



HAL
open science

Améliorer l'accès aux données des bibliothèques sur le web : l'exemple de data.bnf.fr

Emmanuelle Bermès, Vincent Boulet, Céline Leclaire

► To cite this version:

Emmanuelle Bermès, Vincent Boulet, Céline Leclaire. Améliorer l'accès aux données des bibliothèques sur le web : l'exemple de data.bnf.fr. IFLA WLIC 2016 – Columbus, OH – Connections. Collaboration. Community in Session 201 - National Libraries, International Federation of Library Associations and Institutions (IFLA), Aug 2016, Columbus, États-Unis. hal-01393255

HAL Id: hal-01393255

<https://bnf.hal.science/hal-01393255>

Submitted on 7 Nov 2016

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.

Améliorer l'accès aux données des bibliothèques sur le web : l'exemple de data.bnf.fr

Emmanuelle Bermès

Direction des Services et des réseaux, Bibliothèque nationale de France, Paris, France.

emmanuelle.bermes@bnf.fr

Vincent Boulet

Département des Métadonnées, Bibliothèque nationale de France, Paris, France.

vincent.boulet@bnf.fr

Céline Leclaire

Direction des Services et des réseaux, Bibliothèque nationale de France, Paris, France.

celine.leclaire@bnf.fr



Copyright © 2016 by **Emmanuelle Bermès, Vincent Boulet, Céline Leclaire**. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

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

Résumé :

Cet article présente data.bnf.fr, un service créé par la Bibliothèque nationale de France en 2011 qui lui permet d'une part de mieux référencer ses données pour les positionner là où le public passe, et d'autre part de rassembler autour de pages « auteur », « œuvre » et « thème » les ressources de ses autres applications ([Catalogue général](#), [Gallica](#), etc.).

À travers cet exemple, il s'agit de mettre en évidence la façon dont l'adoption des standards du web améliore l'accès aux ressources en les rendant à la fois plus visibles, lisibles et réutilisables. En cela, data.bnf.fr contribue à remplir l'une des missions les plus élémentaires des bibliothèques. Mais il va plus loin en ouvrant l'accès aux données elles-mêmes, ce qui inscrit la BnF dans un écosystème redéfini, propice à de nouvelles formes de partage. Cela implique de continuer à imaginer de nouveaux services à partir de ce point de départ (par exemple en développant les requêtes en SPARQL), et à envisager des sujets de préoccupation spécifiques tels le cadre légal des réutilisations ou la pérennité des données.

Keywords: web de données, web sémantique, *open access*, économie du savoir participative, ouverture des données

Introduction

Créé en 2011 par la Bibliothèque nationale de France (BnF), data.bnf.fr offre un point d'entrée privilégié, depuis le web, à l'ensemble des ressources proposées par l'institution ([Gallica](#), le [Catalogue général](#), [BnF archives et manuscrits](#)...). Son développement s'inscrit

dans un double contexte. En reposant sur des technologies et des formats plus ouverts que ceux utilisés traditionnellement dans les bibliothèques, et en faisant le choix de standards conformes aux standards du web, data.bnf.fr se situe tout d'abord dans la dynamique du web : il s'agit de favoriser le référencement par les moteurs de recherche des données de la BnF afin de les placer sur le chemin des usages des internautes. . Data.bnf.fr répond également aux enjeux de l'open data, à la politique d'ouverture des données publiques culturelles portée par le Ministère de la Culture et de la Communication et, plus généralement, au mouvement de l'*open data* : grâce à data.bnf.fr – premier signe tangible de l'engagement de la BnF dans la voie de l'*open data* – la BnF fut la première institution culturelle à déposer ses données sur data.gouv.fr.

La démarche globale qui a conduit à la création de ce service relève de l'innovation : data.bnf.fr s'est construit dans le but de mettre en pratique, en douceur et à une échelle raisonnable, les réflexions engagées autour de la description et de l'interopérabilité des données dans l'univers du web sémantique. Il s'agissait d' « apprendre en faisant », de tester ici des formes de présentation pour mieux façonner en retour les autres applications de la BnF, à commencer par le catalogue. Preuve de concept à l'origine, data.bnf.fr a obtenu en 2013, avec Gallica (la bibliothèque numérique de la BnF), le [prix de l'innovation](#) décerné par les bibliothèques de l'Université de Stanford. Il continue de surcroît à donner prise à des explorations et expérimentations alors même qu'il a atteint sa pleine maturité et bénéficie aujourd'hui d'une équipe dédiée.

En quoi le web de données est-il une opportunité pour la mission de service public des bibliothèques ? En quoi les données catalographiques d'une bibliothèque peuvent-elles constituer une valeur ? Data.bnf.fr, nous allons le montrer, représente un moyen renouvelé d'honorer l'une des missions fondamentales liées au métier de bibliothécaire : diffuser des contenus au plus près des publics. En reconsidérant les notions d'accès et les contenus effectivement fournis par la Bibliothèque, en réutilisant les atouts des pratiques professionnelles et des données structurées, des identifiants pérennes ARK et des liens présents dans les catalogues de la BnF¹, data.bnf.fr contribue aussi à élargir le panel de ressources et de services proposés. Enfin, cela positionne la BnF dans un écosystème de la connaissance redéfini, caractérisé par des formes collaboratives de production du savoir et par des réutilisations de contenus. Cette inscription originale dans l'univers du web incite les bibliothèques à envisager des sujets de préoccupation spécifiques.

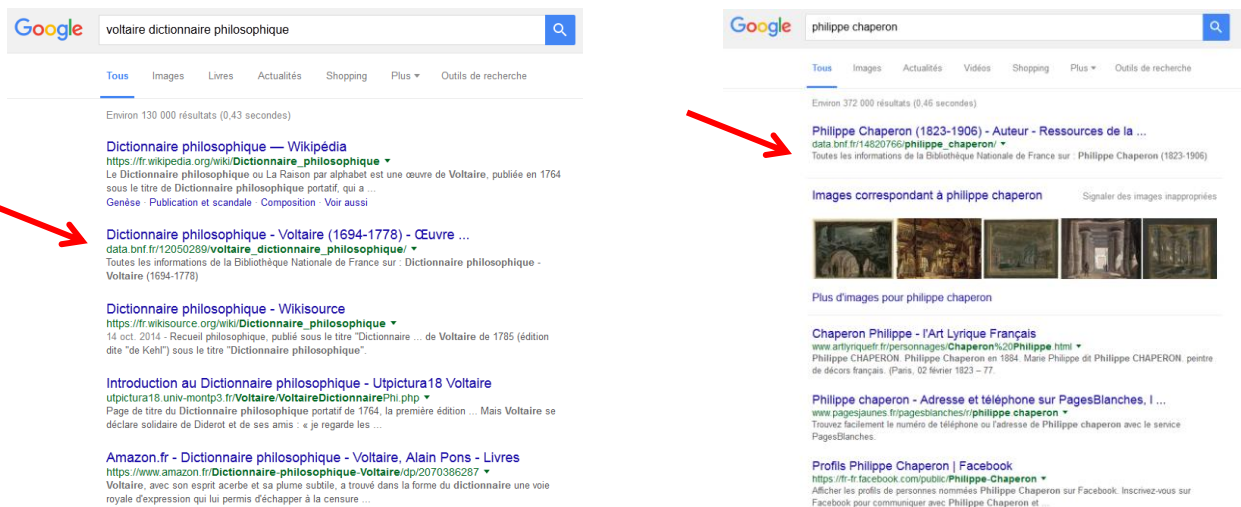
Utiliser les standards du web pour sortir de son périmètre : se rendre plus visible et lisible

Data.bnf.fr vise à mettre en relation les collections et ressources de la BnF avec le public là où il passe : il inscrit pleinement la bibliothèque dans l'environnement du web, en autorisant l'internaute à trouver les ressources de la bibliothèque sans avoir besoin d'identifier celle-ci en tant que source spécifique. Le produit « rejoint donc des principes fondamentaux du métier : fournir des contenus, des liens vers des documents, et offrir un service simple à trouver et à utiliser »².

¹ 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>

² Agnès Simon et Romain Wenz, *BBF*, 2012

Le site en lui-même est le point de rencontre de plusieurs réservoirs de données de la BnF et de partenaires extérieurs (catalogue général, bibliothèque numérique Gallica...), qui se trouvent ainsi décloisonnés. Ces données y sont structurées à l'aide des standards du web, ce qui en améliore le référencement. Ce choix technique et scientifique obéit tout d'abord à la volonté qu'a la BnF d'être trouvée. Data.bnf.fr est une porte d'entrée invisible dans la Bibliothèque : c'est la Bibliothèque qui vient à l'utilisateur. Ainsi, 82% des visites proviennent des moteurs de recherche, et deux tiers des quelque 13 000 visiteurs uniques par jour se redirigent vers les autres applications de la BnF. Cette offre rencontre une audience toujours en forte progression : avec un total de 3,65 millions de visites en 2015, l'audience de data a doublé par rapport à 2014. Ce succès permet de toucher un public plus large, au-delà des lecteurs de la Bibliothèque.

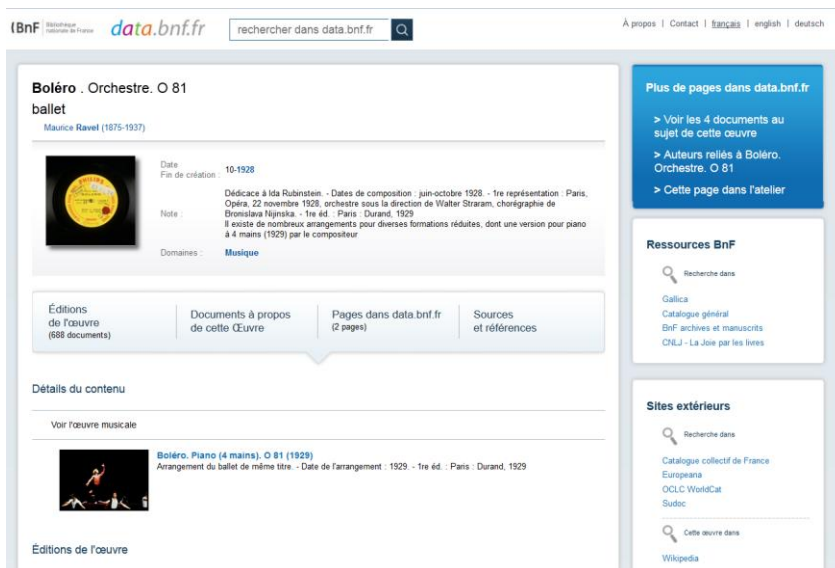


Deux exemples de recherche dans Google : data.bnf.fr apparaît en bonne place dans la liste des résultats.

Un deuxième enjeu consiste à s'adapter aux usages en repensant les catalogues, en rendant plus lisibles les informations proposées. Avant d'être un service en *Linked data* destiné à fournir des jeux de données, data.bnf.fr a pour vocation de proposer de simples pages web, en html, dont la structure aussi bien sur la forme que sur le fond favorise la lisibilité pour les internautes peu familiers de l'organisation des informations en bibliothèque. Ainsi, les informations sont organisées selon trois entrées principales – œuvres, auteurs, sujets – qui s'inspirent des principes de base du modèle FRBR³ et constituent autant de fiches de référence inédites. Les « données des différentes bases de données (BnF catalogue général, BnF archives et manuscrits, Gallica) » sont fédérées « autour de concepts (les œuvres, les auteurs et les sujets), grâce à des algorithmes d'alignement et de regroupement »⁴. Les données d'autorité (sur les personnes et les collectivités, sur les œuvres, sur les sujets, les lieux) jouent donc un rôle pivot dans cette réorganisation de l'information. On observe une prise de distance par rapport à l'organisation classique de l'information dans un catalogue de bibliothèque, dont l'objectif est de rendre l'information plus lisible, aisément compréhensible.

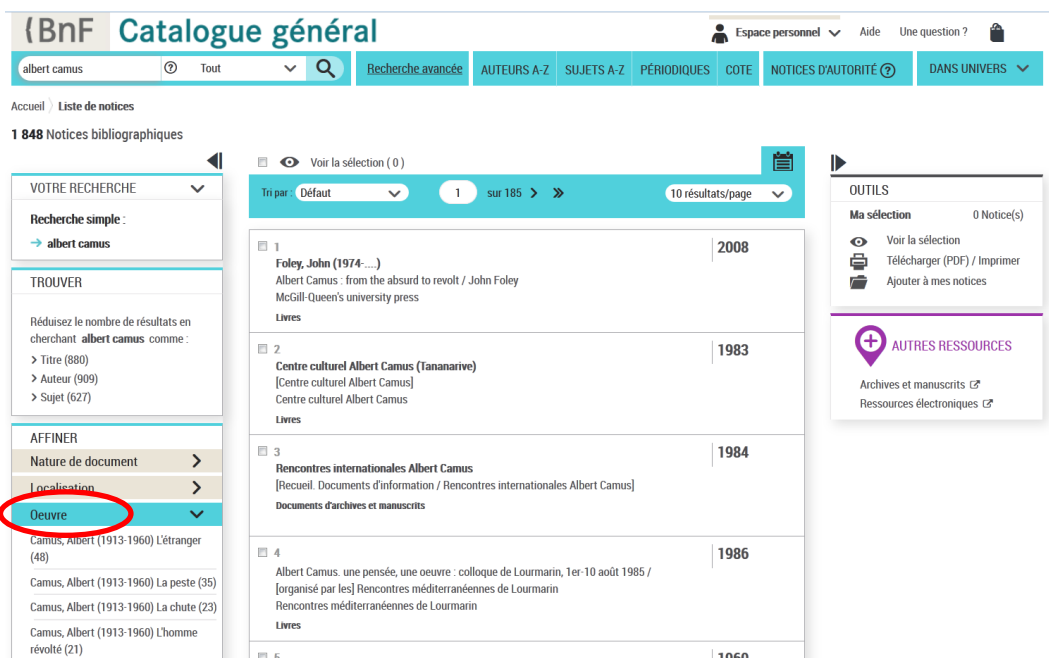
³ Functional Requirements for Bibliographic Records : <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

⁴ Agnès Simon et Romain Wenz, *BBF*, 2012



Une page « œuvre » de data.bnf.fr : le *Boléro* de Ravel

En cela, data.bnf.fr est également le point de départ d'une stratégie plus globale de refonte de l'activité bibliographique de la BnF, et concourt à l'amélioration de l'ergonomie de la recherche dans les catalogues. « Les développements de data contribuent (...) à la FRBRisation des données du catalogue général et notamment au regroupement des manifestations d'une même œuvre : une base de traitement des données du catalogue qui permet de calculer automatiquement de tels regroupements a été déployée fin 2015. (...) S'affirme ainsi autour de data une nouvelle mission, celle d'outil de retraitement de masse des données bibliographiques, au service d'autres applications. »⁵



Grâce à data.bnf.fr, le Catalogue général de la BnF a connu sa toute première FRBRisation. Depuis 2016, il propose d'affiner la recherche selon le filtre des œuvres.

⁵ Rapport d'activité 2015 de la BnF

Enfin, les données des catalogues et applications de la BnF sont liées à des données externes⁶ : l'ajout de liens vers d'autres ressources de la BnF (Gallica) ou extérieures à la BnF (Europeana, WorldCat, Wikipédia...) au sein même de data.bnf.fr vise à faciliter le parcours de l'utilisateur, qui rencontre sur son chemin les informations pertinentes dont il a besoin⁷.

Cette forme de visibilité met davantage en avant ce qui a déjà été numérisé. Elle permet également un décloisonnement des bibliothèques et de leurs contenus, qui ne sont plus enfermés dans un périmètre donné : la bibliothèque s'ouvre, va au-devant des utilisateurs, et son interface utilise les standards généraux du web.

Avec data.bnf.fr, le principe de l'ouverture s'applique non seulement à l'interface mais aussi aux données elles-mêmes, et cela représente une source d'enrichissement du service rendu.

Offrir davantage

Fin 2015, data.bnf.fr exposait l'intégralité des notices d'autorité validées de la BnF (près de 2,4 millions) et pas moins de 8,5 millions de notices bibliographiques rattachées à ces notices d'autorité, ce qui représente plus de 70 % des notices du Catalogue général⁸.

En veillant à accroître sans cesse le volume et la diversité des contenus qu'elle propose⁹ mais aussi en reconsidérant les notions d'accès et les contenus effectivement fournis par les bibliothèques, data.bnf.fr contribue à élargir le panel des ressources et des services proposés par la BnF.

Les jeux de données exposés sont en RDF, ces données sont structurées, librement exploitables, il est possible de pointer des liens vers elles et elles sont elles-mêmes liées à d'autres données : tous ces critères placent cette initiative tout en haut de l'échelle établie par Tim Berners-Lee, l'inventeur du World Wide Web, pour déterminer la qualité des données ouvertes¹⁰.

En s'inscrivant dans cette logique d'ouverture, la BnF témoigne d'une volonté de transparence et crée les conditions nécessaires à l'appropriation par tous de ses richesses, au nom du soutien à la création et du développement des connaissances. Depuis 2014, les métadonnées descriptives de data.bnf.fr sont diffusées sous licence Etalab, la licence ouverte de l'État français, qui encourage leur reproduction et leur réutilisation, y compris à des fins commerciales. Techniquement, la BnF veille à l'amélioration constante des « possibilités de récupération et de réutilisation des données. En particulier, une version plus ergonomique du service de requête sur les données ouvertes en RDF a été déployée en 2015, afin d'en faciliter la réutilisation par des tiers »¹¹.

Saisissant l'opportunité de ces données ouvertes, Frédéric Glorieux, ingénieur informatique au sein du LabEx OBVIL (Paris IV), applique la fouille de textes et de données au domaine

⁶ Jérôme Villeminoz, *Ar(abes)ques*, 2016

⁷ Agnès Simon et Romain Wenz, *BBF*, 2012

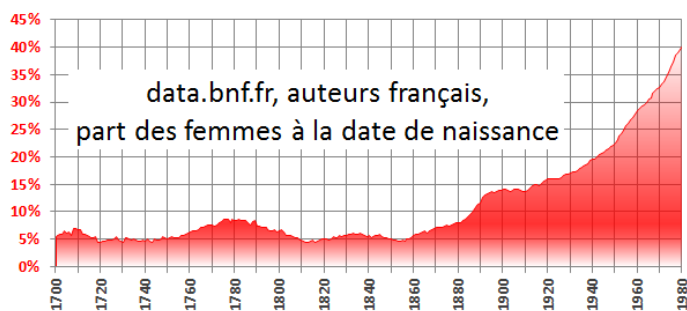
⁸ Rapport d'activité 2015 de la BnF

⁹ Rapport d'activité 2015 de la BnF

¹⁰ https://fr.wikipedia.org/wiki/Open_data

¹¹ Rapport d'activité 2015 de la BnF

de la littérature. Dans son carnet de recherche *J'attends des résultats*, il évoque les autorités « auteurs » de la BnF et explique : « Ces noms, avec souvent des dates de naissance et de mort, constituent une sorte de recensement partiel de la vie culturelle, auquel [peuvent] être [appliquées] les méthodes de la démographie. Ces indications chiffrées sont une base indispensable pour que l'histoire littéraire ne se limite pas à l'épopée des grands auteurs, et connaisse mieux le fond de la toile sur laquelle se détache[nt] les personnages. »¹²



Frédéric Glorieux, « Auteurs français, éléments de démographie historique », Carnet de recherche *J'attends des résultats*

L'exploration permise par les requêtes en SPARQL¹³ laisse imaginer de multiples autres analyses, par exemple à partir de toutes les éditions d'une même œuvre, des ouvrages adaptés pour la jeunesse, de portraits d'auteurs... Une page spécifique est dédiée à l'éditeur SPARQL de data.bnf.fr : <http://data.bnf.fr/sparql/>

Cette ouverture des données et ces requêtes en SPARQL, notamment parce qu'elles permettent de travailler à partir d'identifiants internationaux normalisés (ISBN pour les livres, ISNI pour les personnes...), ouvrent la voie à une meilleure inscription de la BnF dans un écosystème de la connaissance en pleine évolution. Data.bnf.fr constitue une réponse au défi qui se pose aux bibliothèques aujourd'hui : être parties prenantes d'un monde qui, par le web, construit des modèles alternatifs où la connaissance s'élabore grâce à des contributions diverses.

Créer des liens et co-construire

Se profilent de nouvelles formes de dialogue avec des communautés professionnelles très variées, et des perspectives de nouvelles coopérations avec les producteurs de données de tous types (statistiques, géographiques, scientifiques, etc.).

Cette ouverture représente d'abord un enjeu pour la communauté des bibliothèques elle-même, qui travaille de plus en plus avec les musées, les archives, etc. La facilité de transmission et de récupération des données permet en effet d'envisager une meilleure offre en matière de ressources bibliographiques à l'échelle nationale (une autre des missions clés de la BnF) et internationale, les bibliothèques et établissements partenaires pouvant les récupérer ou les utiliser plus facilement via de nouveaux prototypes d'échanges de données. De ce point de vue, l'initiative de data.bnf.fr conduit donc à une économie en matière de production de données à large échelle. Ainsi, RERO (réseau des bibliothèques de Suisse

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

¹³ Langage de requête utilisé dans le cadre du web de données.

occidentale) a conçu un projet d'utilisation du langage RAMEAU¹⁴ pour l'indexation automatique des ressources¹⁵. Ce type de projets amène également à repenser, à terme, les pratiques de production de données à la BnF, qui rejoint ainsi un mouvement international de basculement du centre de gravité des pratiques des producteurs de données culturelles. Nous sommes passés d'un monde où l'on normalisait des notices pour se les échanger à un monde où l'on structure des données pour se les partager. Data.bnf.fr inscrit la BnF dans un flux de données international, qui va au-delà du monde des bibliothèques.

Par exemple, la communauté éducative en tire également un bénéfice : le travail des professionnels de la BnF est par exemple utilisé par AbulÉdu, « un espace numérique de travail pour l'éducation et l'éducation populaire, conçu [en 1998] par des enseignants et réalisé en partenariat avec des informaticiens », qui a mis en ligne une « base de données de ressources éducatives libres francophones »¹⁶ alimentée collectivement. L'explication donnée en ligne est la suivante : « La BNF ouvrant son vocabulaire (RAMEAU) nous avons sauté sur l'occasion : au lieu de laisser libre cours à nos utilisateurs pour saisir les mots clés manuellement nous allons utiliser celui qui est fait par des professionnels du domaine (la BNF) ».

Plus généralement, data.bnf.fr profite à des portails et à des bases de connaissances fédérant plusieurs acteurs, plusieurs sources. [Isidore](http://www.isidore.fr/)¹⁷, un portail d'accès aux données numériques des sciences humaines et sociales, utilise RAMEAU pour la recherche dans le moteur d'autocomplétion et pour l'indexation automatique des ressources. Les données de data.bnf.fr sont également utilisées au sein de [Wikidata](https://www.wikidata.org/wiki/Wikidata:Main_Page), base de connaissances libre, ouverte, multilingue et collaborative, qui centralise les données structurées des projets de la fondation Wikimedia (en particulier de ceux de Wikipédia), mais sert aussi de nombreux autres sites, au-delà¹⁸. Le croisement des données permet des alignements d'identifiants et des enrichissements (sourçage de déclarations externes, ajout de nouvelles informations – par exemple la langue d'expression). Un tel décloisonnement permet donc à d'autres acteurs d'intervenir et donne lieu à des approches plus transversales du savoir.

¹⁴ <http://rameau.bnf.fr/>

¹⁵ http://www.bnf.fr/fr/professionnels/donnees_bnf_recuperables/a.donnees_rdf.html#SHDC__Attribute_BlocArticle9BnF

¹⁶ Voir <https://fr.wikipedia.org/wiki/Abul%C3%89du> et <http://data.abuledu.org/wp/>

¹⁷ <http://www.rechercheisidore.fr/>

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



Preuve de concept conçue par Benoît Deshayes dans le cadre de Wikidata : Carte des lieux évoqués dans les récits de voyage avec texte en français accessible en ligne. Les ressources pointées sont sur Gallica, Wikisource, Rosalis, Internet Archive...

Cette inscription originale dans l'univers du web incite la BnF à s'emparer de sujets de préoccupation spécifiques, telle la question de la pérennisation de cette entreprise et des liens permis. Data.bnf.fr permet d'envisager une continuité dans la mise à disposition des ressources par les bibliothèques sur le web et favorise une relation de confiance avec ceux qui réutilisent ces données, en particulier car ils sont assurés que les liens qu'ils ont établis ne seront pas rendus caduques par une refonte de site. En interne, l'automatisation du traitement de l'information par la machine implique « une politique de long terme de l'établissement sur la structuration des données et la construction de notices d'autorité. »¹⁹

Par ailleurs, les exemples cités supposent de travailler sur de très gros volumes de données : pour répondre à cette exigence précise, la BnF travaille constamment à l'amélioration de data.bnf.fr.

Conclusion

Puisque l'utilisation des données de data.bnf.fr est libre, il est difficile d'avoir une idée précise de tous les emplois qui en sont faits. Ces quelques exemples suggèrent néanmoins leur variété et les enjeux qui leur sont liés : non contente de rendre les ressources plus lisibles, data.bnf.fr crée des circulations. Un « réservoir » initial donne naissance à un véritable écosystème : le rôle de l'institution se modifie, elle permet d'imaginer à terme de nouveaux scénarios de co-production des données avec des communautés professionnelles encore plus diversifiées. Un enseignement pourrait en être tiré : les contenus et expertises ne sont plus suffisants, il faut aussi créer un contexte favorable à la réutilisation et à l'appropriation des ressources, par exemple en définissant un cadre légal pour un modèle ouvert. Surgit alors la question de la médiation, qui vaut autant pour les œuvres que pour les données : une des

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

perspectives actuelles de data.bnf.fr consiste en particulier à explorer les pistes offertes par la data visualisation pour améliorer la navigation dans ces données²⁰.

Références

BnF, *Rapport d'activité 2015* [à paraître en ligne]

« La BnF et le numérique : entretien avec Sylviane Tarsot-Gillery », *Lettre du COEPIA*, n° 28, février 2015. En ligne :

http://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/02/lettre_du_coepia_no28_fevrier_2015.pdf

Simon, Agnès et Wenz, Romain, « Des outils automatiques pour le signalement en bibliothèque », *Bulletin des bibliothèques de France (BBF)*, n° 5, 2012, p. 39-43. En ligne : <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. En ligne :

<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>

²⁰ Jérôme Villeminoz, *Ar(abes)ques*, 2016 : « Il reste également à inventer autour des jeux de données un véritable service aux utilisateurs, au-delà de la seule mise à disposition d'une infrastructure de récupération et d'interrogation : comme toute autre collection proposée par la bibliothèque, ces jeux de données méritent une médiation. Cette médiation passera d'ailleurs par l'évolution de l'interface de recherche et de visualisation, car on ne consulte pas des jeux de données comme on consulte les documents d'une bibliothèque numérique ou même les notices d'un catalogue ; les logiques d'accès et de navigation sont spécifiques, et il reviendra à data.bnf.fr d'expérimenter des techniques de data visualisation. »