

Enrichir son portail et gérer les fichiers d'autorité : l'exemple de la BM de Fresnes, juin 2017

Le contexte de la BM de Fresnes

Dès sa première informatisation en 1992, la bibliothèque de Fresnes (Val-de-Marne) a construit son catalogue à partir des notices bibliographiques et d'autorités de la BnF au format INTERMARC, évitant ainsi les pertes d'information éventuelles provoquées par les conversions de format.

Les fichiers d'autorités ont été intégralement récupérés et mis régulièrement à jour. Les notices bibliographiques produites localement ou récupérées ont été par la suite automatiquement écrasées et remplacées par les notices de la BnF au moyen, à l'époque, du « vendangeur » de la société Opsys, en reproduisant les liens entre les différents types de notices indexées. La base disposait ainsi d'un mode de fonctionnement et des identifiants (ARK) nécessaires à l'établissement des liens entre les entités.

En 2012, la bibliothèque de Fresnes répondait conjointement avec la BnF et la société Logilab à l'appel à projets du ministère de la Culture et de la Communication « Services numériques culturels innovants », pour mener l'expérimentation OpenCat, qui fut une première étape vers l'intégration du modèle FRBR par l'intermédiaire de data.bnf.fr au sein d'un catalogue de bibliothèque publique.

Ce fut un SIGB « atypique » en 2013, un choix délibéré en vue dans un premier temps de se « délester » du format MARC, afin d'imaginer aller plus loin par la suite, c'est-à-dire d'exploiter au maximum les relations exposées par data.bnf.fr, mis en ligne depuis juillet 2011.

Le principe architectural s'appuie sur l'intégration de « dumps » issus de data.bnf.fr au sein du système de gestion, qui vont structurer la base de données et permettre ainsi de gérer les liaisons entre les différentes entités constitutives des notices bibliographiques. Les éventuelles données locales intégrées, qu'elles soient produites en interne ou par des fournisseurs tiers, sont alors considérées comme des calques qui se superposent et héritent de cette structure.

Grâce à cette intégration des alignements au cœur du système, il devenait dès lors possible d'envisager une architecture qui se voulait « innovante » en matière de portail documentaire en bibliothèque publique, avec un développement effectué en interne.

Le portail et la recherche documentaire

Un constat de départ : les catalogues des bibliothèques, du fait de la mise en avant quasi-systématique du bloc ISBD, semblent s'adresser avant tout à un public spécialisé. Ils sont généralement peu clairs dans leur navigation et ne mettent pas assez en évidence le ou les documents et leurs disponibilités.

La mise en retrait du bloc ISBD pour la description d'une notice aura donc été à la base d'un développement personnalisé (en interne) pour le [nouveau site web de la BM de Fresnes](#).

Le titre, l'auteur, l'édition, l'indice systématique, le sujet, la quatrième de couverture (note générale) sont des champs qui renvoient vers des entités (classes d'objets) qu'il devient possible de manipuler à l'aide d'une interface utilisateur qui est délibérément (et évidemment) tournée vers

l'utilisateur. La cote et la disponibilité sont indiquées à l'aide de codes couleurs (propres à la bibliothèque pour la localisation, rouge/vert pour la disponibilité).

Plus généralement, il a semblé important d'éviter un (re)chargement complet de la page à chaque changement de « niveau » lors de l'exploration des entités (œuvres, manifestations, autorités), et donc de favoriser un rafraîchissement partiel de la page en exploitant le principe de la fenêtre « pop-up », ce qui permet de conserver un contexte de recherche tout en rendant sensibles les relations existantes à une échelle plus globale.

Détail de la notice ✕



Vingt mille lieues sous les mers [Texte imprimé] / Jules Verne

Romans
Littérature française du 19e s.

Verne, Jules (1828-1905) (Auteur du texte) Neuville, Alphonse de (1836-1885) (Illustrateur) Chelebourg, Christian (1958-....) (Préfacier)

Paris : Librairie générale française, 1988

■ R VERN xxxx En réserve [1 document]

Filmogr. p. 29. Bibliogr. p. 29-30

D'autres éditions de cette oeuvre...



La recherche « expert » est abandonnée, au profit de groupes de facettes exposant différents critères : sujet, responsabilité, édition, indice systématique... Elles permettent le filtrage d'un flux de réponse issu d'une requête globale d'une manière plus directe.

Par une convention de partenariat signée entre l'Institut Suisse Jeunesse et Medias (ISJM) et la Mairie de Fresnes, nous avons de plus intégré les mots-clés thématiques issus de la [plateforme Ricochet](#) (site d'information et de valorisation de la littérature jeunesse francophone), également rendus accessibles sous forme de facettes.

Accueil Informations pratiques Ma bibliothèque Mon compte Jules verne [Rechercher] Tous les critères

Toutes les sections
Section jeunesse
Section adulte

TOUS (250)

Sujets :







- Verne, Jules (1828-1905) (37)
- Critique et interprétation (9)
- Biographies (6)
- Dans la littérature (5)
- Littérature et sciences (4)
- Correspondance (3)
- Dictionnaires (3)
- Voyages extraordinaires (3)
- +* 1800.....- 1899.....+19e siècle: (2)
- Catalogues (2)

Responsabilités :

- Verne, Jules (1828-1905) (194)
- Bennett, Léon (1839-1916) (11)
- Neuville, Alphonse de (1836-

Au moins un exemplaire disponible
Aucun exemplaire disponible


[1 - 12] / 250 Télécharger la page au format pdf

 Voyage au centre de la terre [Texte imprimé] / Jules Verne. Paris : Éditions courtes et longues, 2016	 La maison à vapeur [Texte imprimé] / scénario, Samuel Figuière. [Allauch] : Clair de lune, DL 2013
 Les enfants du capitaine Grant. 3 [Texte imprimé] / Alexis Nesme. [Paris] : Delcourt, impr. 2013	 Vingt mille lieues sous les mers [Texte imprimé] / Jules Verne. [Sayat] : de Borée, DL 2012
 Le tour du monde en 80 jours (D'après l'œuvre de Jules Verne) [Bande originale du feuilleton TV] - Single.	 Les enfants du capitaine Grant. 2 [Texte imprimé] / Alexis Nesme. [Paris] : Delcourt, impr. 2011


Lorsqu'un lien indexé est présent (autorité, sujet...) une notice propre à l'entité est proposée, avec une sélection d'attributs se limitant à une utilisation « tout public ».

Pour une notice bibliographique, il s'agira des attributs : titre, indice systématique, auteur(s), édition, cote, vedettes Rameau, critique éditeur et/ou interne.

Pour une notice d'autorité, il s'agira des attributs : Nom, notes, formes rejetées, « graphe heuristique » Rameau, chronologie des œuvres liées dont la bibliothèque possède des manifestations, les manifestations rattachées à l'œuvre sélectionnée le cas échéant.

 **Jules Verne (1828-1905)**
Romancier et auteur dramatique
Verne, Jules-Gabriel Verne, Jules
-- Les ressources de la Bnf sur data.bnf.fr

Vingt mille lieues sous les mers



Une chronologie d'oeuvres (accessibles à la bibliothèque)

1869
Vingt mille lieues sous les mers

1864 Voyage au centre de la terre
1872 Le tour du monde en quatre-vingt jours

Le tour du monde en quatre-vingt jours

Voyage au centre de la terre L'île mystérieuse

Gestion et usage des autorités

Une autorité sur le portail est considérée comme étant une entité typée unique, qu'elle soit une notice d'autorité issue du catalogue général de la BnF, ou du thésaurus Rameau. Elle reçoit donc un traitement fonctionnel identique (fonction polymorphe), dont l'affichage final se fait le reflet. Un nom de personne aura la plupart du temps des œuvres rattachées, un nom géographique ou matière d'autres entités géographiques ou matière.

Le « cartouche »

Une illustration à gauche, le « titre » (nom de l'auteur, vedette autorité matière ou nom de personne) suivi de ses éventuelles dates, ses formes rejetées, et un lien vers la source de l'entité, qui peut consister en une version numérisée relative à l'illustration sur le [site de Gallica](#).

Le graphe heuristique Rameau

Dans le but de mettre en évidence le contexte dans lequel une autorité-matière sujet se manifeste, il nous est apparu intéressant de générer une visualisation sous forme de graphe heuristique qui rend compte des relations entre les vedettes, tout en caractérisant la nature du lien (termes génériques, associés, spécifiques).

L'exploitation d'un WS permet en outre d'agréger des liens de flux vidéo issus de [Canal-U](#) (la vidéothèque numérique de l'enseignement supérieur et de la recherche). Ces liens viennent enrichir les thématiques Rameau, et sont le résultat d'une requête par mot-clé. Ils sont sciemment limités aux vedettes noms communs et noms géographiques. En effet, les requêtes WS se faisant par l'intermédiaire d'un « littéral » (chaîne de caractère), les noms de personnes sont générateurs de bruit à cause des homonymies...

La frise chronologique

Lors de l'expérimentation OpenCat, une visualisation des œuvres rattachées à un auteur sous forme de frise chronologique avait été mise en place. Nous avons repris et exploité cette idée en limitant l'affichage des œuvres à celles dont la bibliothèque possède au moins une manifestation. Cela permet en effet de situer l'œuvre dans son contexte chronologique, et d'en afficher les manifestations selon leurs différentes expressions à l'aide d'une interface épurée, si elles sont présentes dans le fonds de la bibliothèque.

Viticulture- Rameau - BnF

Viticulture

Vigne Vigne Viticulture Viniculture

Arborescence du sujet

Glisser, déplacer, double-cliquer sur un terme dans l'arbre ci-dessous ou voir les documents de la bibliothèque sur [Viticulture](#)

Termes associés
Termes spécifiques
Termes génériques
Relation non précisée

Aller plus loin avec canalU

Surveillance de l'exposition aux pesticides utilisés en viticulture par l'analyse de cheveux (PHYTOTIF)

L'ergonomie dans les processus de conception, exemple d'interventions dans le secteur de la viticulture

Le vin se met au vert

La biodynamie dans le vignoble : deux exemples concrets à Bordeaux

Des architectures pour le vin / Raphaël Schirmer

Notre modus operandi

La plupart des requêtes concernant le catalogue sont des appels via des web services exposés par notre SIGB, qui envoient en réponse les éléments « FRBRisés » dès que cela est possible. D'autres appels sous forme de requêtes pouvant inclure une négociation de contenu se font vers des services tiers (data.bnf.fr, canal-U).

Lors d'une requête catalogue « récupération des manifestations » :

- Un WS intégré à notre SIGB est appelé, qui nous renvoie une réponse paginée, en fonction d'un ou plusieurs mots clés, constituée d'une liste de manifestations avec les indications de disponibilité. Une liste de « facettes » est alors exposée à gauche, permettant le filtrage selon différents critères (auteur, sujet, éditeur, indice systématique...).

Lors d'une requête catalogue « affichage d'une notice bibliographique » :

- Dès qu'un utilisateur clique sur une manifestation, une requête est envoyée à un WS SIGB qui renvoie le contenu de la manifestation et tous les liens, y compris l'identifiant pérenne type « ark » qui renvoie vers la ressource concernée. D'autre part, l'ensemble des manifestations liées par l'intermédiaire de l'œuvre sont intégrées à la réponse, et il devient alors possible de créer une fonction récursive qui gèrera l'affichage. La visibilité de ces éléments se fait par un

carrousel horizontal en bas de la fenêtre, et propose ainsi à l'internaute de visualiser ces documents.

Lors d'une requête catalogue « autorité » :

- Lorsqu'un utilisateur clique sur un lien d'autorité indexé, deux requêtes sont envoyées (data.bnf.fr par négociation de contenu et WS SIGB). Ces requêtes sont envoyées de manière synchrone (l'une après l'autre), et pour celle impliquant data.bnf.fr, exploite le numéro BnF rattaché afin de récupérer l'ark correspondant et requête à nouveau afin de récupérer les données afférentes (la raison pour laquelle dans ce cas l'ark n'est pas envoyée nativement par le SIGB n'est pas claire). Si la réponse issue de data.bnf.fr existe et s'avère plus récente, elle a la « prévalence » (notre SIGB, du fait d'un « moissonnage » plus ou moins régulier, n'est pas systématiquement à jour).
- L'illustration, la description, les formes rejetées, les liens Rameau (de « premier niveau ») sont analysés et affichés suivant leur modalité propre.

Dans le cas d'un auteur / préfacier... quelle que soit la fonction :

- Les œuvres éventuelles rattachées à l'auteur sont analysées, triées (par date de création) et filtrées pour ne conserver que celles ayant au moins une manifestation présente à la bibliothèque avant d'être affichées sous forme de frise chronologique s'il y a lieu.

Dans le cas d'un sujet Rameau :

- Les vedettes liées et typées selon leur relation sont récupérées, analysées et affichées dans un graphe heuristique qui exploite <https://d3js.org/>, une bibliothèque javascript open-source de visualisation de données.
- Si la vedette correspondante est de type nom commun, une agrégation est effectuée avec des vidéos issues de canal-U. La requête via un WS de canal-U est purement « littérale » (basée sur une chaîne de caractères) et ne fait appel à aucune relation « sémantisée ».
- S'il s'agit d'une vedette « nom de personne », le traitement est identique à celui d'une autorité personne.

La négociation de contenu consiste à faire une requête puis à récupérer l'URI de la ressource dans le type adapté à la demande. Ainsi, dans le cadre d'une demande de type « httprequest » à partir d'une URI de départ qui serait <http://data.bnf.fr/ark:/12148/cb11928016k>, la réponse du serveur est : http://data.bnf.fr/11928016/jules_verne/ qui correspond à la page « ressource » associée.

Les difficultés rencontrées

A partir du moment où la fourniture des données est effectuée à partir de « dumps », il est essentiel de prévoir des mises à jour régulières et relativement fréquentes. Ainsi par exemple, les données relatives à « Cabu » fournies par la base de données de notre SIGB sont antérieures à janvier 2015. Le problème est éventuellement contourné par l'appel dynamique par négociation de contenu à data.bnf.fr.

Mais le serveur hébergeant data.bnf.fr n'étant pas optimisé pour ces requêtes dynamiques, cela crée des latences (timeouts) rédhibitoires la plupart du temps. L'URI de requête pointe pourtant vers un fichier de type « json » qui ne pèse jamais plus de 5 Ko. Ainsi, un timeout de 3 secondes a été mis en place côté code, délai au-delà duquel la requête est abandonnée. Dans l'idéal, il serait souhaitable que la BnF offre des APIs permettant nativement ce genre de requête dynamique.

Il est à noter que les lenteurs dont souffre data.bnf.fr s'étendent au simple affichage d'une page relative à une ressource sur le site lui-même. La fonction qui gère la requête « machine à machine » automatisée lors de l'appel du portail de la bibliothèque reçoit une réponse qui est littéralement plus de 100 fois moins importante en terme de poids. Il s'agit en effet de ne récupérer qu'un simple fichier ne contenant que du texte.

En tant que système d'information, notre SIGB envoie des réponses pré-formatées selon les prérogatives du système et dans l'optique de servir toutes les composantes de gestion de l'établissement. De ce fait, le portail documentaire ne peut que s'appuyer sur ces données qui ont pour inconvénient de ne pas toujours être le reflet parfait des alignements bruts sous-jacents.

En effet, si les masques mis en place pour laisser transparaître les données locales semblent la plupart du temps ne pas poser de problème, ils interfèrent parfois avec les alignements natifs et peuvent à de rares occasions générer des alignements inexistant dans data.bnf.fr. Une accumulation d'approximations (codes fonctions non considérés, ontologies mal interprétées...) est peut-être également en cause.

Quelles réponses ?

Ainsi, dans l'avenir, nous prévoyons d'intégrer les dumps issus de data.bnf.fr dans un triplestore hébergé sur un serveur local à la bibliothèque. Le requêtage croisé (constitution d'un graphe en SPARQL) s'en verra considérablement allégé en termes de ressources réseau et les réponses bien plus complètes et natives, ce qui offre dès lors bien des possibilités d'enrichir le catalogue.

Mais cela implique d'avoir un serveur relativement puissant, surtout en terme de mémoire, et possédant également un processeur avec un jeu d'instruction étendu (x64) pour prétendre adresser cette mémoire au-delà de 4GB. Cela s'en ressent notamment lors des imports, très gourmands en mémoire à cause de l'indexation des triplets et de l'alignement des inférences à la volée. Est également gourmande la présence de jeux de données en mémoire en vue d'optimiser les requêtes...

Par conséquent, les requêtes en elles-mêmes sont optimisées en termes de temps de réponse, et c'est là que cette solution devient intéressante.

Demeure le problème de l'alignement de ces données avec les données locales issues du SIGB. Dans notre cas cela devrait être simplifié par le fait que les identifiants pérennes préexistent au sein de la base de données. Reste à travailler de concert avec les concepteurs du système de gestion afin de parvenir à un résultat probant, et comme pour toute démarche expérimentale, nous avançons en *terra incognita*.

Mettre en place une intégration du modèle FRBR au sein d'un système demande une extrême rigueur. Concernant notre SIGB, et dans l'acceptation générale d'après ce que nous avons pu en constater, la modélisation FRBR a tendance à se cantonner aux entités du groupe 1 (les WEMI = Work, Expression, Manifestation, Item), alors que les ontologies caractérisant les relations entre les entités des différents groupes ne sont pas toujours bien respectées (les codes fonctions non pris en compte par exemple).

Il existe une « granularité » prévue et exposée par le modèle qu'il est important de respecter. Les connaissances des évolutions du modèle par l'intermédiaire de FRBR-LRM sont alors essentielles.

En effet, FRBR ne peut pas être considéré de manière fragmentaire. Il s'agit d'un modèle qui s'insère dans un écosystème bien plus vaste constitué par le web sémantique, avec toutes les intrications que cela implique. Il est impensable par exemple d'imaginer des procédures de mises à jour partielles, la toile tissée par l'ensemble des liens possédant sa propre cohérence.

Pour conclure

Dans une optique où la notice bibliographique se virtualise entièrement et devient par conséquent uniquement constituée de liens issus de relations sémantiques, la visualisation des données afférentes devient un choix qui n'est pas influencé par un format quelconque.

Il s'agit là d'une grande liberté qui est offerte, mais une fois déblayées les limites confortables qu'offre le format MARC, il est essentiel de suivre de très près les évolutions des modèles prévus dans le cadre de la transition bibliographique.

Ceci en effet, afin de veiller à ce que les acteurs du marché des SIGB ne deviennent pas les forgerons d'un ensemble disparate de « manières de faire » découlant des normes en vigueur, ce qui rendrait caduque toute volonté d'implantation rigoureuse des modèles faisant autorité (la bibliographie nationale demeurant le garant en la matière).

Pour le groupe Systèmes & Données : Pierre BOURNERIE, BM de Fresnes

Précédemment publié sous l'URL : <https://www.transition-bibliographique.fr/17-06-20-enrichir-portail-gerer-fichiers-autorite-exemple-bm-fresnes/>