sources, such as georeferenced and statistical information (Consoli *et al.*, 2014), meteorological information (Villazón-Terrazas *et al.*, 2011), and even data about research projects (De Faira Cordeiro *et al.*, 2011) and the performance of Internet links (Souza *et al.*, 2014).

Finally, the W3C Government Linked Data Working Group has proposed a series of best practices to publish also LOD (www.w3.org/TR/2014/NOTE-ld-bp-20140109 [July 10, 2014]) based on ten general steps. The methodology proposed in this paper is based on these steps, adapting them to small and medium size DMOs. Although previous examples could serve as guidelines for DMOs, the differences among properties of source data and requirements between domains require specific methodologies and examples for tourism LOD.

2.3 Linked Open Data for Destination Management Organizations

LOD can extensively benefit the tourism domain. Firstly, it provides syntactic interoperability enabling a common infrastructure to share tourism data and to facilitate the consumption of these datasets by infomediaries (agents that handle information between providers and consumers) in novel applications. Secondly, links established between datasets can augment data sources with additional knowledge.

Tourism Open Data examples are not difficult to find. For example, the Basque Government published a downloadable file of Points of Interest (POIs) in five languages in 2010 (opendata.euskadi.net [July 10, 2014]) which is annually updated. However, the publication of tourism LOD is still mainly related to initiatives led by research institutions. For instance, Larrinaga *et al.* (2013) describe a use case of the application of LOD in the tourism sector. Data related to destinations, accommodations and restaurants has been gathered from Open Data Euskadi and has been enriched with user reviews from social networks. Bacciu *et al.* (2014) also have transformed Open Data provided by the DMO about accommodations in Tuscany to LOD. In these examples, researchers have directly published LOD without the intervention of the DMO. Generally speaking, the two main data categories published are related to statistical data and POIs of the destination.

Focusing on LOD available at repositories from DMOs, authors have found few examples of tourism LOD with active URIs: Zaragoza (5-star) (www.zaragoza.es/ciudad/risp [July 10, 2014]) and National Parks of Oklahoma (4-star), (data.ok.gov [July 10, 2014]). Some initiatives have published tourism LOD, but do not have active URIs to be linked from other datasets as Barcelona (opendata.bcn.cat [July 10, 2014]), Italy (www.linkedopendata.it [July 10, 2014]) or statistical data of Serbia (elpo.stat.gov.rs [July 10, 2014]).

Furthermore, the TourMIS project contains information about the arrivals, bed nights and capacity tourism indicators, recorded from 1985 onwards, about over 150 European cities and in connection to 19 major markets (Sabou *et al.*, 2012). The project combines this data with data from other sources (Eurostat, UNWTO) to support decision-making scenarios on the tourism domain, establishing links between DBpedia resources and corresponding destinations.

The lack of appropriate methodologies, publication examples and tools for small and medium size DMOs can be due to the lack of best practices of tourism LOD published by DMOs. This paper details a methodology with the corresponding Open Source software tools targeting these DMOs to increase the availability of tourism LOD.

3 Methodology for publishing Linked Open Data

The proposed methodology aims at supporting small and medium size DMOs on the publication of 5-star tourism LOD, covering all the steps related to the transformation and triplification process from source files, and the linking of annotated triples to other datasets and their publication. The methodology is based on a set of well-known Open Source tools available for the publication of LOD to maximise the reuse of existing tools.

Before starting the publishing process, several non-technical and technical issues related to the data should be defined. Regarding the non-technical issues, the first decision is related to the selection and categorization of the data that will be published. Ideally, the dataset should be selected on the basis of its higher potential value for infomediaries and final users of the data. Furthermore, it is important to specify the type of license in which data will be published. Recommended licenses are the ones proposed by Open Data Commons (opendatacommons.org [July 10, 2014]), which have been designed to specifically tackle the legal issues related to OD. The technical decisions are mainly related to the correct design of the URIs that will identify the data. When data should be available in more than one language, technical decisions also include the selection of a multilingual data publication pattern (Gracia *et al.*, 2012).

The publication process consists of three main steps: pre-processing, triplification and publication. First, the data pre-processing step is responsible for extracting, cleaning and normalizing data from structured sources in a variety of data formats (CSV, JSON, Excel files). Although data should be normalised and properly stored, initial trials have shown that data may present inconsistencies and lack a standardised format. Thus, original data should be transformed into a source of structured and coherent data, typically a CSV file.

Several transformations could be applied at this step, including the definition of a coherent format for strings and numbers (i.e. telephone numbers with and without international prefix, decimal numbers separated by commas); the format for multilingual values (i.e. Bilbao, Bilbo, Bilbo – Bilbao, Bilbao – Bilbo, Bilbao / Bilbo); the correct storage of some values that could be doubled (i.e. two values in the telephone number field); and the detection of errors or non-existing values.

The LOD Refine software is proposed to perform the transformations required at this step. This programme extends Open Refine, which can be considered as a spreadsheet application that allows transforming data from a source file (CSV, JSON, Excel files...) based on formulas expressed on GREL (Google Refine Expression Language). The resulting file can be exported to a new file (CSV, RDF, Excel file,...) and all the transformations that have been applied can be stored in a JSON file.

Once initialised, the application of these transformations to new source files with the same structure or new versions of the same source file is an automatic process which only requires selecting the stored JSON file.

Secondly, the triplification step is the core of the publication process, as during this step, the properties and entities are created, expressed in triples, identified by URIs and semantically annotated with an ontology or a vocabulary. This step transforms the data obtained in the previous step into a RDF file containing semantic annotations and links.

This step begins with an analysis of the domain of the data to find related ontologies, vocabularies and linkable LOD. There are several tourism ontologies which focus on supporting tourist-centric applications. For example, QALL-ME provides a model to describe tourism destinations, sites, events as well as transportation (Ou *et al.*, 2008). The Harmonise ontology focuses on tourism events and accommodation types (Fodor and Werthner, 2005), while the Hi-Touch ontology models tourism destinations and their associated documentations (Mondeca, 2004). Moreover, general ontologies such as vCard or even DBpedia have also been proposed to publish tourism data as LOD (Bacciu *et al.*, 2011). If existing ontologies are not fully applicable to the data, new ontologies and vocabularies should be designed using an ontology editor such as Protégé, the most popular Open Source editor.

Once the ontology has been defined, the triplification process starts. The process transforms (or generates) data in RDF triples format in order to instantiate individuals based on defined ontologies. Finally, the linking process expresses the relation between items of the source dataset and items of third-party datasets.

The LOD Refine programme also includes tools to triplify and link data, generating RDF files from source files. First, external LOD (such as DBpedia) can be directly linked from external data sources in a process known as reconciliation. Then, RDF skeletons are generated defining the semantic annotations and the links to external LOD. Finally, triplified data is automatically exported as a RDF file. Once again, the whole process can be stored in a JSON file, which can be applied straightforward to new source files with the same structure or new versions of the source file.

The final step consists of the publication of the data, both as LOD and Open Data. Firstly, the previous RDF is published on a triple store, a semantic server that provides access to LOD resources directly through the URIs, or at a SPARQL endpoint. Published resources are identified with unique URIs. These URIs will return information about the resources and could be linked by other datasets. The latter allows running SPARQL queries.

There are many triple stores that provide basically similar functionalities, although they differ in the query processing time and extra features. The Open Source version of Virtuoso has been selected as it fulfils the expected requirements from small DMOs. Virtuoso has been successfully implemented by different LOD initiatives such as the Spanish LOD repository (datos.gob.es [July 10, 2014]) and other research examples (Auer *et al.*, 2011; Souza *et al.*, 2014).

Secondly, within the publication step, a dataset is created at the Open Data repository of the DMO and described with metadata adhering to the Data Catalog Vocabulary

(DCAT), a RDF vocabulary recommended by the W3C to describe OD catalogues. Then, the resources of the dataset are uploaded. These resources are composed by the published data on different formats: the transformed CSV file, the RDF file, the URL of the SPARQL endpoint, and optionally any related resource (for example, a PDF containing a report based on the data).

Following the Open Source approach, the DKAN open data platform has been selected as the OD repository for DMOs. DKAN has a simple user interface to manage the datasets, including the addition of metadata, the upload of resources, and the update of information. Moreover, it also allows third parties to look for and access data externally through an HTTP Application Programming Interface (API) called Datastore API, as well as to visualise and filter the data. Finally, DKAN is based on Drupal, a popular PHP Content Management System (CMS), making its deployment, customization and maintenance by DMOs easier.

Table 1 summarises the steps of the publication methodology and its tasks, and the software tools proposed for each step.

Table 1. Summary of the methodology

Step	Task	Tool
Configuration	Select data	-
	Select the license to publish the data	-
	Design the URI scheme	-
	Select a multilingual data publication pattern	-
Pre processing	Clean and normalise the data	LOD Refine
Triplification	Select existing ontologies, vocabularies and LOD	-
	Define new ontologies and vocabularies (if required)	Protégé
	Triplification	LOD Refine
	Link to external LOD	LOD Refine
Publication	Upload the RDF file to a triple store	Virtuoso
	Create the dataset and add metadata	DKAN
	Upload the resources of the datasets	DKAN

4 Application to a small DMO

The proposed methodology has been validated by the publication of LOD about different POIs of a small destination. Furthermore, a mobile prototype has been implemented as an example of added-value data reuse.

4.1 Publishing Linked Open Data

The configuration step involved meetings with the DMO to decide the technical and non-technical issues related to the data.

For this first prototype, a subset of 143 POIs available at the regional DMO Website (ww.urolaturismo.net/en [July 10, 2014]) in four languages (Basque, Spanish, English and French) and clustered in five categories (restaurants, accommodations, visitor centres, museums and monuments) has been selected. Source data has been provided as a set of CSV files generated by the CMS of the Website (Joomla). For each language, the following information was available for each of the POIs: name, description, category, city, address, telephone(s), email address, and Web page.

After analysing the alternatives, the Open Data Commons Public Domain Dedication and License (PDDL) was selected, as it has no requirements on attribution (include mentions to the source of data) or share-alike (keep the license of the original data) on data derived from the source data.

Regarding the technical decisions, the URI of the resources was designed as *BASE_URL/data/tourism/identifier* in order to reflect the fact that resources were related to tourism. Each URI identifies one POI, including the information about the POI in all the languages. As POIs were available in four languages, a multilingual design pattern was defined including the Basque name of the POIs as descriptive URIs, and using labels for multilingual information fields such as descriptions. The only non-ASCII Basque character is the letter "*ñ*", that was decided to be replaced by the combination "in" for URIs. It was also decided to replace blank spaces by '_' characters for URIs.

In order to transform data in the pre-processing step, three types of transformations were applied, exporting the resulting CSV to a new file. First, names of the POIs were in uppercase, so they were transformed to title case, i.e. from ERREXIL to Errexil. Secondly, international prefix was added to telephone numbers and the blank spaces between numbers were removed. Thirdly, secondary mobile numbers existing only for some POIs were stored in a new column.

During the triplification step of the data, tourism related ontologies and vocabularies were analysed (QALL-ME, Hi-Touch, Harmonise). Furthermore, there is an ongoing effort by the AENOR's Technical Standardisation Committee on Intelligent Cities AEN/CTN 178 in Spain to generate a norm about Open Data. Thus, the section about LOD of the preliminary version of the norm was also analysed. Based on the limited amount of basic information available for each POI, the general ontologies vCard and Dublin Core were selected to model the data. Related existing LOD instances were identified for the location (Geonames and DBpedia) and category (DBpedia) fields, generating links to the corresponding instances. POIs were triplified as shown in Fig. 1.

Within the publication step, the exported RDF file was uploaded to Virtuoso to load the triples; activate the URIs of the POIs; and provide a SPARQL endpoint for the dataset. The URI of the previous POI is *http://212.81.220.68:8385/data/tourism/isidro#this*.

```
<http://212.81.220.68:8385/data/tourism/isidro#this>
  a vcard:Organization ;
  vcard:fn "Isidro" ;
  vcard:adr [
    a vcard:Address ;
    vcard:street-address "Aingeru Kalea,16" ;
    vcard:locality "Azkoitia" ;
    vcard:country-name "España" ;
    vcard:region "Gipuzkoa"
  ] ;
  vcard:tel "(+34)943852003" ;
  dc:issued "2014-09-09T11:44:32" ;
  dc:description "Ordutegia: Astean zehar ... itxita."@eus,
  " Horario: Diario: ... Lunes, cerrado."@es,
  "Opening Hours:Week ... Monday closed"@en,
  "Horaire:En cours de ... Lundi: fermé"@fr ;
  dc:spatial <http://sws.geonames.org/6358108> ;
  dc:subject <http://en.wikipedia.org/wiki/Category:Restaurants> ;
  vcard:category "Azkoitian jan"@eus, "Places to eat in Azkoitia"@en,
  "Manger à Azkoitia"@fr, "Comer en Azkoitia"@es ;
  dc:isPartOf <http://dbpedia.org/resource/Azkoitia> .
```

Fig. 1. RDF of a POI (Isidro restaurant) located in Azkoitia.

Finally, a new dataset described with DCAT metadata (license, author, modified date...) was created on the data repository, which is available at *http://opendataazkoitia.gobernuirekia.net/eu/dataset/urola-turismo*. Resources were attached to the dataset (Fig. 2): a CSV file with POIs in four languages (Basque, Spanish, English and French); a RDF file; a PDF brochure of a tourism map of the area; and a link to the SPARQL endpoint.

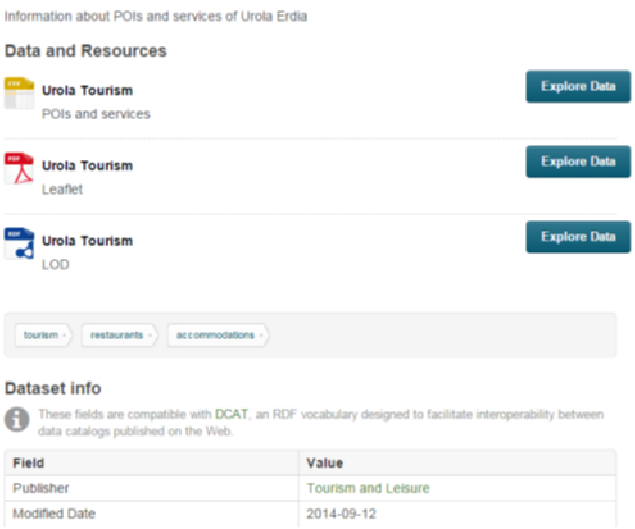

Fig. 2. Publication of resources about POIs.

4.2 Mobile application as an added value for reusing LOD

As a proof of concept of the possibilities related to LOD, a HTML5 mobile prototype describing POIs of the DMO has been implemented (Fig. 3). The prototype shows a list of the POIs and presents details about each of them, including links to further information about the location of the POIs from Geonames and DBpedia.

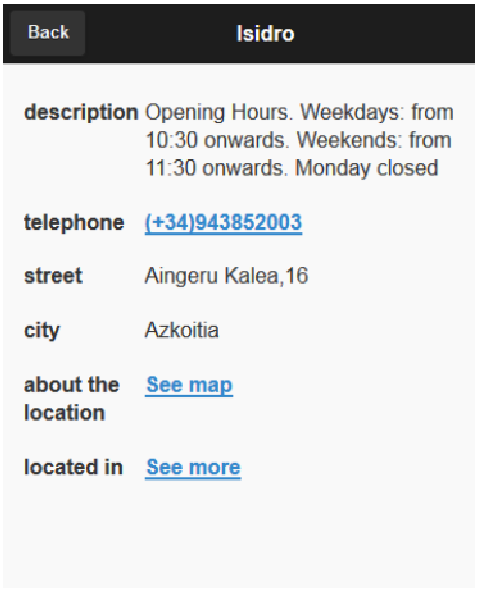

Fig. 3. Example of reusing the POIs.

The proposed software stack allows accessing data through different channels (direct download, SPARQL API and Datastore API). Developers can directly download the dataset from the repository and import it to their application. It is also possible to launch SPARQL queries using the HTTP API of the Virtuoso SPARQL endpoint in real-time from the mobile application to access only the data that satisfies some filtering conditions. Finally, the Datastore API of DKAN offers real-time access to the resources of the dataset through a regular HTTP REST API.

In order to validate the flexibility of the proposed software stack, three versions of the mobile application have been developed. Although each version presents the same interface and functionality, each one accesses data through a different channel (direct download, SPARQL API and Datastore API).

5 Conclusions and further work

The publication of LOD datasets by DMOs opens new possibilities such as the reduction of the development costs of applications (mobile guides, websites) on top of the data, or the provision of innovative added-value services which combine multiple data sources. However, the lack of integrated LOD transformation and publication methodologies, best practice examples or tools focused on the tourism domain is still a barrier for small and medium DMOs.

This paper has presented a methodology for small and medium size DMOs to publish 5-star tourism LOD. The methodology proposes an initial configuration phase and three steps: pre-processing, triplification and publication of data. The methodology has been implemented based on existing Open Source tools: Protégé to define new ontologies and vocabularies; LOD Refine to process, triplify and link data; Open Link Virtuoso to publish RDF files at SPARQL endpoints; and DKAN to publish OD in a data repository.

Data about POIs from the Web page of a small DMO from Gipuzkoa (Spain) have been used as an example to validate the methodology. Moreover, three variants of a mobile application accessing data through three different channels (direct download, SPARQL queries and REST API) have been developed as a showcase of the advantages of LOD publication for DMOs.

The methodology is currently being implemented by the DMO. Although the integration of the publication process of tourism LOD on the workflow of the DMO has already started, the publication of official datasets by the DMO has been planned to the first quarter of 2015. Thus, results of the validation of the impact of the methodology on the publication and consumption of tourism LOD are expected by the end of 2015.

Future work will tackle the publication of statistical data about tourism indicators (arrivals, occupancy rates...), as it has been already asked by the DMO. Although the methodology can cope with this change, the RDF Data Cube Vocabulary must be applied in order to model statistical data and to establish guidelines to manage historical data updated periodically such as the monthly arrival of tourists.

Moreover, new software tools should be added to publish data stored at relational databases as LOD, as the Open Source version of Virtuoso does not connect to relational databases. Successful initial tests have been done with D2RQ, a triple store for this type of data. Finally, once the AEN/CTN178 norm may be published, its recommendations and best practices for ontologies, vocabularies and generation of LOD links will be integrated in the methodology to be compliant with the norm.

References

Auer, S., Bühmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P. N., Nuffelen, B. V., Stadler, C., Tramp, S. & Williams, H. (2012). Managing the Life-Cycle of Linked Data with the LOD2 Stack.. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. X. Parreira, J. Hendler, G. Schreiber, A. Bernstein & E. Blomqvist (eds.), *International Semantic Web Conference* (pp. 1-16). Springer.

Bacciu, C., Duca, A., Marchetti, A., & Tesconi, M. (2014). Accommodations in Tuscany as Linked Data. *Ninth International Conference on Language Resources and Evaluation (LREC)* (pp. 3542–3545). Reykjavik, Iceland: European Language Resources Association (ELRA).

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22.

Consoli, S., Gangemi, A., Nuzzolese, A. G., Peroni, S., Presutti, V., Recupero, D. R., & Spampinato, D. (2014). Geolinked Open Data for the Municipality of Catania. *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)* (p. 58).

De Faria Cordeiro, K. F. F. De, Pereira, B. D. O., Freitas, A., Ribeiro, C. E., Vitor, J., & Boas, V. (2011). An approach for managing and semantically enriching the publication of Linked Open Governmental Data. *Proceedings of the 3rd Workshop in Applied Computing for Electronic Government (WCGE), SBBD* (pp. 82–95).

Fodor, O., & H.Werthner. (2005). Harmonise: A Step Toward an Interoperable E-Tourism Marketplace. *International Journal of Electronic Commerce*, 9(2), 11–39.

Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., & McCrae, J. (2012). Challenges for the multilingual Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11, 63–71.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2), 199-220.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268.

Kaschesky, M., & Selmi, L. (2013). Fusepool R5 linked data framework: concepts, methodologies, and tools for linked data. *14th Annual International Conference on Digital Government Research* (pp. 156–165).

Larrinaga, F., Lizarralde, O., Serna, A., & Gerrikagoitia, J. K. (2013). Caso de uso de Open Data y Linked Data en Turismo. *tourGUNE Journal of Tourism and Human Mobility*, pp. 31–39.

Mondeca. (2004). Semantic Web Methodologies and Tools for Intra-European Sustainable Tourism. White paper.

Ou, S., Pekar, V., Orasan, C., Spurk, C., & Negri, M. (2008). Development and Alignment of a Domain-Specific Ontology for Question Answering. *6th Edition of the Language Resources and Evaluation Conference* (pp. 2221–2228).

Sabou, M., Brașoveanu, A. M. P., & Arsal, I. (2012). Supporting Tourism Decision Making with Linked Data. In H. Sack & T. Pellegrini (Eds.), *I-SEMANTICS '12 Proceedings of the 8th International Conference on Semantic Systems* (pp. 201–204). Graz, Austria: ACM New York.

Salas, P., Viterbo, J., Breitman, K., & Casanova, M. (2011). Stdtrip: Promoting the reuse of standard vocabularies in open government data. In D. Wood (Ed.), *Linking Government Data* (pp. 113–133). Springer New York.

Souza, R. F., Cottrell, L., White, B., Campos, M. L., & Mattoso, M. (2014). Linked Open Data Publication Strategies: Application in Networking Performance Measurement Data. *Academy of Science 2014 Bigdata / SocialCom / Cybersecurity*.

Villazón, B., & Vilches, L. M. (2011). Methodological Guidelines for Publishing Government Linked Data. In D. Wood (Ed.), *Linking Government Data* (pp. 27–49). Springer New York.

Acknowledgements

Authors would like to thank both the Basque Government and the Regional Government of Gipuzkoa for partially funding this work; as well as the companies Elkarmedia, Eko.tx and Iraurgi Berritzen who have been part of the consortium of the project.