



## Data.bnf.fr: FRBR and Linked Data at the French National Library

Agnès Simon, Romain Wenz

### ► To cite this version:

Agnès Simon, Romain Wenz. Data.bnf.fr: FRBR and Linked Data at the French National Library. 2011. hal-03246107

**HAL Id: hal-03246107**

**<https://hal.science/hal-03246107>**

Submitted on 2 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Data.bnf.fr: FRBR and Linked Data at the French National Library

*Agnès Simon and Romain Wenz, Bibliothèque nationale de France*

The Bibliothèque nationale de France (BnF) has designed a new project, data.bnf.fr [1], in order to make its resources more useful on the Web. Currently the BnF provides free access to more than 1.5 million digital items in Gallica, its digital library, and contains more than 12 million records in its catalogues. But this information is hard to reach for the general public. The BnF decided to provide access to these resources by building

simple Web pages, in HTML, that would gather the links to all the resources owned by the library, about authors and works. A first version of data.bnf.fr was published online in July 2011. It is frequently improved, and provides data that are part of the Web: fully accessible for machines and for users. The pages are created for works that have been studied at least once. As of December 2011, there are more than 35.000 Web pages, providing links to 750.000 references, which makes it one of the biggest library services freely available.

The pages are built using Semantic Web techniques which match the resources and bring them together automatically. The original data was created in various formats such as INTERMARC for the main catalogue, XML-EAD for archives inventories and Dublin Core for the digital library and virtual exhibitions. Therefore, the first challenge was to use persistent identifiers and to implement matching techniques in order to build the links. We rely on the authority files of the BnF (persons, works, and subjects). These authority files are used as a basis, with persistent (ARK) identifiers. This set of named entities also provides alternative labels, dates, abstracts, thumbnails and other information from the different catalogues, which we match and align together. That way, it is possible, for instance, to find on a single page the abstract of a book, the link to the original manuscript, the list of the editions and the digital copies available online for free. For example, the editions and performances of Notre-Dame de Paris:

[http://data.bnf.fr/11967596/victor\\_hugo\\_notre-dame\\_de\\_paris/](http://data.bnf.fr/11967596/victor_hugo_notre-dame_de_paris/)

This web interface gathers resources (data and links to digital documents) around the concepts of “works”, “authors,” and “subjects,” using the conceptual framework of the *Functional Requirements for Bibliographic Records* (FRBR). However, the “Expression” level is only partially implemented. A page devoted to a “Work” collocates all the editions of this work, sorted by content type, and displays the relationship between the work and its author. A page devoted to an “Author” lists works created by this author, but also the author’s contributions to expressions of works by other persons, e.g. as translator or editor (see: Charles Baudelaire as the French translator of Edgar Poe [2]).

The web pages are automatically created and rely on our bibliographic and authority records, about persons, corporate bodies, works and RAMEAU

subjects. “Manifestations” associated with an author are gathered around his works, thanks to the hypertext link between the original bibliographic records and Title authority records (The BnF uses Title authority records for some of the Works). We also align “Manifestations” that are not explicitly linked to the Title authority records, if they are explicitly linked to an Author authority record in the bibliographic record, and if the character strings of these manifestations match the work’s title, and if publication dates and links to other contributors are not conflicting with the information provided by bibliographic records linked to the Title authority record. However, some manifestations remain unaligned. We are taking a next step by working on advanced algorithms in order to:

- ❖ group manifestations that couldn’t be gathered via the algorithms mentioned above, with works described in a Title authority record;
- ❖ gather manifestations around new “pages”, when there is no Title authority record.

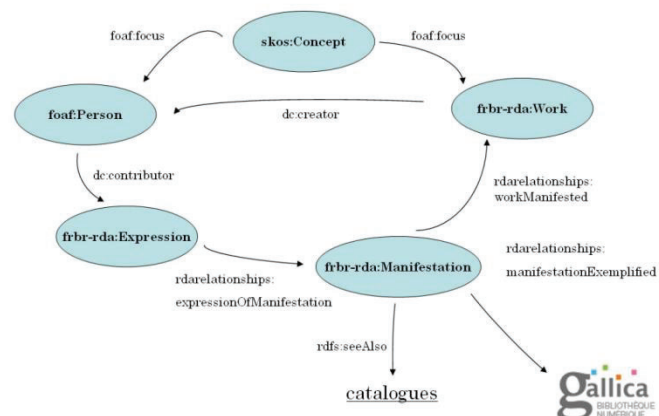
We carried out this research in the belief that it will help us design new tools for the implementation of the FRBR model in our catalogues.

We want to create interoperability between our own datasets and with external datasets, to be able to exchange, in the future, not only with other libraries in France and abroad, but also with other cultural institutions and communities. We made vocabularies available, with all of the “RAMEAU subject headings” in RDF, in SKOS. Data.bnf.fr provides HTML pages based on the FRBR model and raw data in an RDF framework, following the principle of linked open data architecture. This RDF data follows the “Semantic Web” principles, by using persistent URIs, and existing vocabularies. We use Web standards, which are popular, not only in libraries but also in other communities: **SKOS** for concepts, **FOAF** for authors, **Dublin Core** and **RDA** describing the work’s main elements. For example, the raw data of “Les travailleurs de la mer”, by Victor Hugo:

[http://data.bnf.fr/11992081/victor\\_hugo\\_les\\_travailleurs\\_de\\_la\\_mer/rdf.xml](http://data.bnf.fr/11992081/victor_hugo_les_travailleurs_de_la_mer/rdf.xml). Some elements had to be described with frbr-rda properties: for instance, relationships between work, expression and manifestation. Indeed, data.bnf.fr is not a new catalogue following all the FRBR requirements, but a specific service. We had to adapt this domain model to our application, so that users can “pick and choose” the part of the data they want to use.

The current data.bnf.fr pages are built that way, which shows that these techniques are scalable.

Data is organised in this simple model:



We also link our data to external data sets, such as the Library of Congress, and the Deutsche Nationalbibliothek, but also VIAF and DBpedia. data.bnf.fr is built with an open software: CubicWeb, and the RDF data is available using an Open licence (Attribution) so that the data is not only technically usable, but also freely available. data.bnf.fr is described in the French Governmental portal, data.gouv.fr [3], and part of the Linked Open Data movement.

[1] <http://data.bnf.fr/>

[2] [http://data.bnf.fr/11890582/charles\\_baudelaire/](http://data.bnf.fr/11890582/charles_baudelaire/)

[3] <http://www.data.gouv.fr/>