

Nouveautés et perspectives des données diffusées par la BnF

5 ans après, contour actuel et usages des données de data.bnf.fr¹

(Raphaëlle Lapôtre et Aude Le Moullec-Rieu)

Le projet actuel : cinq entités...

Data.bnf.fr repose sur cinq entités autour desquelles la BnF organise la présentation de son catalogue et son modèle de données :

- auteur (environ 2 millions d'autorités – personnes physiques et organisations)
- sujet
- œuvre (autorités titres, périodiques, spectacles, auxquelles sont rattachées environ 8 millions de manifestations)
- lieu
- date : cette entité a été créée de toutes pièces pour data.bnf.fr. Elle n'existe pas dans le catalogue général, contrairement aux quatre autres.

Ces cinq entités permettent d'agréger d'autres données de manière dynamique et de procéder à des alignements avec d'autres ressources extérieures (Wikidata, Library of Congress).

Les données présentes dans data.bnf.fr sont celles qui ont été validées par la BnF : n'y figurent donc pas les pré-notices ni les notices non révisées.

... et deux visages

Les fonctionnalités actuelles de l'application se déploient selon deux modalités :

- une interface grand public
- une version RDF qui permet de récupérer les données (en récupérant les données sur la page, en interrogeant le SPARQL endpoint, en utilisant un dump...)

S'appuyant sur les données du catalogue général de la BnF, l'interface de data.bnf.fr – à la fois complémentaire et indépendante des données – hérite de la dimension visuelle très prégnante qui a caractérisé la description bibliographique dès son origine : organisées d'abord sur des fiches cartonnées selon une logique qui tient à la fois du tableau et de la carte, les informations s'inscrivent dans un système qui comprend aussi les modes de classement et les cotes.

Cette réalité se retrouve dans l'interface de data.bnf.fr.

De même, la modélisation FRBR, pensée du point de vue de l'œil de l'utilisateur, comporte aussi un aspect graphique. Le modèle FRBR a vocation à s'intégrer dans l'économie du web. Il repose sur le RDF, lui-même à mi-chemin entre le langage humain et le graphisme : le langage se visualise complètement.

Pour résoudre le paradoxe soulevé par l'économie de l'attention – l'œil humain est limité alors que

¹ <http://data.bnf.fr/>

L'information est surabondante – deux options sont possibles :

- soumettre des listes de résultats à la subjectivité humaine
- organiser la pléthore d'informations en proposant des facettes, en gérant le bruit

Modéliser une représentation du réel ou modéliser le réel ? Le futur de data.bnf.fr

Data.bnf.fr n'est pas une fin en soi : elle est une application expérimentale qui permet de tester le traitement de données et la visualisation. Robot-données, une application issue de data, sera par exemple rendue autonome et servira aux autres applications de la BnF.

Entre représenter pour le web et représenter finement les objets et ressources de la BnF, data.bnf.fr choisit de travailler à décrire plus finement les relations, ce qui permettra de mieux

- lier un roman au film qu'il a inspiré
- lier les différents niveaux dans le cas des séries
- lier les entités entre elles (par exemple visualiser tous les lieux sur lesquels un auteur a écrit, toutes les thématiques qu'il a abordées...)
- proposer de nouveaux types de visualisation pour représenter ces liens
- créer un espace d'échanges et de dialogue
- offrir un outil d'alignement permettant à quiconque de déposer un jeu de données et d'en tirer des alignements

Etudier les usages : restitution d'un atelier tenu avec des réutilisateurs des données de data.bnf.fr

Un premier atelier a été mené avec les fournisseurs de services aux bibliothèques et éditeurs de systèmes d'information. Il était axé sur les données elles-mêmes. Sa réussite encourage la BnF à poursuivre cette étude des différents utilisateurs.

Trois enseignements peuvent être dégagés.

1. Data.bnf.fr est bien perçu comme un lieu d'expérimentation et d'innovation

Les partenaires soulignent le rôle important de l'interface pour le public, même pour celui qui sait manipuler les données, et l'utilité d'un tel outil expérimental lorsqu'il s'agit de discuter avec les tutelles territoriales : ils appellent donc data.bnf.fr à devenir pérenne.

Leur expertise – les SIGB –, et leur connaissance d'autres offres – celle d'OCLC – les conduisent à souhaiter que data.bnf.fr propose aussi des outils de gestion des collections via les données, par exemple pour faciliter le désherbage. Mais les pistes d'amélioration concernent surtout la fraîcheur, la complétude et l'exhaustivité des données, ou encore la granularité des modes de récupération. Enfin, ils encouragent la BnF à renforcer le niveau « expression » (qui est pour l'instant une copie du niveau « manifestation »), clé de voûte de la visualisation pour le public.

2. La grande force de data.bnf.fr est de jouer le rôle d'une bibliothèque de liens

Si les données de data.bnf.fr sont plébiscitées, c'est parce que leur richesse et leur fiabilité permettent un enrichissement des notices. Le premier enrichissement consiste en des liens vers les notices d'autorité. Rameau est également apprécié en tant que source d'enrichissements, même si se pose le problème de la navigation dans ce modèle lorsque certains domaines sont sous-représentés².

Deux pistes de développement sont évoquées : les alignements dans Wikidata et l'expression de la

² Il manque alors des nœuds... Cela se produit dans les bibliothèques offrant une couverture documentaire globale, non spécialisée.

provenance.

3. Un besoin émerge : un espace d'échanges avec la BnF mais aussi entre utilisateurs.

Un tel espace pourrait avoir pour premier objet la mise à disposition d'un outil collaboratif d'alignement manuel des données, qui centraliserait toutes les nouvelles propositions d'alignements, par exemple pour alimenter Rameau avec d'autres référentiels qui le complètent. Cet outil ouvert à tous proposerait des tableaux de concordance (*mapping*³). Il ferait écho au projet de fichier national d'entités mené actuellement avec l'ABES, voire à des entreprises de partage de ce travail d'alignement entre les différentes bibliothèques nationales.

Pour qu'un espace collaboratif fonctionne, il faudrait réunir plusieurs conditions : maintenir toutes les données exposées sous licence ouverte, y compris celles qui seraient reversées par les partenaires, développer des modes de communication en plus du simple courriel (par exemple un réseau social), et multiplier les formes de médiation, notamment pour favoriser la prise en main du SPARQL endpoint.

Toutes ces observations rejoignent les axes de travail retenus par l'équipe de data.bnf.fr elle-même.

Une dernière perspective

Elle consisterait à comparer différents échantillons obtenus dans le cadre de la FRBRisation via data.bnf.fr, les SIGB des bibliothèques municipales..., soit via des données hétérogènes. Il serait intéressant pour data.bnf.fr de tester des visualisations à partir de jeux de données plus proches de ceux d'une bibliothèque publique, et de repérer les formules les plus pertinentes.

³ Un travail d'alignement avec des référentiels autour des enzymes et des poissons est en cours. L'objectif n'est pas nécessairement d'opérer un rapprochement entre des concepts présents dans les deux référentiels, mais plutôt d'affiner RAMEAU à partir de référentiels plus précis que lui (par exemple sur les instruments de musique, etc.).