Semantic-based Open Data Portal for the Spanish National Library

What need was solved?

Historically, libraries have been responsible for storing, preserving, cataloguing and making available to the public large collections of information resources.

The Spanish National Library has a collection of information sources such as:
1. Authority records: These records contain metadata describing people, organizations, work titles, and subject headings.
2. Bibliographic records: These records form a representative subset of the BNE catalogue, and provide metadata belonging to modern and ancient monographs, electronic records, manuscripts, periodical publications, maps, engravings, photographs, printed music, sound and audiovisual recordings.

The Spanish National Library requested to the Universidad Politécnica de Madrid to create an open data portal based on semantic technologies that integrates all the data coming from the above heterogeneous sources and connect the Spanish library data with other data from relevant international libraries.

What services were provided?

The service provided is the development of the datos.bne.es, an open data portal for the Spanish National Library. The portal is in production 24*7, allowing the general public and librarians to access the catalogue using advanced features for searching over different types of entities (e.g., authors, works, translated works, or subjects), navigation across different types of connected entities (such as the author of a work) using advanced features for data indexing, ranking, retrieval and visualization. In addition to the portal itself, the following services are provided.

1. A network of ontologies based on more than ten different bibliographic ontologies. The BNE ontology builds on the IFLA FRBR model (Functional Requirements for Bibliographic Records), although it reuses several terms from other ontologies such as the RDA ontologies, Dublin Core, or the Bibo ontology. This rich ontology network is published following the Linked Data principles and publicly available datos.bne.es/def.
2. To transform a large catalogue composed by more than 4 million MARC 21 bibliographic records, more than 3 million MARC 21 authority records (e.g., providing personal information, classification information), and to more than 100,000 digitized materials in RDF triples. Data curation and integration are also supported.
3. Linking the Spanish National Library data to the following library datasets: The Virtual International Authority File (VIAF), The Authority File of the German National Library (GND), The French National Library (DBpedia), The authority and bibliographic catalogues of the National Library of Sweden (Libris), and SUDOC. The number of links to external datasets was significantly increased up to 1,395,108 links. Additionally, 108,834links to digital materials were added.
4. The release of a large and highly interlinked dataset under a public domain license. The dataset is query-able using the SPARQL language through a public endpoint datos.bne.es/sparql, and through a standard Linked Data front-end that provided access to the data in different formats using content-negotiation.

5. An end-user interface was developed to give access to the vast amounts of interconnected entities. This user interface was built exclusively using the Linked Data and leveraged the data connectivity and the underlying ontology to index, rank, present, and arrange information catalogues to rich ontology-based data, and their publication and exploitation on the Web.

The relation with digitization

In recent years, the amount of semantically structured knowledge available on the Web as part of the so-called *Linked Open Data (LOD) cloud* has seen a substantial growth in the domain of cultural heritage and, particularly, in digital libraries. Indeed, libraries, museums and archives are showing great interest in publishing their data as Linked Data (LD). Initiatives like Europeana LOD or VIAF also stress the opportunities offered by digitalizing cultural heritage archives using Linked Data.

The benefits of publishing library data as Linked Data have been summarized in the by the W3C Incubator Group on Library Linked Data. In particular, the following key benefits of *Library Linked Data (LLD)* with regards with digitalization have been identified: i) provides enhanced and more sophisticated navigation through information, ii) increases the visibility of cultural data, iii) supports integration of cultural information and digital objects into research documents and bibliographies, iv) offers a more durable and robust semantic model than metadata formats that rely on specific data structures, v) facilitates re-use across cultural heritage datasets, thus enriching the description of materials with information coming from outside the organization's local domain of expertise, and vi) allows developers and vendors to avoid being tied to library-specific data formats such as MARC.

The datos.bne.es system is the result of a project that started with the goal of applying semantic and Linked Data technologies to the catalogues of the National Library of Spain. The system is in production being fully available 24x7 to citizens and librarians.

Customer

Spanish National Library
www.bne.es
Spain