

Le Web de données à la BNF

Dans quel sens faut-il le dire ? Est-ce la BNF qui s'intègre dans le web de données, ou le web de données qui gagne la BNF ? Il est certain en tout cas que le web de données fait penser et agir différemment, et que pour l'information bibliographique en général, il apparaît désormais comme un horizon naturel.

À la Bibliothèque nationale de France (BNF), l'intention est nette et affichée. Dans son *Contrat de performance pour les années 2014 à 2016*^[1], elle inscrit clairement son action bibliographique « dans le mouvement d'ouverture des données publiques (*open data*) comme dans le développement du web sémantique » (p.8), ainsi que dans « la politique d'ouverture des données publiques culturelles poursuivie par le ministère de la Culture et de la Communication. » (p.10). Une priorité est énoncée en appui de cette direction : le développement de data.bnf.fr^[2].

Le *Schéma numérique* publié en mars 2016^[3] cite lui aussi le web de données, conjointement au mouvement d'ouverture des données publiques, comme « une réelle opportunité technologique pour le développement des missions de service public qui incombent à la BNF » : l'entrée dans le web de données est à la fois l'occasion et le fruit d'un décloisonnement pour les bibliothèques ; elle rend nécessaire le dialogue avec d'autres communautés

professionnelles et ouvre la voie à de nouvelles coopérations, avec des opérateurs culturels comme des producteurs de données de tout type : statistiques, géographiques, scientifiques, etc. Au-delà de l'expression des données en RDF et de l'ouverture juridique, le *Schéma numérique* insiste particulièrement sur deux axes, dont l'association est présentée comme le socle de ce qui pourrait être un nouveau contrôle bibliographique universel : « la normalisation et la dissémination d'identifiants de confiance », et la mise à disposition des données d'autorité sous la forme de vocabulaires ou de référentiels.

Data.bnf.fr et l'exposition en RDF, les identifiants, les référentiels d'autorité : on retrouve là les trois modalités principales d'investissement de la BNF dans le web de données.

DATA.BNF.FR

À son lancement en 2011, data.bnf.fr était un projet expérimental, l'occasion pour la BNF de se frotter au web de données. Il était fondé sur l'intuition qu'avec les données d'autorité, les catalogues de bibliothèques disposent de référentiels pivots cor-

respondant parfaitement à la logique et à la structure du web de données. Ses objectifs étaient multiples : améliorer la visibilité des ressources de la BNF depuis les moteurs de recherche du web, favoriser la réutilisation des données et décloisonner les catalogues de la BNF, en les articulant entre eux et en liant leurs données avec des données externes. [Data.bnf.fr](http://data.bnf.fr) se présentait sous une double forme : des pages web visibles par les internautes, et des jeux de données en RDF destinés aux machines et à la réutilisation. Les premiers résultats se sont avérés encourageants, en termes d'audience, d'appropriation par les équipes et d'image institutionnelle. La réutilisation a quant à elle avancé notablement lorsque l'intégralité du répertoire d'autorité matière Rameau a été exposée et mise à disposition en Skos^[4] : la complétude de ce jeu de données et les perspectives d'utilisation

Data.bnf.fr et l'exposition en RDF, les identifiants, les référentiels d'autorité : on a là les trois modalités principales d'investissement de la BNF dans le web de données.

multilingue qu'offrait son alignement avec des référentiels anglo-saxon (LCSH^[5]) et allemand (SWD^[6]) ont sans nul doute joué en sa faveur.

Fort de ces bilans convaincants, data.bnf.fr a pu se développer jusqu'à exposer la totalité des données d'autorité validées de la BNF, et l'ensemble des données bibliographiques qui leur sont liées (soit 70 % du Catalogue général). Aujourd'hui, les 2,8 millions de pages de data.bnf.fr attirent 12 à 14 000 visiteurs uniques par jour, dont les deux tiers se redirigent vers les autres applications de la BNF. Désormais, après une phase marquée par un objectif principal de croissance, l'enjeu pour le site est de faire jouer à plein les ressorts du web de données, en développant les liens et les alignements de données par une politique active de partenariats. 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

[1] http://www.bnf.fr/documents/contrat_performances_2014.pdf

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

[3] http://www.bnf.fr/documents/bnf_schema_numerique.pdf

[4] *Simple Knowledge Organization System* : recommandation du W3C pour représenter du vocabulaire contrôlé

[5] *Library of Congress Subject Headings, autorités de la Bibliothèque du Congrès*

[6] *Schlagwortnormdatei*, thésaurus de la Bibliothèque nationale allemande (DNB)

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*.

RELIURES, BP16 ET DOREMUS

En complément de data.bnf.fr, la BNF a expérimenté la structuration en RDF et la publication dans le web de données pour deux sites spécialisés, *Reliures* et *BP16*. Le premier (reliures.bnf.fr^[7]) propose une sélection de reliures numérisées accompagnées de descriptions détaillées, et le second (bp16.bnf.fr^[8]) est une édition sous une nouvelle forme de la *Bibliographie des éditions parisiennes du 16^e siècle*. Ici encore, les résultats sont concluants, car le RDF offre des facilités de navigation interne et permet que ces sites, dont la spécialisation aurait pu être une cause d'isolement, soient intégrés à un écosystème dans lequel ils sont reliés à d'autres ressources. Cependant, cette expérience a aussi montré qu'elle serait dans l'immédiat difficile à rééditer, en raison du coût actuel de la maintenance et de la mise à jour de telles applications.

Enfin, la BNF s'est engagée dans le projet Doremus, aux côtés de partenaires culturels et scientifiques, afin d'explorer des modèles de données plus fins que ceux qui ont été utilisés pour data.bnf.fr. De fait les modèles choisis pour data.bnf.fr ont été volontairement simples, de façon à favoriser la réutilisation par des communautés du web. Le projet Doremus a des besoins de plus grande précision : il a pour objectif de doter les institutions culturelles musicales, les éditeurs, les distributeurs et les communautés de passionnés de modèles et de référentiels communs pour publier dans le web de données, partager et enrichir des catalogues d'œuvres et d'événements musicaux. C'est donc une extension du modèle FRBRoo (*Functional Requirements for Bibliographic Records – object oriented*) qui a été retenue ; sa souplesse permettra de gérer différents niveaux de granularité, jusqu'à restituer dans toute leur richesse les données de catalogues spécialisés.

ISNI

Du côté des identifiants, on peut indiquer l'investissement de la BNF sur l'ISNI, qui présente de nombreux atouts. Identifiant attaché aux identités publiques de créateurs, personnes ou collectivités, sans limitation de discipline ou de domaine, et dont la qualité est garantie par l'assignation au sein d'une base centralisée, selon des règles d'attribution

contrôlées : autant d'aspects qui font de l'ISNI un puissant levier d'interopérabilité pour le web de données. La BNF est donc devenue agence ISNI, de façon à pouvoir attribuer des ISNI conformément à ses besoins, et travaille pour le moment à des scénarios d'échanges vertueux de données avec le monde de l'édition dans le cadre du dépôt légal.



Phot. R2hox / CC-BY-SA 2.0 (Flickr)

 **Data path**
« Exposition de données »

WEB DE DONNÉES ET FICHIER NATIONAL D'AUTORITÉS

Sur le front des autorités enfin, comme l'Abes l'a annoncé lors de ses journées 2016, le temps est venu d'explorer l'opportunité et la faisabilité de constituer un fichier national. Le principe en est acquis, les tutelles en conviennent ; il reste cependant beaucoup à préciser : le périmètre d'un tel fichier, sa gouvernance, son modèle, ses formats, son architecture...

La route est longue, mais il est clair qu'elle mérite d'être empruntée : si les stratégies des deux agences bibliographiques passent par le web de données, leurs référentiels seront d'autant plus des référentiels qu'ils auront fusionné. Le web de données aura alors procuré, en plus de l'opportunité de développer les missions de signalement, les conditions d'une substantielle économie de production.

JÉRÔME VILLEMINOZ

Directeur adjoint du Département
des métadonnées à la BNF
jerome.villeminoz@bnf.fr

[7] <http://reliures.bnf.fr>

[8] <http://bp16.bnf.fr/>