La Rioja Turismo: The Construction and Exploitation of a Queryable Tourism Knowledge Graph

Ricardo Alonso-Maturana[(✉)], Elena Alvarado-Cortes,
Susana López-Sola, María Ortega Martínez-Losa,
and Pablo Hermoso-González

GNOSS: RIAM I+L LAB. S.L.,
C/Piqueras nº 31 4ª planta, 26007 Logroño, La Rioja, Spain
{riam,elenaalvarado,susanalopez,
mariaortega}@gnoss.com, pablohermoso@gmail.com

Abstract. The institutional website for La Rioja tourism (https://lariojaturismo.com) is a working example of the construction and exploitation of a tourism Knowledge Graph where all digital contents referring to attractions, accommodation, tourism routes, activities, events, restaurants, wineries, etc., are semantically represented in RDF/OWL. The construction of the Knowledge Graph was carried out through the conceptualization of a Digital Semantic Model that hybridized and extended several existing ontologies and vocabularies (mainly Harmonise, OnTour, Geonames, Rout, FRBR and rNews). The overarching objective was to generate a digital space where information retrieval was simpler, more useful and more practical, offering a much more friendly and satisfactory website experience. At present, the Knowledge Graph is made up of more than 7,000 digital contents; 67,284 entities; 472,361 relations; and 675,368 triples. The digital space receives more than 40.000 visits per month. The most important Knowledge Graph exploitations are associated with the existence of a metasearch engine, faceted search engines for each knowledge object, contextual information systems, and Graph visualization systems through combination of map and semantic geo-positioning.

Keywords: La Rioja · Knowledge Graph · GNOSS · Semantic technologies
Cognitive platform · Smart tourist destination · Digital Semantic Model

1 Introduction: The La Rioja Turismo Portal as a Queryable Knowledge Graph for Building a Smart Tourist Destination

In 2014, La Rioja, the smallest autonomous community in Spain, decided to represent its tourism content semantically. The main reason they built a tourism Knowledge Graph was to contribute to the configuration of La Rioja as a smart tourism destination. A technological strategy was utilized that offered tourists a simple, useful, practical and user-friendly experience with the information, while facilitating the more efficient management of tourism information by various stakeholders (tour operators, journalists,

the regional tourism sector). Configuring the Knowledge Graph allowed foundations to be laid in a simpler and more agile way so as to be able to later extend and expand the model, as well as to integrate and link (linked data) [1] other digital contents whose primary purpose is not tourism, as opposed to cultural or journalistic contents, yet could in fact exercise this function and enrich the first graph created. The possibility of configuring a website with tourism data that could be related to other semantically-expressed data did not exist with other technological approaches.

The project integrated and linked the region's tourism contents through the construction and exploitation of La Rioja's tourism Knowledge Graph[1], implementing a modification from the traditional operational mode that is still common in most digital projects for public and private institutions. In this model, contents are labeled and managed in traditional content management systems (CMS) such a Drupal, Workpress, Liferay, etc. The approach used for La Rioja was based on the creation of a Digital Semantic Model[2], which identified existing entities, their properties and relations. Existing ontologies were hybridized and extended with the interests of the end user in mind and an integrated Semantic Content Manager allowed data to be published natively in RDF/OWL. This creates advantageous exploitation possibilities together with a much more useful and efficient web experience.

This project contributed to the development of a smart tourist destination, as could be read in September 2015 in "Smart Tourism Destinations: constructing the future" [2], a report prepared by *Sociedad Estatal para la Gestión de la Innovación y las Tecnologías Turísticas, S.A.* (SEGITTUR), within the framework of the *Plan Nacional de Ciudades Inteligentes* [Spanish National Smart Cities Plan] by Spain's Digital Agenda institution.

The creation of the Knowledge Graph and its exploitation needed to increase the value of its contents, that is, to make them more known, more accessible, better positioned on the internet for search engine retrieval. They had to be able to be linked so as to offer tourists a website user experience wherein a search transforms into a path to learning and discovery.

This enriched user experience stimulates the desire not just to spend more time on the travel destination website, but above all, to physically visit the place. In other words, increasing the value of digital content means increasing the conversion rate, transforming the mere search and inquiry for information into the visit of and consumption and purchase at the specific destination.

This objective has been met; portal usage data shows that the portal's usage ratio increased by 42.66% from 350,589 users in 2014 to 503,017 users in 2017. These users visit 2.5 million web pages and spend an average of 2.5 min on the website.

---

[1] "Web Semántica y Turismo" (2014), https://es.slideshare.net/gnoss/web-semntica-y-turismo-caso-la-rioja-turismo-gnoss-2014. Accessed 2018/04/03.

[2] Museo del Prado. Digital Semantic Model https://www.museodelprado.es/en/modelo-semantico-digital/el-prado-en-la-web.

2 The Knowledge Graph and Its Main Applications

2.1 Definition: The La Rioja Turismo Knowledge Graph

The La Rioja Turismo Knowledge Graph is the semantically-represented (RDF/OWL) system comprising the region's more than 7,000 tourist resources (accommodation, restaurants, wineries, tourist services, events, tourism routes, attractions, and villages and urban centers) using entities and entity attributes. The Graph understands facts about them, as well as any object potentially linked to them (news, brochures, activities, schedules, bids, multimedia resources). In particular, it understands the way that this set of entities is interconnected.

The La Rioja Turismo Knowledge Graph now integrates 7,045 digital resources with 67,284 entities and 472,361 relations. Entities are used to understand the meaning of the term that the user searches for and offers a system by which to explore all of its resources. This tool is based on a faceted search engine, among other utilities, and avails the user of all possible navigation modes concerning that set of entities. The number of triples in the Knowledge Graph is 675,368.

2.2 Properties of the La Rioja Turismo Knowledge Graph

The La Rioja Turismo Knowledge Graph has four main properties that make it uniquely efficient at meeting specific objectives. The Knowledge Graph is:

- **Unified** because it enables data that are hosted in scattered, heterogeneous and diverse management systems to be integrated. These data may exist in a variety of formats and be structured to several different degrees. For example, content originating from distinct data sources such as CRM, ERP, Document Managers and Content Managers could be integrated using this "semantic layer", similarly to the Prado Museum[3]. In the case of La Rioja Turismo, this step was simplified since only one data source was used, from which all resources were obtained. The contents of the old portal content manager were migrated. From that point on, the new contents were created directly using semantic CMS. A metadata layer represented semantically in RDF/OWL has been generated for all contents. The contents are connected within a Digital Semantic Model that represents the entities, their attributes and their relations, linking all contents into a single graph, independently of their sources of departure.
- **Queryable**, **by both people and machines.** With the former, humans are able to generate information retrieval systems that enable reasoning. We demonstrate this using faceted searches [3], whose purpose is to provide a means by which to retrieve information for anyone who visits the website, and to do so according to their interests and intentions. Faceted search enables iterated interrogations, and therefore the ability to refine searches as much as desired. In turn, it becomes possible to provide users with well-organized, enriched and contextualized faceted

---

[3] The Museo del Prado's Knowledge Graph. https://www.museodelprado.es/en/modelo-semantico-digital/el-grafo-de-conocimiento-del-museo-del-prado.

results [4] that correctly correspond to their search. The specific aim is to provide access to the portal to the public with an intuitive, personalized, semantically meaningful and effective smart browsing and search experience, thus encouraging them to continue exploring the graph. These exploitations may be carried out in all case studies completed by GNOSS.

- **Expressive:** the semantic representation of contents in RDF/OWL can be as rich and expressive as its semantic model; a machine will be able to "understand", interpret, and therefore make use of the entities, attributes and relations that define a given digital resource. For example, in the case of an event, his includes where it takes place, on which dates, what restaurants are nearby, etc. Thus, machines are able to understand what each digital resource means and thereby help users to connect data with data. In addition, this enables interoperability with other systems.
- **Extensible,** because new entities, the properties of existing or new entities, and new relationships between them can be agilely and flexibly incorporated. This makes it possible to integrate and link data with new data from the same source or with other data banks and contents in an appropriate way. The Digital Semantic Model's level of abstraction facilitates extension to new data, other data sources and other sets of content much simpler way than the traditional data model.

2.3   La Rioja Turismo's Ontological Model

The configuration of the Digital Semantic Model for the La Rioja Turismo web portal was the first step in constructing the La Rioja Knowledge Graph and its advantageous exploitation for the end user. An exhaustive study of state-of-the-art semantic standards, ontologies and existing ontological models was carried out as part of model creation. The goal was to take advantage of all existing developments that were useful for mixing, hybridizing, extending and combining them in the best manner possible, focusing on the construction of a useful, practical, simple web experience that stimulates visitors when questioning, inquiring and discovering information.

The ontological project carried out with La Rioja Turismo for the construction of its Knowledge Graph hybridized a wide range of domain ontologies, integrating them into a common ontological framework that represented the contents and activities in the tourism field. These range from content related to accommodation, dining and activities to tourism routes, towns and attractions, as well as tourism services, news and related articles. Below we present the set of hybridized vocabularies in the La Rioja Turismo project.

Semantic Digital Model

The hybrid ontology discussed here has been consolidated into what could be called the Rioja Turismo Digital Semantic Model. It is composed of a set of vocabularies developed to represent a large portion of the contents and tourist activities. The following diagram represents the first La Rioja Turismo Digital Semantic Model, which displays the set of ontologies that were combined in the ontological story.
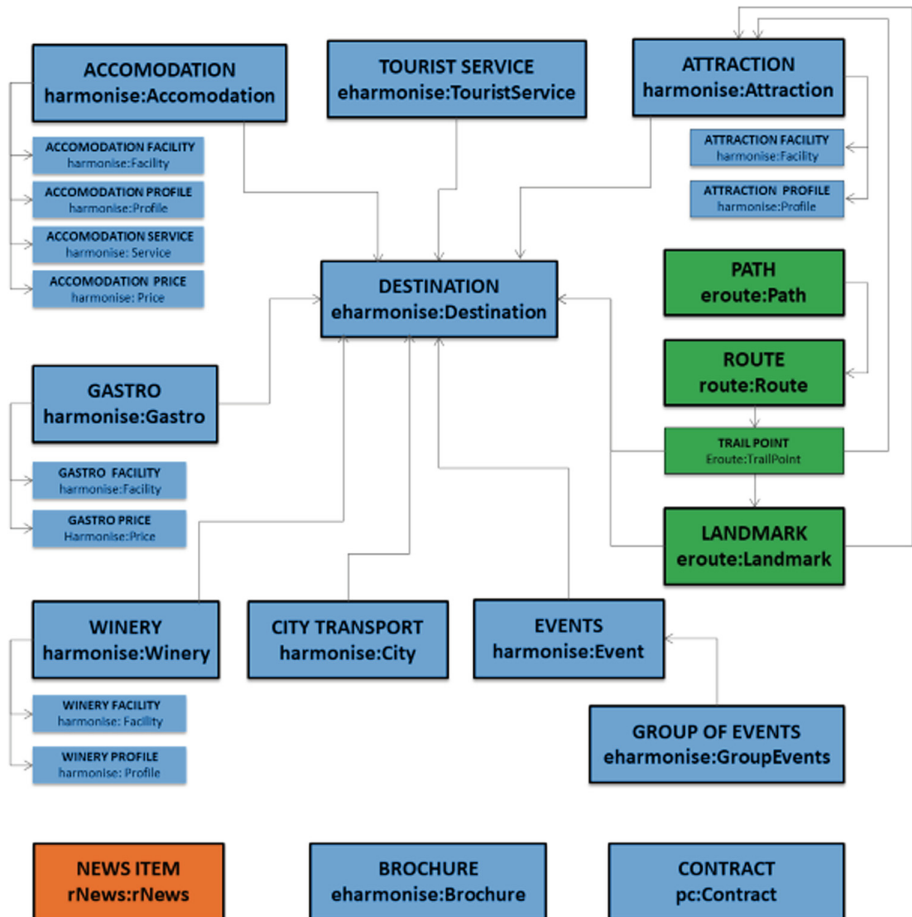
The following ontological models were analyzed [5] in order to construct the Digital Semantic Model. The original bibliography consulted included:

- Harmonise Ontology
- Mondeca Tourism Ontology
- Hi-Touch Ontology
- QALL-ME Ontology[4]
- DERI e-Tourism Ontology
- cDott Ontology
- Cross Ontology
- Contur Ontology

---

[4] https://hlt-nlp.fbk.eu/technologies/qall-me-ontology.

- *Ontología sobre Rutas Turísticas (a Pie o en Bicicleta) por Espacios Naturales* [Ontology on Tourism Routes through nature (on foot or by bicycle)]
- EON Traveling. Although this was one of the first ontologies, it seems to have fallen into disuse
- GETESS Ontology
- ANOTA ontology
- Tourism ontology schema
- World Tourism Organization (WTO) Thesaurus

After the contents of the La Rioja Turismo website and the various ontological models were analyzed, the main basis used for the development of the La Rioja Turismo Digital Model was the Harmony ontology. It was chosen for its maturity, the fact that it had been tested in several projects (including Turespaña) with a positive outcome, and because it includes concepts that were able to represent a majority of the contents required by the La Rioja tourism website. This was the preferred model upon which to base the Qall-me ontology, which also includes Harmonise concepts. Qall-me is a much more hierarchical ontology (and therefore more difficult to maintain) and is not as mature as Harmonise in production projects (beyond other prototypes). The cDott ontology, which could have represented a large advance, appeared to be experimental; with the exception of the bibliography where the model is described, it was impossible to locate the complete ontology. The Mondeca and Hi-Touch ontologies were rejected due to private ownership. Where the model required extension at certain points, however, the Qall-me, Deri and cDott ontologies were used as a reference (destination/location).

Finally, an expansion using other ontologies was required in order to represent entities that were not included in the aforementioned models: tourism tours or routes (the UOC [*Universitat Oberta de Catalunya*] model was the reference for the first approach), recipes, wines, wineries or other tourism services (service providers). It should be noted that where model enrichment was deemed suitable due to interoperability issues or for specific operations, the model was extended using other ontologies such as Geonames (for geographic data) or SKOS (for categories).

In summary, the Knowledge Graph was built upon the foundation of a Digital Semantic Model designed for the La Rioja tourism web portal in addition to hybridized and extended several ontologies and existing vocabularies:

- Harmonise ontology[5], to represent tourist destinations, attractions, events, services, accommodation, restaurants and wineries.
- OnTour ontology, to express certain properties not encompassed by Harmonise.
- Rout, to model routes[6].
- Geonames[7] for locations and WGS84 to express latitude and longitude.

---

[5] http://www.harmosearch.com/.

[6] http://openaccess.uoc.edu/webapps/o2/bitstream/10609/2284/1/igutierrezl_articulo.pdf.

[7] Geonames: http://www.geonames.org/.

- Functional Requirements for Bibliographic Records (FRBR)[8] to represent document data
- rNews[9], for news.

2.4   Main Exploitations of the Knowledge Graph: Advantages

For the end user, the main advantages of browsing www.lariojaturismo.com are:

- A graph querying system where information can be found in a much more precise, useful, and practical way, which saves time and finds what the user wants to find. This can be observed through the metasearch engine (a search engine that performs global searches of all website content) and the specific faceted search engines for each case, together with specific facets according to the type of element selected (for example, search engines specific to wineries and restaurants, locations and attractions, routes, activities, or accommodation, among others) [6].
- Its system for generating informative contexts. When information on a winery, restaurant, hotel or activity is presented, the graph displays the most pertinent related resources based on shared attributes. La Rioja Turismo contexts are configured so that once the tourist views information about a resource, they are able to see what they can do, where to eat and where to sleep in relation to the content visited.
- Graphic visualization and enriched information on a map thanks to semantic geolocation. Visualizing the Knowledge Graph allows the user to combine the map view and locate information of interest to them. Faceted search both permits filtering in order to refine searches. It also displays information about other attributes related to the object sought. For example, as the user zooms in on the map to view wineries in Haro, other winery-specific attributes appear on the screen, such as their portfolio of services, hours, the languages in which they offer their services, types of facilities, and more. Latitude/longitude is one of many attributes of the entity 'winery' and one can "reason" using these attributes. By zooming in on the map, the data corresponding to the defined area are reconfigured and adapted to the selected space. Corresponding valuable information is made available as the search progresses. Likewise, summarized information may be displayed for each element located on the map without having to leave the map.

3   Conclusion

The www.lariojaturismo.com website has served as a case study in the efficient and practical exploitation of a built Knowledge Graph. The conceptual and technological difficulties, as well as the corresponding opportunity that arose from them, arose from the lack of a complete ontological model for the expression of an institutional tourism

---

[8] http://www.sparontologies.net/ontologies/frbr.

[9] http://dev.iptc.org/rNews.

website. This created the need for a Digital Semantic Model where several existing ontologies hybridized and mixed.

The main challenge of this project was convincing of the need to dispense with the limits and obligations linked to transactional logic and filing documents. SQL logic prevented expressing the content and its meaning in an expressive way, and therefore limited the ability to discover information, query as a human would, and make inferences of interest.

Through the linking of semantically-represented data, the extensibility of the Knowledge Graph enables the advancement and incorporation of new information relevant to the next stages, such as cultural data that have value in tourism.

It also enables the enrichment of information through connection with other datasets such as those in the DBPedia. The next steps concern the extension and deepening of the Knowledge Graph "inward": generating a semantic marketplace for tourism resources, improving the information access system via a system accompanying the tourist on a "here and now" visit that provides them with more useful information when planning their visits, and generating a digital backpack where interesting tourism resources can be shared with other users through the increasingly intense exploitation of the social graph.

La Rioja Turismo shows how the creation and exploitation of Knowledge Graphs in the tourism sector is a winning strategy when it comes to addressing the future of smart organization search and query systems. A queryable Knowledge Graph enables the implementation of personalized search strategies based on reasoning and the capacity to contextualize systems.

References

1. Serna, A., Murua, I., Gerrikagoitia, J.K., Alzua, A., Lizarralde, O., Larrinaga, F.: Economía de Datos en Turismo proceso para la publicación de LOD en turismo. tourGUNE J. Tour. Hum. Mobil. (2013)
2. Alonso Maturana, R.: El valor de los contenidos digitales en los destinos turísticos inteligentes. Informe Destinos Turísticos Inteligentes: construyendo el futuro, Segittur, pp. 126–129 (2015)
3. Suominen, O., Viljanen, K., HyvÄnen, E.: User-centric faceted search for semantic portals. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 356–370. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72667-8_26
4. Ferré, S., Hermann, A., Ducassé, M.: Semantic faceted search: safe and expressive navigation in RDF graphs. Research report (2011). ISSN 2102-6327
5. Kathrin Prantner, Y., Michael Luger, Z.: Austria. "Tourism ontology and semantic management system: state-of-the-arts analysis". In: IADIS International Conference WWW/Internet 2007 (2007)
6. Dal Mas, M.: Faceted semantic search for personalized social search. In: Computing Research Repository, abs-1202-6685 (2012). http://arxiv.org/abs/1202.6685