

## Data.bnf.fr

### *Une application open data à la BnF<sup>1</sup>*

MAXIME BARRÉ

Depuis plusieurs années, les bibliothèques investissent le web afin de mettre leurs documents à la disposition des internautes. Il s'agissait au départ de rendre accessibles les catalogues sur le web, en facilitant la consultation à distance. Cette première étape, entreprise dès les années 1990 au travers des OPAC (*online public access catalogues*), a permis une diffusion plus large en donnant la possibilité aux lecteurs de préparer leur venue en bibliothèque, et de disposer gratuitement des diverses références. Mais cette démarche voit ses possibilités s'accroître : les techniques du web sémantique, avec les traitements automatiques et la création de liens, facilitent l'exposition des informations de façon plus pertinente, plus simple et plus rapide. Le mouvement d'ouverture des données publiques (*open data*) constitue une opportunité nouvelle de diffusion gratuite et ouverte des ressources auprès du plus grand nombre.

---

1. Pour des raisons éditoriales, les coordinateurs ont dû réduire le texte proposé par Maxime Barré avec son autorisation.

## Mouvement international de l'open data en bibliothèques

Des collectifs comme le « Linked Open Data in Libraries, Archives, and Museums » (LOD-LAM)<sup>2</sup> et l'« Open bibliography » projet piloté par L'Open Knowledge Foundation, œuvrent à la promotion du Linked open data. Ils offrent des outils, un cadre juridique et fixent les principes<sup>3</sup> de l'open bibliographic data.<sup>4</sup>

Depuis peu de temps, plusieurs organismes culturels mettent à disposition leurs données, citons la Library of Congress<sup>5</sup>, la British Library<sup>6</sup>, Europeana<sup>7</sup>, le musée d'Amsterdam<sup>8</sup>, etc. La Bibliothèque nationale de France (BnF) a décidé de s'inscrire dans ce mouvement, en valorisant le caractère spécifiquement culturel et patrimonial de ses données. Le portail *data.bnf.fr* a vu le jour en juillet 2011.

Quel type de données, ces institutions culturelles peuvent-elles mettre à disposition ?

Il s'agit de données bibliographiques ou des données descriptives, souvent appelées « métadonnées », qui servent à identifier et à gérer les collections de documents. Cette distinction entre les « métadonnées » et les « documents » est essentielle. Elle correspond à une question d'accès et à des conditions juridiques. À titre d'exemple, les métadonnées de *data.bnf.fr* sont réutilisables gratuitement<sup>9</sup>, même dans un cadre commercial. Ce qui n'est pas le cas des œuvres numérisées dans *Gallica*. Ils dépendent de la loi n° 78-753 du 17 juillet 1978 qui autorise les utilisations à but non lucratif des œuvres, mais prévoit une rémunération de l'institution en cas d'utilisation commerciale. Cette distinction entre métadonnées et contenus est liée à une vision de long terme : les données

---

2. <http://lodlam.net/about/>

3. Principles on Open Bibliographic Data. <http://goo.gl/ovLRj>

4. Le projet Open bibliography est piloté par un groupe de travail de l'Open Knowledge Foundation, association à but non lucratif fondée en 2004 dont le but est de promouvoir le savoir libre.

5. LC Linked Data Service. <http://id.loc.gov/>

6. Metadata services. <http://www.bl.uk/bibliographic/datafree.html>

7. Europeana Linked Open Data (LOD). <http://pro.europeana.eu/linked-open-data>

8. Open data. <http://www.amsterdammuseum.nl/open-data>

9. Conditions d'utilisation. <http://data.bnf.fr/licence>

descriptives servent à fournir un accès auprès du plus grand nombre, tandis que l'usage des documents numériques par le public doit pouvoir être suivi pour que les bibliothèques puissent rendre compte de leur action.

### Data.bnf.fr, le choix des outils du web sémantique

Pour être aussi visibles sur le web que dans la ville, les bibliothèques doivent diffuser leurs données de manière efficace, donc s'appropriier de nouvelles techniques. Comment valoriser le travail unique qu'elles effectuent dans la signalisation de l'information ? La BnF, en plus de *Gallica* et de son catalogue général, a créé data.bnf.fr qui, en sus de permettre aux internautes de télécharger les métadonnées, offre une nouvelle approche pour consulter les ressources de la BnF. Ce portail s'inscrit dans une politique d'ouverture des données publiques qui vise à rendre accessibles et réutilisables les données descriptives donnant accès à tous types de documents : livres, manuscrits, archives, ou bibliothèque numérique. Il répond à plusieurs objectifs : le regroupement des données bibliographiques depuis diverses bases, l'alignement sur des sources extérieures, et la diffusion des données auprès du plus grand nombre.

Pour ce faire, data.bnf.fr utilise les outils du *Linked Open Data (LOD)* ou web de données<sup>10</sup> ouvertes. Elle vise « à favoriser la publication de données structurées sur le web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations »<sup>11</sup>. Les chargés du projet ont fait le choix du format RDF<sup>12</sup> (*Resource Description Framework*) et opté pour les standards du web sémantique par l'utilisation de bonnes pratiques pour la diffusion des données bibliographiques. C'est un cadre technique permettant de décrire les ressources de façon fiable et interopérable en attribuant aux documents des identifiants uniques qui sont exposés sur le web. Ce format améliore le service rendu au public en offrant une visibilité aux informations enfouies dans des bases de données. *Data.bnf.fr* s'inscrit dans le mouvement du *Linked Open Data*, c'est-à-dire de données ouvertes liées.

---

10. Le web de données à la BnF : data.bnf.fr. <http://goo.gl/3QJRy>

11. BnF, (2012). Web sémantique. <http://goo.gl/nDZqb>

12. *Resource Description Framework (RDF)*. <http://www.w3.org/RDF/>

Le projet essaie le mieux possible de correspondre aux exigences édictées par Tim Bernes-Lee, un des fondateurs du W3C<sup>13</sup>. De ce fait, les relations entre les auteurs peuvent ainsi être exprimées, tout comme les différents intervenants dans l'existence d'une œuvre, un préfacier ou un traducteur. Par exemple, les pages « thèmes » de data.bnf.fr sont alignées sur le répertoire des thèmes de la Library of Congress. Un internaute qui recherche tous les ouvrages sur le thème « Italie-conditions économiques et sociales au XX<sup>e</sup> siècle » peut également effectuer cette recherche sur les ouvrages indexés par la Library of Congress sur ce même thème. Les avantages de l'alignement sont multiples : les pages *Wikipédia* des auteurs sont également reliées à data.bnf.fr par ce mécanisme.

L'investissement de la BnF sur le web à travers de nombreuses applications comme le catalogue général, le dépôt légal web, ou la base Bibliothèques Archives et Manuscrits entraîne une diversité des points d'accès aux ressources. Cette organisation permet à un internaute de trouver tous les ouvrages d'un auteur ou bien toutes les éditions d'une œuvre. Data.bnf.fr permet de centraliser les différentes données de la bibliothèque, et de lier les métadonnées et documents numérisés puisque l'adresse du document numérisé est attachée à sa description bibliographique. Il s'agit donc d'extraire les données des bases et de permettre une recherche sur l'ensemble des données produites et possédées par la bibliothèque.

### Des données ouvertes

L'internaute peut consulter des ressources, des catalogues, des inventaires d'archives et de manuscrits et des documents numériques. Les métadonnées décrivant un livre, son auteur, son éditeur, etc. sont extraites des bases pour être accessibles sur une plateforme ouverte. La BnF a fait le choix de la Licence ouverte promue par Étalab<sup>14</sup>.

Si les données de la BnF s'adressent en priorité aux professionnels des bibliothèques, afin que ceux-ci puissent, par exemple, construire plus facilement un catalogue, leur utilisation n'est pas limitée à une sphère

---

13. Berners-Lee, Tim (2006). Linked data. <http://goo.gl/Hsj2I>. À ce propos voir également dans ce numéro l'article de Denis Berthault et Bernard Marx [ndlr].

14. Étalab. Licence ouverte/Open licence. <http://goo.gl/GTbe7>

particulière. La diffusion des données auprès du plus grand nombre repose sur un double mécanisme : d'une part, la publication de pages web (HTML) simples à utiliser et structurées en incluant des balises qui indiquent où se trouve l'information et quelle est sa nature<sup>15</sup> ; d'autre part, la fourniture de données en RDF, ouvertes techniquement et juridiquement, qui permettent à des tiers de réutiliser librement l'ensemble des informations du site.

## Conclusion

Par l'utilisation de ces technologies, data.bnf.fr propose une ouverture et une visibilité à une échelle nouvelle, celle d'internet. C'est en exposant ses données sur le web qu'elle peut répondre aux attentes de ses usagers. Le monde des bibliothèques s'est distingué par l'utilisation de nombreuses technologies structurées, mais utilisant des formats spécifiques. Les protocoles d'échanges des données des bibliothèques sont tout à fait efficaces et adaptés, mais l'usage de ces outils est nécessairement restreint. Le *Linked Open Data* libère les données des silos fermés des bibliothèques, mais ouvre également vers d'autres institutions : la BnF fait le pari que l'usage de ces technologies se répandra de plus en plus largement. D'autres bibliothèques, comme nous l'avons vu auparavant, ont de la même façon fait ce choix.

Les technologies sémantiques reposent sur l'accessibilité des données : data.bnf.fr est un des points de rencontre entre le web sémantique et le mouvement d'ouverture des données publiques.

*Les coordonnateurs de ce numéro tiennent à remercier Romain Wenz, Chef de projet data.bnf.fr du Département de l'Information bibliographique et numérique de la BnF pour sa relecture.*

---

15. À propos de data.bnf.fr. <http://data.bnf.fr/about>

