Motivation

- Primary motivation was finding a tool to achieve multilinguism, a main topic in the Spanish library environment.
- Expose bibliographic and authority data as Linked Open Data, in order to:
  - Enhance and enrich the user experience of navigating the data, using internal and external sources.
  - Make rich library data open and reusable.
Project features

- Achieved between BNE and the Ontology Engineering Group. (Universidad Politécnica de Madrid)
- The work began at early 2011 as a small proof of concept: work by and related to Cervantes.
- Eventually, it evolved into a bigger part of the catalogue. It was presented in December 2011, as datos.bne.es
Potential sources of data

- More than 4 millions bibliographic records
- More than half million authority records
- More than 100000 digital objects
- Other sources: virtual exhibitions, archive
Modelling and vocabularies

- Proof of concept of exhaustively using IFLA models and vocabularies: FRBR, FRAD, FRSAD and ISBD ontologies.
  - Widely agreed upon by the library community, potentially with no loss of data
  - Provides a rich framework for exposing, sorting and connecting library data.
  - Sustainability: backed by an international association.
  - Enable data to be related with other theoretical models from other similar disciplines.
Advantages of FRBR and FRAD as models

- **Order**: grouping of related data that leads to powerful and neat visualization schemes
- **Relationship**: discovery of related resources
- **Abstract models can be mapped to other models**: CIDOC-CRM
IFLA vocabularies

- All of them are in RDF in the OMR with the status *published*
- The elements are denominated by numeric designations:
  
  \[\text{http://iflastandards.info/ns/isbd/elements/P1004}\]
  
  \[\text{Isbd:P1004} \rightarrow \text{“has title proper”}\]
  
  - Goal: linguistic neutrality
  - The “human side” has been translated into many languages
- Additional elements from various vocabularies were added: DC, RDA, MADS/RDF,
IFLA vocabularies

- `<skos:scopeNote xml:lang="en"> The title proper includes any alternative title, but excludes parallel titles and other title elements of subseries or parts, the title proper can consist of two or more components: the monographic resource, the dependent title and the dependent title designation. For resources containing several individual works and lacking a collective title, has its own title proper. </skos:scopeNote>

  <rdfs:label xml:lang="es">tiene título propiamente dicho</rdfs:label>

- `<skos:definition xml:lang="es"> Relaciona un recurso con el título propiamente dicho (el nombre principal de un recurso fuente preferida de información para el recurso). </skos:definition>

- `<skos:scopeNote xml:lang="es"> El título propiamente dicho incluye cualquier título alternativo, pero excluye títulos para sección o algunos suplementos y algunos títulos de subseries o partes, el título propio (o el título de la serie principal o del recurso monográfico multiparte), el título dependi
Building FRBR/FRAD and Linked Data from MARC21

Authority records -> frbr:person, frbr:corporate body

frbr:work

frbr:expression

Bib records -> frbr:manifestation

http://bne.linkeddata.es/mapping-marc21/index.html

Clasification

100 $a -> Person
100 $a $t -> Work

Annotation

260$a Publication place

Relation

100 $a author of Work $t
Conversion process

- Transformation was carried out with the tool Marimba, designed by the OEG group.
- This tool supports the mapping through user-friendly spreadsheets.
- Allows librarians to map manually the pieces of information from MARC records (fields and subfields codes) to the ontologies of their choice.
- Marimba pre-processes MARC records and present librarians with the current metadata elements used in their catalogue, and they assign the correspondences.
Results

- It has been migrated:
  - Authority file (persons, corporate bodies)
  - Modern Monographs: 1,947,332
  - Ancient Monographs: 107,803
  - Notated music: 162,519
  - Sound recordings: 172,484

- Building of datos.bne.es:
  - Datadumps (from the website and Datahub)
  - Sparql end-point
Results (2)

- Data released under CC0 license.
- More than 58 millions of triplets.
- Owl:sameAs links (downloadable also as a single datadump) to VIAF, GND, LIBRIS, SUDOC and DbPedia.
Reception

- Data and FRBR constructing and mapping has been used, studied and cited by JISC (BibSoup) or ABES.
- Reviews from OpenBiblio:
  - “A complicating matter from a data wrangler’s point of view is that the field names are based on IFLA Standards, which are numeric codes and not ‘guessable’ English terms like DublinCore fields for example. This is more correct from an international and data quality point of view, but does make the initial mapping more time consuming.”
- No known reuse of data beyond library community.
Problems found and future developments

- Mapping problems.
  - Not all FRBR basic relations could be extracted or inferred from MARC records.
  - MARC ambiguities make exact matching difficult.

- Future developments:
  - Visualization tools for human consumers.
  - Mapping refinements.
  - New links to selected datasets.
  - New sources of data: digital objects.
  - Better coding and cataloging procedures to improve future matchings.
Other developments: subjects in SKOS vocabulary

- Subjects records were in large part left aside in the first migration.
- SKOSification of BNE Subject List (2013)
- Accessible from the Sparql end-point from datos.bne.es.
- Data not yet fully integrated with the bibliographic record.
- Despite being in beta phase and lack of promotion, has been already reused, although in library domain.
Other developments: BNE Escolar  
(BNE for schools)

- Aim: semantic exploitation and educational valorization of Spanish cultural goods.
- Tool for students and teachers of primary and secondary education.
- Digital objects selected from the BNE Digital Library, related to enrich and complement educational contents of the intermediate education curriculum.
- Collection of over 8,500 digitalized documents of the Hispanic Digital Library Selection (books, drawing, engravings, maps…)
- It is part of the larger Didactalia, a big platform of educational resources, built upon semantic mark-up. Allows social tagging from teachers recommending contents.
- Records are taken from MARC21 and converted into Learning Objects Models (LOM) ontology, and enriched to other sources.
Notas y correspondencia de vocablos " rabes, cursive."
Library-specific models and vocabularies (namely, FRBR-based ontologies), provides a rich way of publishing data, and an appropriate frame for relationships to other collections from libraries and other heritage institutions.

At the same time, library models and vocabularies seem not appropriate to accomplish one of the Library Linked Data goals, that is, make library data understandable and thus reusable to other parties.

Currently, National libraries serve a wide audience with different needs, so different approaches are suitable, as our experience shows.