

Making the most of library data on the web: the data.bnf.fr example

English translation of the original paper: “Améliorer l'accès aux données des bibliothèques sur le web : l'exemple de data.bnf.fr”.

Translated by Emmanuelle Bermès and Céline Leclaire, National Library of France (BnF), Paris, France.

The text of this document has been translated into English and differences from the original text may occur. This translation is provided for reference purposes only.

Emmanuelle Bermès

Networks and services Directorate, France's National Library, Paris, France

emmanuelle.bermes@bnf.fr

Vincent Boulet

Metadata department, France's National Library, Paris, France

vincent.boulet@bnf.fr

Céline Leclaire

Networks and services Directorate, France's National Library, Paris, France

celine.leclaire@bnf.fr



This is an English translation of “Améliorer l'accès aux données des bibliothèques sur le web : l'exemple de data.bnf.fr” copyright © 2016 by Emmanuelle Bermès and Céline Leclaire. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/4.0>

Abstract

This paper deals with data.bnf.fr, a service launched in 2011 by the National Library of France (BnF) to make better-referenced data easily findable by internet users and to act as a hub for resources from other applications (the [general catalogue](#), [Gallica](#)) by providing web pages dedicated to authors, works and subjects.

Such an example highlights the way web standards, once taken into account by a library, may improve access by making resources more visible, readable and reusable. Data.bnf.fr thus perfectly responds to one of the primary missions of libraries. Moreover data itself is open, which gives the BnF a position within a globalized ecosystem, likely to allow for new crowdsourced production methods. This experimental starting point must be furthered toward new services (for instance developed SPARQL queries). Meanwhile, specific issues will be considered to ensure data's longevity and legal reuse.

Keywords: web of data, semantic web, open access, collaborative knowledge, open data

Introduction

Launched in 2011 by the National Library of France (BnF), data.bnf.fr aims to offer a unique entry point, from the web, to all BnF resources ([Gallica](#), the [general catalogue](#), [BnF archives and manuscripts](#) etc.). It was implemented within a twofold context.

First, to comply with the web's dynamic environment, data.bnf.fr relies on technologies and formats more open than those traditionally used by libraries and that conform to web standards. This strategy enables search engine optimization: BnF resources reach out internet users within their navigational patterns.

Second, data.bnf.fr meets Open data challenges, in line with the global Open Data movement and especially the French Ministry of Culture and Communication policy for openness and sharing of cultural public data. Data.bnf.fr can be considered the first visible move of the BnF towards Open Data and the BnF was the first cultural institution to join the French data.gouv.fr program.

A global innovative approach has led to create this BnF service, which uses semantic web technologies to smoothly experience, on a limited scale, how to match the requirements of data description and interoperability. Through the “learning by doing” method, the BnF aimed to test new data presentations in order to enrich its other services in return, especially its catalogue. Proof of concept to begin with, data.bnf.fr obtained in 2013 the [innovation prize](#) awarded by the libraries of Stanford University, along with Gallica (the BnF digital library). Though the service is currently thriving and is now managed by a dedicated team, it still proves a fertile ground for experiments.

To what extent is the web of data an opportunity for libraries' public service mission? How can the high value of library data be revealed? Thanks to data.bnf.fr, as it'll be shown below, one of librarians' primary missions can be achieved in a renewed way – to disseminate content as close as possible to users. Data.bnf.fr also enriches the BnF services by extending both accessibility and resources actually provided, and by reusing its assets (librarians' longstanding expertise, structured data, ARK persistent identifiers and links within its catalogues). Finally, it gives the BnF a place in a knowledge environment thoroughly changing, where communities co-create and reuse content. Due to their original position in the web ecosystem, libraries may thus consider specific issues.

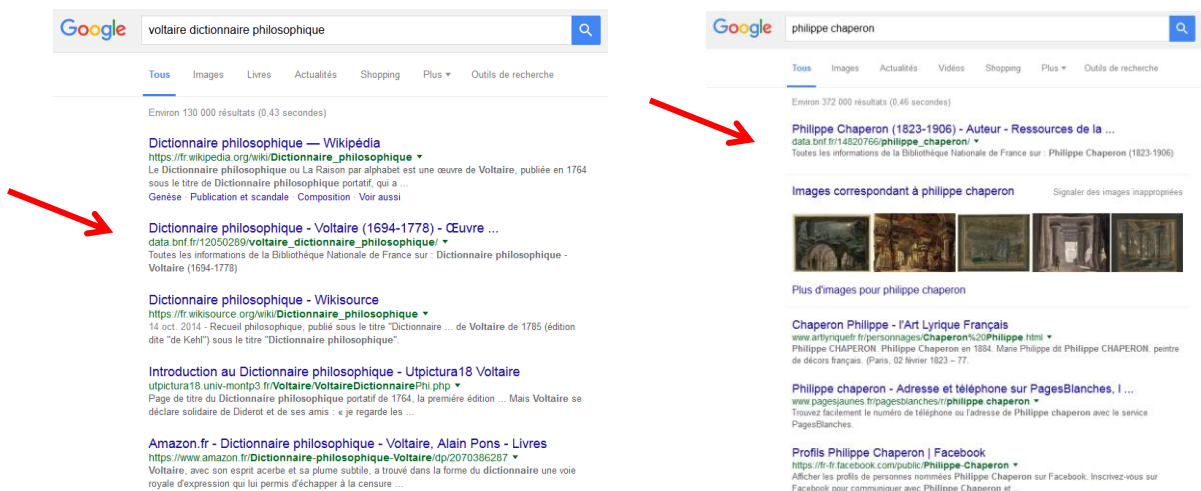
Data.bnf.fr makes the BnF resources more visible and readable, beyond its usual scope

By displaying BnF data and resources on the internet users' path and allowing them to find its content with no need to identify the Library as a specific source, data.bnf.fr aims to fully adapt the BnF to the web environment. Its design is “in line with librarianship's key principles: providing content and links towards documents through an easily findable and user-friendly service”¹.

The website itself brings together several data silos from the BnF and its partners (the general catalogue, Gallica etc.). Moreover, it merges and opens datasets. Structured therein following web standards, this data gets fitter for search engines. Above all, through such a technical and scientific measure, the BnF wishes to be found. Data.bnf.fr acts as an invisible gateway that allows internet visitors to enter the Library: the BnF reaches out to its users. 82% of

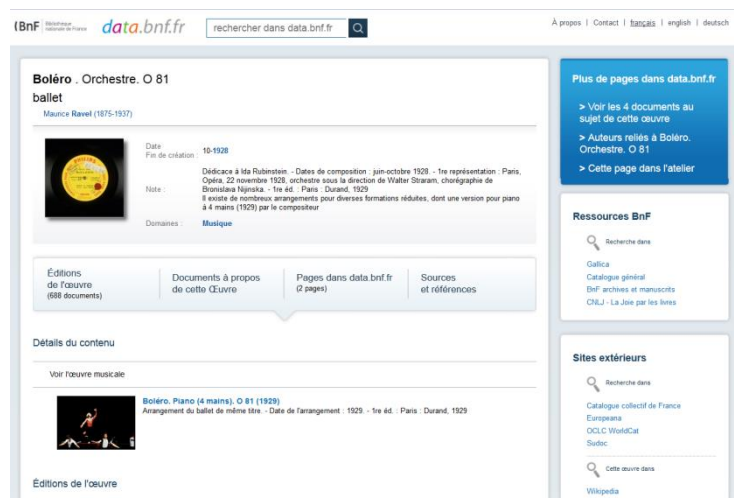
¹ Agnès Simon et Romain Wenz, *BBF*, 2012

data.bnf.fr visits come from a web search engine and 2/3 of the 13,000 individual visitors per day also visit another BnF application. This service engages an increasing audience: in 2015, data.bnf.fr cumulated 3.65 million visitors, twice as much as in 2014, and a considerably larger audience than that currently enjoyed by libraries.



Two Google queries proving data.bnf.fr's high ranking

What is also at stake is to evolve along with usage by challenging catalogues and making information they provide more readable. Data.bnf.fr is not only a service intended to offer datasets in linked data format, but also wishes to propose simple html web pages, whose intellectual and formal scheme allows internet users unfamiliar with library information management to easily understand content. Information is organised according to three core concepts (works, authors and subjects) in a way that complies with FRBR², leading to innovative reference pages. "Data from various databases (BnF general catalogue, BnF archives and manuscripts, Gallica)" is unified "according to concepts (works, authors and subjects) thanks to alignment and clustering algorithms"³. Authority data (about people, corporate bodies, works, subjects and places) acts as a hub in such an information rearrangement. This strategy, taking a fresh look at the traditional display of information in library catalogues, seeks to make it more attractive.



A « work » page on data.bnf.fr: the *Boléro* by Maurice Ravel

² Functional Requirements for Bibliographic Records : <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

³ Agnès Simon et Romain Wenz, *BBF*, 2012

In this respect, data.bnf.fr is the starting point of a more global rethinking of the BnF's bibliographical activity. As a result, the other catalogues' search options and user-friendliness are improved. "The developments of data.bnf.fr contribute (...) to organize the BnF general catalogue's data according to the FRBR model, especially to gather the different manifestations related to the same work: a data processing tool able to automatically calculate such clusters was implemented in late 2015. (...) Data.bnf.fr was thus given an additional mission – to mass reprocess bibliographic data for the benefit of other applications."⁴

The screenshot shows the BnF Catalogue général search results for 'albert camus'. The search results are displayed in a list format, showing the following entries:

Tri par	Défait	1 sur 185	10 résultats/page
1	Foley, John (1974-...) Albert Camus : from the absurd to revolt / John Foley McGill-Queen's university press Livres	2008	
2	Centre culturel Albert Camus (Tananarive) [Centre culturel Albert Camus] Centre culturel Albert Camus Livres	1983	
3	Rencontres internationales Albert Camus [Recueil. Documents d'information / Rencontres internationales Albert Camus] Documents d'archives et manuscrits	1984	
4	Albert Camus. une pensée, une oeuvre : colloque de Lourmarin, 1er-10 août 1985 / [organisé par les] Rencontres méditerranéennes de Lourmarin Rencontres méditerranéennes de Lourmarin Livres	1986	
5		1960	

Thanks to data.bnf.fr, the FRBR model was implemented in the BnF general catalogue for the first time. Since 2016, you may narrow searches by restricting them to works

Furthermore, catalogue data and BnF applications are linked to external data⁵. By providing links to other BnF resources (Gallica) and beyond (Europeana, WorldCat, Wikipedia etc.), data.bnf.fr proves to be user-friendly: useful information comes out on the user's path within a single website⁶. Thanks to this content visibility strategy, digitized documents get better highlighted. It enables both libraries and their content to be reached beyond their usual borders.

The data.bnf.fr openness encompasses not only the interface, but also data itself, which represents an enriched service.

Data.bnf.fr enriches the BnF offer

At the end of 2015, every BnF authority record checked by a librarian (about 2.4 million) and more than 8.5 million bibliographic records linked to these authority records had been made available on data.bnf.fr, which represents 70% of the BnF general catalogue's records⁷.

⁴ BnF Annual Report 2015

⁵ Jérôme Villeminez, *Ar(abes)ques*, 2016

⁶ Agnès Simon et Romain Wenz, *BBF*, 2012

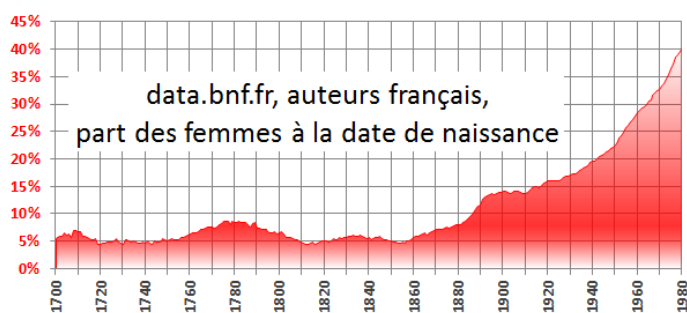
⁷ BnF Annual Report 2015

By constantly increasing the volume and diversity of the resources it makes available⁸ as well as by redefining access and content actually provided by libraries, data.bnf.fr is committed to a significant development of the Library offering.

Datasets are exposed on the web in RDF. Structured data is freely reusable and linkable: people can point at it and it's itself linked to other data. This service fully meets the requirements of the 5-star deployment scheme suggested by Tim Berners-Lee, the inventor of the World Wide Web, to assess the quality of open data.

Through this openness strategy, the BnF wishes to strengthen its engagement in transparency and enables the reuse of cultural heritage resources, thus fostering creation and knowledge. Since 2014, the metadata published via data.bnf.fr is placed under the public open licence recommended by the French Government Open Data mission (Etalab) for its data.gouv.fr program⁹. It encourages reuse and reproduction, including for commercial purposes. On a technical level, the BnF ensures constant improvement of “data download and reuse options. In particular, a more user-friendly version of the query service on RDF open data was implemented in 2015”¹⁰.

This open data opportunity has enabled Frédéric Glorieux, an IT engineer involved in the OBVIL LabEx¹¹ (Paris), to experiment text and data mining in literature domain. In his research log *J'attends des résultats*, he explains about the BnF “authors” authorities that demographic methods may be used to analyse this information as it includes birth and death date and offers a partial census of cultural life – so to speak. According to him, thanks to these essential figures, literary historians won't study only great authors but also overlooked elements in the background¹².



French female authors ratio according to birth date.
Frédéric Glorieux, « Historic demography of French authors », Research log *J'attends des résultats*

Such SPARQL¹³ queries results may lead to other explorations and analyses, for instance on all the editions of a work, on the works that have been adapted for young people, and on portraits of authors etc. Data.bnf.fr can be queried in SPARQL via a dedicated web page: <http://data.bnf.fr/sparql/>

⁸ BnF Annual Report 2015

⁹ This licence is an attribution licence, similar to CC BY or ODC-BY.

¹⁰ BnF Annual Report 2015

¹¹ Cluster of excellence combining the best research teams and laboratories in a field

¹² <http://resultats.hypotheses.org/518>

¹³ SPARQL is a RDF query language used in the semantic web world.

This open data approach and especially SPARQL queries rely on international standard identifiers (ISBN for books, ISNI for people and organisations), which enable the BnF to better take part in an evolving knowledge economy. Data.bnf.fr responds to the challenge libraries are currently facing: to get involved in a web environment generating crowdsourced data and collaborative knowledge.

Thanks to data.bnf.fr, new links and relationships can be created, supporting collaborative knowledge

New relationships with various professional communities can emerge, along with new cooperation opportunities with producers of all types of data (statistic, geographical, scientific etc.).

First, open data represents a major lever for the library community itself as it supports increasing work with museum and archive communities. Shareable and reusable data ensures a better bibliographic resource offering, both at a national and international level, all the more if libraries and partner organisations can collect and use it via new data exchange protocols. The data.bnf.fr project thereby contributes to a wide-scale data production workflow, saving time and effort. For example, in Switzerland, the RERO network¹⁴ conceived a RAMEAU-based¹⁵ project allowing for automatic indexing of its resources. The data production process is also being redefined at the BnF, which undergoes a worldwide move in this domain like many other cultural data providers: records used to be standardized in order to be exchanged whereas data is structured to be shared and combined. Via data.bnf.fr, the BnF supplies an international data flow, beyond the library community.

Library data may be shared with other communities for educational purposes: for instance, the virtual workspace AbulÉdu benefits from the BnF librarians' work. It was created in 1998 by teachers and computer scientists. They launched a collectively enriched, free French-speaking educative resources database that relies on RAMEAU subject headings.

More generally, data from data.bnf.fr is reused within portals and knowledge bases co-managed by several stakeholders and unifying several sources. [Isidore](#)¹⁶ is a one-stop access portal to digital data in the humanities and social sciences. It complies with the principles of the semantic web and uses the RAMEAU thesaurus for word completion and resources automatic indexing: “When one or more words are entered, linked by an implicit AND, a list of search suggestions appears, giving keywords from the Rameau and Pactols thesauruses and authors' names”¹⁷. The free and open knowledge base [Wikidata](#) also benefits from data.bnf.fr. This base “acts as central storage for the structured data of its Wikimedia sister projects”¹⁸, and supports many other websites. Data matching allows for identifiers alignment as well as enrichments such as additional information regarding the language, references statements, etc. Such openness engages new stakeholders and ensures a more transversal approach of knowledge.

¹⁴ Library Network of Western Switzerland

¹⁵ RAMEAU is an authoritative repository for subject headings managed by the BnF

¹⁶ <http://www.rechercheisidore.fr/>

¹⁷ https://www.ined.fr/en/resources-methods/selected_links/portals-web-directories/isidore/

¹⁸ https://www.wikidata.org/wiki/Wikidata:Main_Page



Proof of concept designed by Benoît Deshayes within Wikidata framework: different places evoked online in travelers' writings in French language have been mapped together. Gallica, Wikisource, Rosalis and Internet Archive are pointed at.

Due to its special position in the web environment, the BnF considers issues such as the longevity either of data itself or of the links created. Data.bnf.fr guarantees that resources available on the web via this media will remain so. Those who reuse this data are encouraged to trust libraries as they're assured that the links they have added won't be deleted if the pointed website is rebuilt. In the back office, the information automatic processing implies "a long term policy regarding structuration of data and authority records construction"¹⁹.

In the examples presented above, very large amounts of data need to be processed. The BnF keeps developing data.bnf.fr to match this requirement.

Conclusion

Since data from data.bnf.fr is freely reusable, it's difficult to precisely survey all types of reuse made from these growing datasets. These few examples nevertheless suggest their variety and the challenges they carry: what is at stake is to create circulations and networks, and not only to make resources more readable. From the "silo" whence it first flows away, data bolsters a significant ecosystem. The library role evolves towards new data co-production schemes with ever diversified professional communities. Awareness must be raised on the fact that content and expertise aren't enough any more: it is essential to provide a user-friendly environment, for instance by defining a legal framework for an open model, in order to allow people to appropriate, disseminate and reuse library resources. Finally, one data.bnf.fr upcoming development will consist in improving content (or data) curation – another essential issue – and the BnF intends in particular to create new insights thanks to datavisualisation methods in order to improve navigational patterns.

¹⁹ Agnès Simon et Romain Wenz, *BBF*, 2012

References

BnF, *BnF Annual Report 2015*, <http://webapp.bnf.fr/rapport/index.html> [Key facts and figures in English: http://webapp.bnf.fr/rapport/pdf/rapport_2015_synthese_en.pdf]

“La BnF et le numérique : entretien avec Sylviane Tarsot-Gillery”, *Lettre du COEPIA*, n° 28, february 2015, http://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/02/lettre_du_coepia_no28_fevrier_2015.pdf

Illien, Gildas, “Are you ready to dive in? A case for Open Data in national libraries”, 2012, <http://www.ifla.org/past-wlic/2012/181-illien-en.pdf>

Simon, Agnès and Wenz, Romain, “Des outils automatiques pour le signalement en bibliothèque”, *Bulletin des bibliothèques de France (BBF)*, n° 5, 2012, p. 39-43, <http://bbf.enssib.fr/consulter/bbf-2012-05-0039-008>

Stanford Prize for Innovation in Research Libraries (SPIRL): Application from the Bibliothèque nationale de France (BnF) for Gallica (gallica.bnf.fr) and Data (data.bnf.fr), 2012, <https://library.stanford.edu/sites/default/files/Bibliotheque%20nationale%20de%20France.pdf>

Villeminoz, Jérôme, “Le web de données à la BnF”, *Ar(abes)ques*, n° 83, 2016

Vincent Boulet, “L’infrastructure de l’information et les besoins des utilisateurs : tout le pouvoir aux données structurées!”, 2012, <http://www.ifla.org/past-wlic/2012/80-boulet-fr.pdf>

Wenz, Romain, “Linked open data for new library services: the example of data.bnf.fr”, *Italian Journal of Library, Archives, and Information Science (JLIS.it)*, *Global Interoperability and Linked Data in Libraries: Special issue*, vol. 4, n° 1, 2013, <http://leo.cineca.it/index.php/jlis/article/view/5509/7919>