

Notice bibliographique ou manifestation ? Notice d'autorité titre ou œuvre ? Point en juin 2017

La Transition bibliographique, on en parle depuis quelques années, on l'attend et on l'annonce, on se demande ce qu'il faudrait faire pour s'y préparer – et on en oublie parfois qu'en réalité, elle est déjà en cours. Mais comme il n'y aura pas de grand soir catalographique, il faut donc identifier les différents signes, les pas qui nous y font progresser, ces petites évolutions successives dans les données de nos catalogues.

Parmi ces signes, ces pas successifs, pour la BnF une étape particulièrement importante, à la fois au niveau symbolique et à un niveau très concret, a été franchie il y a quelques mois en enrichissant massivement plusieurs milliers de notices bibliographiques du catalogue de la BnF, avec des informations venues de data.bnf.fr.

Techniquement, des liens ont été générés entre des notices bibliographiques et des notices d'autorité Titre. *Dit comme ça, ça ne semble pas forcément très innovant...*

Notice bibliographique

Notice Au format Unimarc ▼

000 cam 22 3 450
001 FRBNF391514040000006
003 <http://catalogue.bnf.fr/ark:/12148/cb39151404j>
010 .. \$a 0-385-50420-9 \$b rel.
100 .. \$a 20040329d2003 m y0frey50 ba
101 0. \$a eng
102 .. \$a US
105 .. \$a |||z 00|||
106 .. \$a r
200 1. \$a The Da Vinci code \$b Texte imprimé \$e a nov
210 .. \$a New York \$c Doubleday \$d 2003
215 .. \$a 454 p. \$d 25 cm
500 11 \$3 14580209 \$a The Da Vinci code \$m anglais
700 .| \$3 14529449 \$a Brown \$b Dan \$f 1964-.... \$4 07
801 .0 \$a FR \$b FR-751131015 \$c 20040329 \$g AFNOR
930 .. \$5 FR-751131010:39151404001001 \$a 2004-133

Notice de titre conventionnel

Notice Au format Unimarc ▼

000 cx h22 45
001 FRBNF145802098
005 20080405
009 <http://catalogue.bnf.fr/ark:/12148/cb14580209w>
100 .. \$a 20041109afrey50 ba0
101 .. \$a eng
102 .. \$a XX
103 .. \$a 2003
106 .. \$a 010
152 .. \$a AFNOR Z 44-061 \$c 2
240 .. \$a Brown \$b Dan \$f 1964-.... \$9 0 \$7 ba0yba0y \$8 freeng \$t The
300 |. \$a Roman policier
440 .. \$a Brown \$b Dan \$f 1964-.... \$7 ba0yba0y \$8 frefre \$t Le Da Vin
440 .. \$a Brown \$b Dan \$f 1964-.... \$7 ba0yba0y \$8 frefre \$t Code Da V
440 .. \$a Brown \$b Dan \$f 1964-.... \$7 ba0yba0y \$8 freita \$t Codice Da
530 .. \$5 xxc \$0 A pour adaptation : \$3 15076975 \$a The Da Vinci code
686 .. \$a 800 \$c Littératures \$2 Note de regroupement par domaine
801 .. \$a FR \$b FR-751131015 \$c 20080405
810 .. \$a Le "Code Da Vinci" décrypté / Simon Cox, 2004

Si on reprend les concepts du modèle FRBR, ce sont plus de **150.000 liens entre manifestations et œuvres qui ont ainsi été créés.**

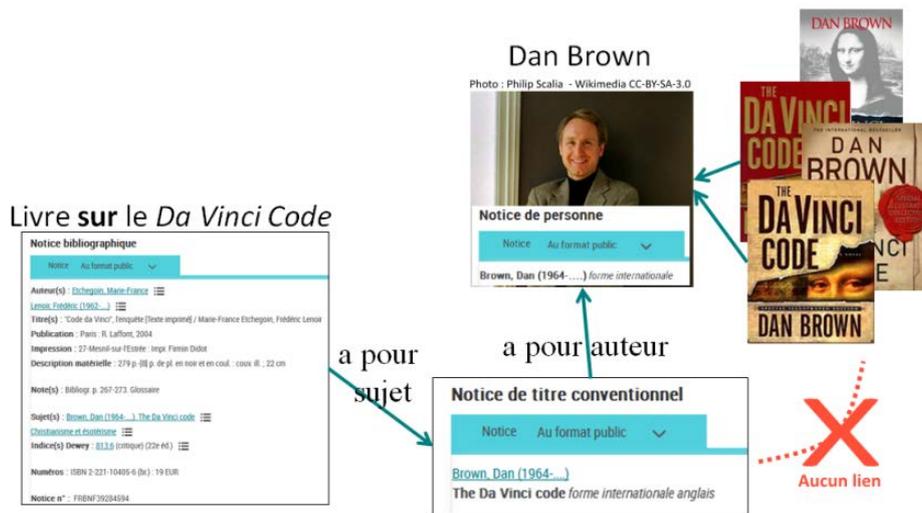
Le contexte

Depuis 2011, la plate-forme data.bnf.fr est un outil d'expérimentation des technologies du web sémantique pour tester la modélisation en FRBR et l'exposition des données en RDF.

Les (méta)données sources sont celles préexistantes dans les bases de production de la BnF, à savoir prioritairement le [catalogue général](#) et [BnF Archives et manuscrits](#).

Dans le catalogue général, on trouve notamment 13 millions de notices bibliographiques (le niveau Manifestation du modèle FRBR) et 250.000 notices d'autorité Titre, créées pour les besoins du catalogue :

- identifier correctement **les œuvres musicales et les classiques de l'Antiquité et du Moyen âge**. Dans ces cas-là, le catalogueur qui décrit l'édition d'une de ces œuvres ajoute manuellement un lien depuis la notice bibliographique vers la notice d'autorité Titre.
- indexer **les documents qui parlent d'une œuvre**. Ces notices d'autorité Titre ont été créées à des fins d'indexation (*une étude sur Les Misérables*, ou *sur le Da Vinci Code*) – mais un grand nombre des œuvres ainsi décrites sont aussi présentes dans les collections de la BnF. Sauf qu'ajouter un lien entre une édition des *Misérables* (ou du *Da Vinci Code*) et la notice d'œuvre n'a jamais fait partie des consignes de catalogage...



Lors du reversement périodique des données dans data.bnf.fr, le lien entre notice bibliographique et notice d'œuvre a été systématiquement calculé : pour chaque ouvrage, un algorithme récupère la forme de titre, et regarde si une notice d'autorité titre (ou notice d'œuvre) de forme de titre comparable est associée au même auteur.

Si c'est le cas, elle crée une relation **rdrelationships:workManifested** de la manifestation vers l'œuvre.

manifestation	relation	oeuvre
http://data.bnf.fr/ark:/12148/cb30037683v	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb12048477w#frbr:Work
http://data.bnf.fr/ark:/12148/cb302836671	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb14438345r#frbr:Work
http://data.bnf.fr/ark:/12148/cb30340029f	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb15952515b#frbr:Work
http://data.bnf.fr/ark:/12148/cb314540424	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb15127684n#frbr:Work
http://data.bnf.fr/ark:/12148/cb34560972c	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb12436638j#frbr:Work
http://data.bnf.fr/ark:/12148/cb34565241c	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb16711740t#frbr:Work
http://data.bnf.fr/ark:/12148/cb34628566m	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb13777805v#frbr:Work
http://data.bnf.fr/ark:/12148/cb34656233z	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb120083602#frbr:Work
http://data.bnf.fr/ark:/12148/cb34731215k	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb120083602#frbr:Work
http://data.bnf.fr/ark:/12148/cb34804647k	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb120082823#frbr:Work
http://data.bnf.fr/ark:/12148/cb34873887k	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb13334743w#frbr:Work
http://data.bnf.fr/ark:/12148/cb34913208q	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb11974350n#frbr:Work
http://data.bnf.fr/ark:/12148/cb34992249n	rdrelationships:workManifested	http://data.bnf.fr/ark:/12148/cb121010016#frbr:Work

Le reversement dans le catalogue

Pendant plusieurs années, ces informations, **ces liens, n'ont existé que dans data.bnf.fr**.

En 2014-2015, la BnF a souhaité en faire bénéficier les données du catalogue, qui sont davantage partagées, diffusées [via les produits bibliographiques](#). Pour l'occasion, l'algorithme a été revu pour être plus restrictif (générer quelques erreurs à la volée dans data.bnf.fr est moins gênant que d'inscrire ces erreurs en dur dans les notices du catalogue).

En novembre 2015, 167.000 liens manifestation-œuvre ont été versés dans les notices bibliographiques du catalogue de la BnF, après un an de spécifications, contrôles et corrections. Désormais, cet algorithme est rejoué tous les ans : en février 2017, 7.000 nouveaux liens ont été ajoutés de la même manière.

Précision : L'algorithme a pu identifier un certain nombre de liens à générer, mais le processus de chargement dans le catalogue BnF a ajouté des restrictions. Ainsi, toute notice identifiable comme agrégat (agrégat = danger !) a été retirée du traitement. Du moins pour le moment...

Pour la suite...

Le même algorithme a permis d'identifier des propositions de liens, jugées dans un premier temps trop incertaines, qui n'ont pas été reversées mais pourront l'être en partie après des vérifications supplémentaires.

De même le traitement suggère l'ajout de liens entre les notices de spectacles et les notices d'œuvres ([ces liens existent déjà dans data.bnf.fr](#) mais pas encore dans le catalogue).

Du point de vue des utilisateurs des produits bibliographiques

Comment faire pour récupérer les notices enrichies par ces traitements ? Elles sont identifiables (et donc requêttables !) par un code propre à chacun des deux chargements :

- traitement de fin 2015 : utiliser la [clef de recherche 20160116PRR1V04](#)
- traitement de début 2017 : utiliser la [clef de recherche 20170111PRR1V04](#)

Ces termes peuvent servir de critères de recherche [via le serveur Z39.50 de la BnF](#), et donc les notices récupérées par ce biais. Cela dit, le code lui-même n'est pas visible dans l'affichage en Unimarc (ou les 2 codes sont convertis en libellé "\$2 lien automatique" dans la zone 500 générée) :

```
500 11 $3 11939715 $a The Lord of the rings $m français $2 lien automatique
```

mais la recherche par les codes ci-dessus fonctionne aussi quand on souhaite récupérer les notices en format UNIMARC, quand bien même le code n'apparaîtrait pas dans la notice récupérée.

Par ailleurs, un grand nombre de ces liens concerne des œuvres musicales qui font l'objet de dépouillements spécifiques par le département de l'Audiovisuel de la BnF, sous forme de "notices analytiques" ([exemple](#)).

Or ces sous-notices s'accommodent mal du protocole Z39.50. Pour le moment, elles sont récupérables au format UNIMARC "à plat" (les informations sont dans des zones de la notice principale, donc une [notice contenant plusieurs œuvres musicales aura plusieurs zones 500](#)) :

```

Records: 1
[!OUT-ANA1-UTF8]Record type: Unimarc
01433ncm 22003733n 450
001 FRBNF433494100000001
003 http://catalogue.bnf.fr/ark:/12148/cb43349410t
039 $o MUS $a FG059801340001PA
100 $a 20110907d1916 u y0frey50 ba
101 0 $a fre
102 $a FR
106 $a r
128 $a uu
200 1 $a Choix de grands succès soigneusement transcrits pour flûte seule...
$b Musique imprimée
210 $a Paris $c Hengel $d [1916-]
215 $a In-8o
801 0 $a FR $b FR-751131015 $c 20110907 $g AFNOR $h FRBNF433494100000001 $2 int
ermrc
930 $5 FR-751021003:43349410001001 $a 4-UM10-13 $b 759999999 $c Richelieu - M
usique - Magasin $d N
300 $9 a001000 $a Mléditation
464 1 $9 a001000 $t Thalys
500 10 $9 a001000 $3 13916056 $a Thalys $2 lien automatique
701 1 $9 a001000 $3 13027102 $a Massenet $b Jules $f 1842-1912 $4 230
300 $9 a002000 $a Entr'acte
464 1 $9 a002000 $t Mignon
500 10 $9 a002000 $3 13920376 $a Mignon $2 lien automatique
701 1 $9 a002000 $3 13700400 $a Thomas $b Ambroise $f 1811-1896 $4 230
300 $9 a003000 $a Pizzicati
464 1 $9 a003000 $t Sylvia
500 10 $9 a003000 $3 13927265 $a Sylvia $n Musique de ballet $2 lien automatique
701 1 $9 a003000 $3 12519700 $a Delibes $b Léo $f 1836-1891 $4 230
300 $9 a004000 $a Aragonaise
464 1 $9 a004000 $t Le Cid
500 10 $9 a004000 $3 13925628 $a Le Cid $2 lien automatique
701 1 $9 a004000 $3 13877189 $a Massenet $b Jules $f 1842-1912 $4 230

```

... et au-delà ?

La BnF a pour projet d'augmenter le nombre de ces liens manifestations-œuvres (il sera question d'expressions une autre fois). Pour cela, il faut :

- **trouver de nouveaux liens.** Un nouvel outil, baptisé "RobotDonnées", permet de travailler sur des sous-ensembles du catalogue, pour ajuster les paramètres de l'algorithme d'alignement en fonction des corpus traités – afin d'obtenir un plus grand nombre de propositions de liens sans pour autant générer de faux positifs.
- **générer de nouvelles notices d'œuvres.** Car la grande majorité des ouvrages conservés n'ont évidemment pas fait l'objet d'études, donc n'ont pas de notice d'œuvre auxquelles les associer. Le même outil doit permettre de générer de nouvelles notices d'œuvres, par calcul, à partir des métadonnées présentes dans les notices bibliographiques.

Et c'est ainsi qu'en associant le travail de la machine et l'expertise du catalogueur – expertise portant à la fois sur les données sources et sur le modèle cible – que nous avançons peu à peu dans ce processus de transformation des données.

Pour le groupe Systèmes & Données : Étienne CAVALIÉ, BnF

Précédemment publié sous l'ancienne URL : <https://www.transition-bibliographique.fr/17-06-20-notice-bibliographique-manifestation-notice-autorite-titre-oeuvre/>