

S'appuyer sur la structure des données et les liens : le format MARC comme tremplin pour le Web de données : l'exemple de data.bnf.fr

Vincent Boulet¹

Département Information bibliographique et numérique, Service Coordination et administration des métadonnées, Pôle données d'autorité - Bibliothèque nationale de France
E-mail vincent.boulet@bnf.fr



Copyright © 2013 by **Vincent Boulet**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Résumé :

Les formats UNIMARC et INTERMARC (format de travail utilisé à la BnF) permettent de structurer de manière fine les données. Des liens divers et nombreux peuvent être établis entre différentes notices bibliographiques, différentes notices d'autorité et entre notices d'autorité et notices bibliographiques. Ce sont des atouts majeurs pour propulser le contenu des notices dans le Web de données. La BnF a pour cela lancé un projet, data.bnf.fr, dont le périmètre et les fonctionnalités s'enrichissent progressivement. Il tire parti de la mise en réseau des données ainsi que la qualification des relations permises par le format INTERMARC pour exploiter les données du catalogue. Data.bnf.fr utilise pour cela les bonnes pratiques du Web de données, priorise les ontologies communes. Cela permet aux données d'être facilement réutilisables d'une manière très large, pour les usages extrêmement variés. De plus, data.bnf.fr, par des opérations de croisement et d'alignement de données, crée du contenu et enrichit les données existantes. Le format MARC est donc utilisé comme tremplin pour le Web de données.

Mots clés: INTERMARC, UNIMARC, Resource Description Framework, Web de données

¹ Vincent Boulet est conservateur à la Bibliothèque nationale de France (depuis 2007), au département Information Bibliographique et Numérique. Il est expert en autorité et membre du comité permanent de la section de bibliographie de l'IFLA. Il a présidé le Conseil VIAF (2012-2013).

Introduction

Les pratiques professionnelles du monde des bibliothèques peuvent s'enorgueillir d'une longue tradition de données structurées. Il s'agit de leur meilleur atout pour répondre aux défis posés par le Web et la question des *Linked data*. Il est possible d'appliquer non seulement le modèle conceptuel FRBR mais également les technologies du Web sémantique en s'appuyant sur les atouts des formats MARC. Ces derniers permettent de structurer finement les données. Ils permettent en outre l'utilisation de différents liens, entre notices bibliographiques et notices d'autorité, mais également entre notices bibliographiques elles-mêmes et notices d'autorité elles-mêmes. Cela rend possible d'extraire l'information et d'organiser les données selon le modèle FRBR et selon la syntaxe des triplets RDF par opération automatique. En fait, il n'est pas nécessaire de réécrire les catalogues.

I- Les données existantes : le cas d'INTERMARC et de ses atouts

1) Vous avez dit INTERMARC ?

Au commencement était le MARC et plus précisément INTERMARC, qui est le format de travail utilisé à la BnF pour créer, stocker et gérer les données bibliographiques et d'autorité. La première version fut élaborée en 1974² par un groupe de travail européen qui avait pour objectif d'établir un format d'échange commun pour une bibliographie francophone. Ses principes sont proches du MARC anglo-américain d'alors³. La publication de la première version de ce dernier empêche INTERMARC de devenir le format de travail en France. Il reste cependant utilisé à la Bibliothèque Nationale et y est enrichi pour les besoins de descriptions très variées, notamment pour les documents spécialisés : musique imprimée, cartes et plans, documents audiovisuels... Quand la Bibliothèque Nationale devient Bibliothèque nationale de France, un nouveau système informatique est mis en place, centré sur l'idée d'un catalogue unique. Dans ce contexte, le format INTERMARC est modifié en profondeur. Les formats INTERMARC spécifiques pour les documents spécialisés sont fondus en un seul format, qui est appelé « INTERMARC intégré », en 1998.

Le format INTERMARC intégré se compose de deux volets : INTERMARC (A) pour les notices d'autorité et INTERMARC (B) pour les notices bibliographiques.

INTERMARC (A) est un format neuf. Il s'applique pour l'ensemble des données d'autorité (noms de personne, collectivités, titres textuels et musicaux, noms géographiques, classification Dewey, marques audiovisuelles, RAMEAU). Une seule autorité est établie par entité décrite, qu'elle soit en accès signalétique ou analytique. Les données d'autorité sont donc centrées sur l'entité elle-même, ou pour les noms de personne sur son identité publique, et non sur l'utilisation de la notice dans le contexte restreint du catalogue à un moment donné.

² Les versions actuelles d'INTERMARC sont : la version 9.1 pour le format de données bibliographiques et la version 4.1 pour les données d'autorité (depuis 2009). En France, outre la BnF, le format INTERMARC est utilisé par la Bibliothèque municipale de Fresnes, dans la banlieue de Paris, et par les bibliothèques de conservatoire.

³ MARC I et MARC II élaborés par la Bibliothèque du congrès en 1966 et 1968, BNB Marc par la British Library en 1969. Le format MONOCLE qui fut la première tentative pour adopter les formats MARC est utilisé comme source.

Le format INTERMARC (B) est le format de référence pour l'ensemble des données bibliographiques, quel que soit le support du document, pour les ressources qui entrent dans un processus éditorial⁴. Cela implique qu'INTERMARC n'est plus centré sur les documents imprimés.

2) La fine structuration d'UNIMARC et d'INTERMARC

INTERMARC est plus proche de MARC 21 que d'UNIMARC si l'on considère l'organisation des zones en blocs et le choix des étiquettes. Il est cependant plus proche d'UNIMARC si l'on prend en compte la structuration des données à l'intérieur même des zones. INTERMARC permet une granularité de l'information plus fine que MARC 21, ce qui est un atout important pour la problématique qui nous occupe ici. Il est de ce point de vue similaire à UNIMARC.

Pour les notices bibliographiques, on peut prendre l'exemple de la structuration de la zone de titres et mentions de responsabilité (zone 200 en UNIMARC, 245 en MARC 21 et INTERMARC), qui est fine en UNIMARC et en INTERMARC, mais faible en MARC 21.

INTERMARC : 245 1 \$a Passionate observer **\$d** Texte imprimé **\$e** Eudora Welty among artists of the thirties **\$e** [exhibition, Mississippi museum of art, Jackson, April 6 - June 30, 2002] **\$f** edited by René Paul Barilleaux **\$g** essays by Suzanne Marrs, Patti Carr Black, Francis V. O'Connor

UNIMARC : 200 1 \$a Passionate observer **\$b** Texte imprimé **\$e** Eudora Welty among artists of the thirties **\$e** [exhibition, Mississippi museum of art, Jackson, April 6 - June 30, 2002] **\$f** edited by René Paul Barilleaux **\$g** essays by Suzanne Marrs, Patti Carr Black, Francis V. O'Connor

MARC 21 : 245 10 |a Passionate observer : |**b** Eudora Welty among artists of the thirties / |**c** René Paul Barilleaux, editor ; essays by Suzanne Marrs, Patti Carr Black, Francis V. O'Connor.

UNIMARC et INTERMARC peuvent, dans la sous-zone \$g distinguer les mentions de responsabilité suivante de la première mention de responsabilité indiquée en \$f. Ceci est impossible en MARC 21. UNIMARC et INTERMARC respectent donc l'entièreté de la structuration des données édictée en ISBD, qui distingue la mention de responsabilité principale des mentions de responsabilité suivantes.

L'exemple de la structuration des formes du nom dans les notices d'autorité est révélateur. On aura ainsi :

INTERMARC : 100 \$w.O..b....\$a Baudelaire **\$m** Charles **\$d** 1821-1867

UNIMARC : 200 |\$7ba0yba0y\$8fre \$9 0 \$a Baudelaire **\$b** Charles **\$f** 1821-1867

MARC 21 : 100 1_ |a Baudelaire, Charles, |**d** 1821-1867

UNIMARC et INTERMARC permettent de distinguer pour un nom de personne l'élément d'entrée (ici : le nom de famille, Baudelaire) de l'élément rejeté (ici : le prénom, Charles) de la forme retenue de la notice d'autorité. Ces éléments sont entrés dans des sous-zones particulières. En revanche, MARC 21 ne permet pas une telle distinction et entre les différents éléments du nom dans une même sous-zone.

⁴ A l'exception donc des documents d'archives et des manuscrits, qui sont décrits en EAD.

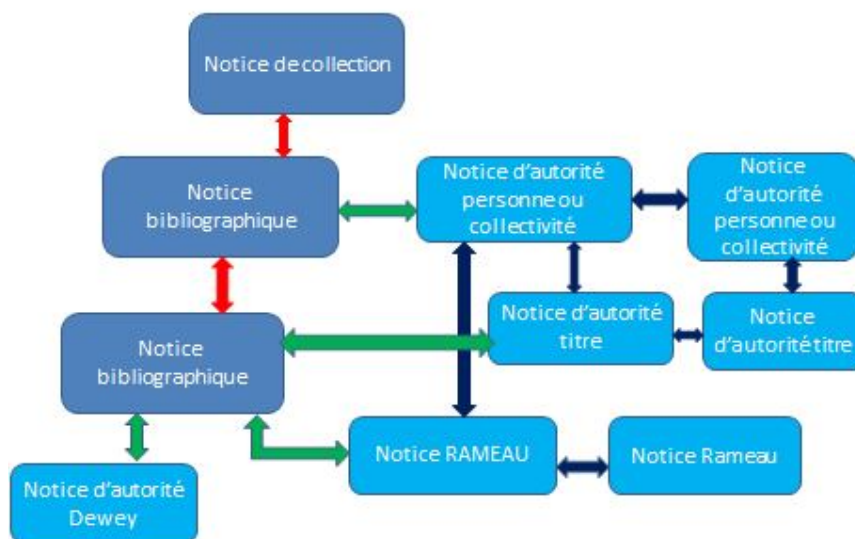
3) Données textuelles et référentiels codés

Les données en MARC ne sont pas seulement des données textuelles, mais également des données codées qui renvoient à des référentiels internationaux. C'est le cas par exemple du référentiel pour les langues, qui s'appuie sur la norme internationale ISO 639-2b « Codes pour la représentation des noms de langue - Partie 2 : Code alpha-3 ». La classification Dewey présente un cas de figure similaire. La BnF gère un fichier d'autorité Dewey, qui renvoie à la version de référence de la classification au niveau international⁵.

Ces référentiels sont importants car ils constituent une des portes d'entrées du MARC dans le Web de données. Ils ont été intégrés dans le Web de données avant les données bibliographiques. La classification Dewey a été utilisée par l'espace de noms dewey.info⁶ et la norme ISO des codes de langues par l'espace de noms loc.gov⁷.

4) Mise en réseau des données

Les possibilités de liens dans le format INTERMARC sont riches et variés. Ils sont exposés dans les tableaux donnés en annexe et résumés dans le schéma suivant :

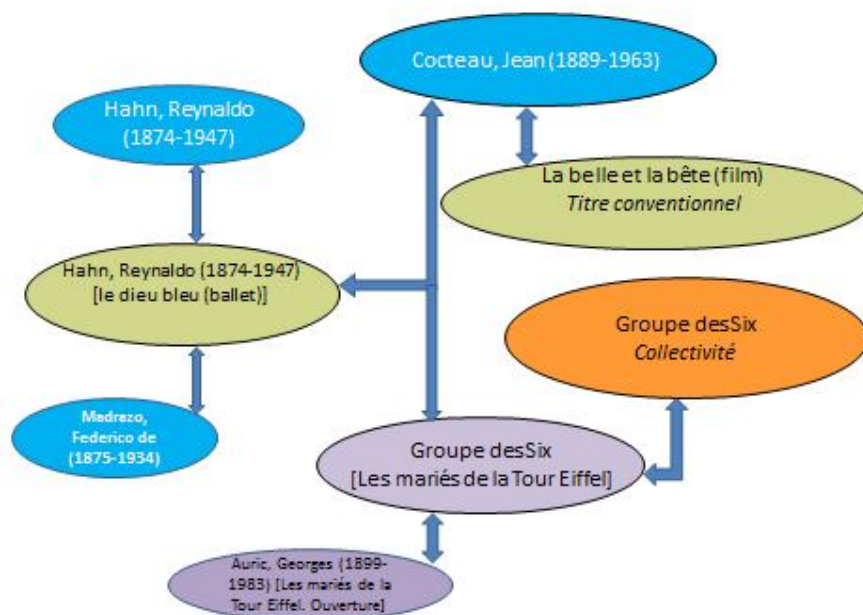


Ces différents liens permettent de mettre les données en réseau et de reporter l'information d'une notice à l'autre. Cela doit être particulièrement souligné pour les notices d'autorité, qui peuvent être liées les unes aux autres, indépendamment des notices bibliographiques. Les données d'autorité prennent une certaine autonomie envers les notices bibliographiques et peuvent être utilisées pour elles-mêmes.

⁵ <http://catalogue.bnf.fr/ark:/12148/cb13645241s/PUBLIC> (consulté le 2013-07-25).

⁶ <http://dewey.info/> (consulté le 2013-07-25).

⁷ <http://www.loc.gov/marc/languages/> (consulté le 2013-07-25).



Les liens sont réciproques : quand un lien est créé d'une notice vers une autre, le lien miroir de cette autre notice vers la première s'établit automatiquement.

5) Qualifier les relations

Les liens en soi ne sont pas suffisants. Il est nécessaire de les qualifier afin de leur donner du sens.

- a) **Par l'étiquette du champ MARC.** Par exemple, pour les liens entre notice bibliographique et notice d'autorité. La zone 100 d'une notice d'autorité indique la forme retenue d'un nom de personne. Le contenu et la structuration de la zone est transférée automatiquement de la notice d'autorité à la notice bibliographique en récupérant l'identifiant de la notice d'autorité.

notice bibliographique

100	\$3	12434010	\$w 0 b..... \$a Auric \$m Georges \$d 1899-1983 \$4 0110
245	1	\$a	Lettre de Georges Auric à Francis Poulenc, Paris, 8 avril 1961 \$d Texte manuscrit

Le même principe s'applique pour les notices d'autorité titres uniformes textuels ou musicaux auteurs/titres. Le nom de l'auteur n'est pas saisi mais est récupérée par le lien avec la notice d'autorité personne.

notice d'autorité titre uniforme textuel

100	\$3	12406064	\$w 0 b.....\$a Hahn \$m Reynaldo \$d 1874-1947
144	1	\$w....b.fre.	\$aLe dieu bleu \$e ballet

- b) **Par des codes de fonction qui précisent cette relation.** Ces codes de fonction renvoient à un référentiel, utilisé aussi bien pour les liens d'une notice d'autorité à une notice bibliographique en MARC qu'à un instrument de recherche en EAD. Dans l'exemple ci-dessus,

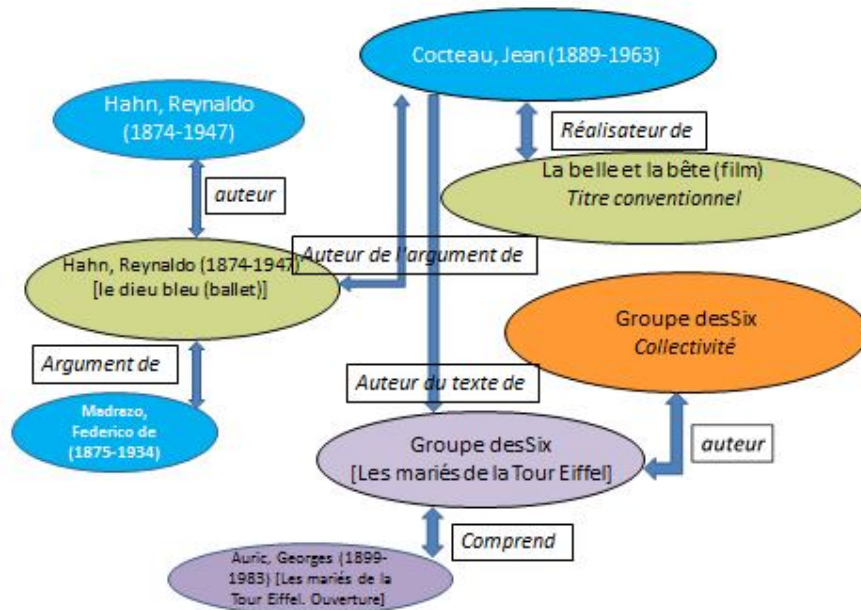
la nature du lien entre la notice bibliographique « Lettre de Georges Auric à Francis Poulenc, Paris, 8 avril 1961 » et la notice d'autorité « Auric, Georges (1899-1983) » est définie et précisée par le code de fonction « 0110 » indiqué dans la sous-zone \$4 de la zone de lien 100. Ce code signifie « auteur de lettres »⁸.

- c) **Par un indicateur qui vient au besoin préciser pour un champ MARC donné la nature de la relation.** Par exemple, une relation entre une personne et une collectivité est précisée dans le champ INTERMARC 511. Il s'agit d'une relation technique, qui déclare un lien entre une notice d'autorité personne et une notice d'autorité collectivité. Les indicateurs portent sur la nature intellectuelle de la relation entre les entités décrites.

```
100 $w 0 b $a Poulenc $m Francis $d 1899-1963
511 1 $3 13195155 $w 20..b.....$a Groupe des six
```

L'indicateur « 1 » du champ 511 précise que Francis Poulenc « est membre » du Groupe des six.

Grâce à la manière dont sont qualifiées les relations, on obtient un réseau de données d'autorité et bibliographiques non seulement liées entre elles mais également mise en relations intellectuellement entre elles. On peut représenter ces relations de la manière suivante :



Les données en MARC, par leur structure, les liens et la sémantisation des liens, permettent d'utiliser le modèle de données existant pour le réinvestir dans le web de données. La BnF a pour ce faire lancé un projet, data.bnf.fr. en 2010.

⁸ Voir la liste de l'ensemble des codes utilisés : http://www.bnf.fr/documents/intermarc_ref_fonctions.pdf (consulté le 2013-07-24)

II- Data.bnf.fr : méthodes et objectifs⁹

Data.bnf.fr répond à plusieurs objectifs, qui peuvent être résumés par l'idée de sortir les données du catalogue pour les placer là où passent les utilisateurs du Web, en utilisant les standards communs du Web et les technologies du Web sémantique. Il s'agit également de mettre en relation les données des catalogues avec des ressources extérieures. L'objectif est de créer des correspondances entre référentiels équivalents ou complémentaires par le biais d'opérations d'alignements automatiques.

1) Objectifs de data.bnf.fr : utiliser les bonnes pratiques du Web pour sortir les données des catalogues

Data.bnf.fr rend les données des catalogues accessibles directement par les moteurs de recherche, selon les principes du Web de documents. Les données sont également accessibles en utilisant les technologies du Web sémantique et les bonnes pratiques du Web de données. Data.bnf.fr fournit donc des pages lisibles et exploitables par les humains (pages HTML) et des données brutes exploitables par les machines, en proposant des sorties de données en RDF et JSON.

L'application des principes du Web de données à data.bnf.fr se fait grâce à l'utilisation des URI http, par exemple les identifiants pérennes ARK, et à la mise à disposition des données RDF par négociation de contenu à partir des pages de data.bnf.fr. Afin de faciliter l'indexation des pages par les moteurs de recherche, les URL comprennent des labels explicites. L'URL des pages œuvre, auteur, thème est formée suivant le principe : <http://data.bnf.fr/ID/label>. Le label est issu des zones MARC correspondant à la forme retenue de la notice d'autorité, en privilégiant toutefois l'ordre usuel :

Exemple :

Personne : [http://data.bnf.fr/ID/prénom\(100\\$m\)_nom\(100\\$a\)](http://data.bnf.fr/ID/prénom(100$m)_nom(100$a))¹⁰

Organisation : [http://data.bnf.fr/ID/nom\(110\\$a\)](http://data.bnf.fr/ID/nom(110$a))¹¹

Œuvre : [http://data.bnf.fr/ID/titre\(245\\$a_ou_14X\\$a\)_complément_de_titre\(\\$i\)](http://data.bnf.fr/ID/titre(245a_ou_14Xa)_complément_de_titre($i))¹²

Pour Victor Hugo :

http://data.bnf.fr/11907966/victor_hugo/

De plus, l'amélioration des données de data.bnf.fr par des liens vers des jeux de données extérieurs fait partie intégrante de l'application des bonnes pratiques du Web de données. C'est ainsi que les données de data.bnf.fr sont alignées avec les référentiels suivants :

⁹ Voir pour de détails : les pages « A propos » sur le site data.bnf.fr. On peut également se reporter à Vincent Boulet, L'infrastructure de l'information et les besoins des utilisateurs : tout le pouvoir aux données structurées!, 78^e congrès de l'IFLA, Helsinki, 2012, <http://conference.ifla.org/past/ifla78/80-boulet-fr.pdf>

¹⁰ Zone 100 : forme retenue d'une notice d'autorité nom de personne, structurée en sous-zones \$a élément d'entrée \$m élément rejeté. Exemple : 100 \$a Hugo \$m Victor.

¹¹ Zone 110 : forme retenue d'une notice d'autorité nom de collectivité

¹² Zone 245 \$a : zone de titre des notices bibliographiques. Zone 14X \$a : forme retenue d'une notice d'autorité titre : 141 pour un titre uniforme textuel, 144 pour un titre uniforme musical, 145 pour un titre conventionnel

Référentiel BnF	Référentiel aligné
Notices RAMEAU	STITCH RAMEAU ¹³
Notices RAMEAU	Thésaurus W ¹⁴
Notices RAMEAU	LCSH
Notices d'autorité Dewey	Dewey.info
Langues	id.loc.gov languages
Types de manifestation	Dcmi : type
Notices d'autorité Personne	VIAF
Nationalités	id.loc.gov countries
Noms géographiques	Geonames
Notices d'autorité Personne	IdRef ¹⁵

L'objectif est de respecter les quatre grands principes du *linked data* :

- a) nommer les ressources avec des URI ;
- b) utiliser des URI http (ou « URI déréférencables ») de façon à ce qu'on puisse utiliser ces URI pour accéder à des informations sur les ressources ;
- c) renvoyer des informations utiles grâce à RDF lorsqu'on déréférence une URI ;
- d) se relier avec d'autres URI pour créer un réseau de liens.

2) L'exploitation des données des catalogues

Ces traitements de données sont rendus possibles par les données structurées et par les liens permis par le format MARC, entre notices d'autorité et notices bibliographiques. Le modèle de données de data.bnf.fr est centré sur les notions d'œuvres, d'auteurs et de thèmes. Le contenu des pages œuvres provient des notices d'autorité Titres (titres uniformes textuels¹⁶, titres uniformes musicaux¹⁷, titres conventionnels). Celui des pages auteurs des notices d'autorité personne¹⁸ et collectivité¹⁹. Les codes de fonction qui qualifient le lien entre une notice d'autorité et une notice bibliographique sont également exploités, ce qui permet un tri des résultats²⁰.

À partir des liens avec les notices d'autorité, les principales bases documentaires de la BnF sont exploitées par data.bnf.fr : les catalogues BnF catalogue général (en MARC) et BnF archives et

¹³ Projet expérimental mené par TELplus pour exprimer Rameau en SKOS (<http://www.cs.vu.nl/STITCH/rameau/>).

¹⁴ Thésaurus des Archives de France pour la description des archives contemporaines dans les Archives Départementales.

¹⁵ IdRef est une application Web des données d'autorité des catalogues des bibliothèques des institutions d'enseignement supérieur en France. Elle est développée et maintenue par l'ABES. Elle permet à des utilisateurs d'enrichir et d'utiliser ces données, notamment par l'utilisation des technologies du Web sémantique.

¹⁶ Par exemple : http://data.bnf.fr/13189630/jean_cocteau_la_machine_infernale/. Les données proviennent de la notice d'autorité : <http://catalogue.bnf.fr/ark:/12148/cb131896305/>

¹⁷ Par exemple : http://data.bnf.fr/13963731/groupe_des_six_les_maries_de_la_tour_eiffel/. Les données proviennent de la notice d'autorité : <http://catalogue.bnf.fr/ark:/12148/cb139637319/>

¹⁸ Par exemple : http://data.bnf.fr/11897143/jean_cocteau/. Les données proviennent de la notice d'autorité <http://catalogue.bnf.fr/ark:/12148/cb11897143s/PUBLIC>

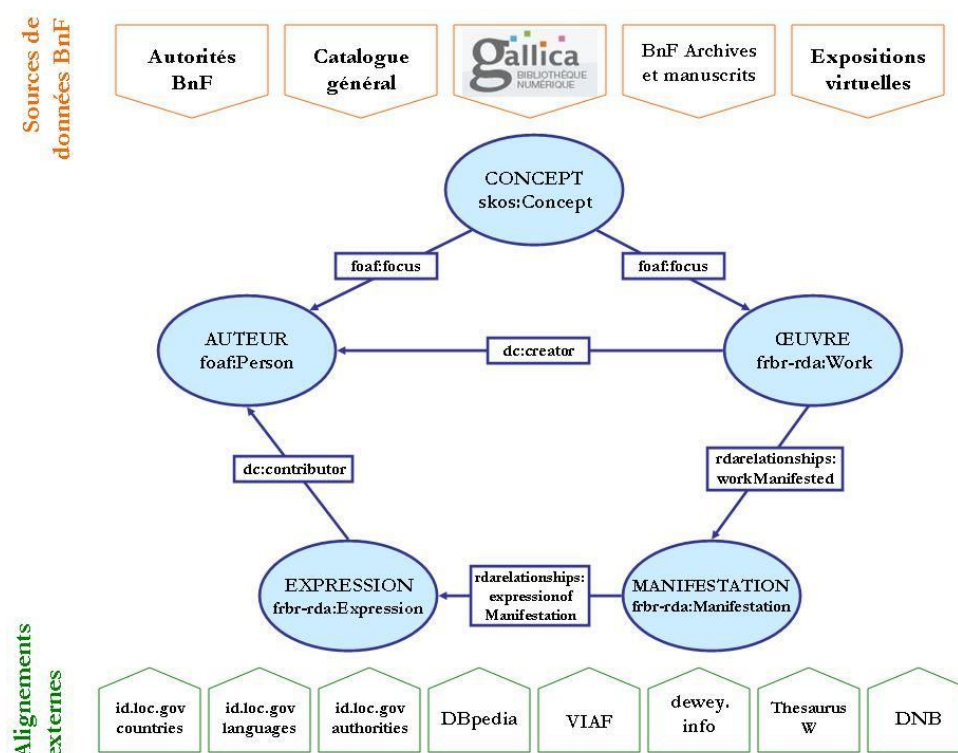
¹⁹ Par exemple : http://data.bnf.fr/13195155/groupe_des_six/. Les données proviennent de la notice d'autorité : <http://catalogue.bnf.fr/ark:/12148/cb13195155c/>

²⁰ Voir par exemple la fiche de Jean Cocteau : http://data.bnf.fr/11897143/jean_cocteau/ dans l'onglet « ses activités ».

manuscrits (en EAD), la bibliothèque numérique Gallica (grâce aux liens entre Gallica et les deux catalogues) et les expositions virtuelles²¹.

Data.bnf.fr ne couvre pas l'ensemble des données des catalogues, et ne les remplace pas. Le périmètre doit être maîtrisé afin de pouvoir garantir la qualité des données fournies, et le degré de confiance qu'elles peuvent inspirer du fait du label « BnF ». L'extension du périmètre de data.bnf.fr est progressif. La première mise en ligne, en juillet 2011, couvrait les auteurs et les œuvres citées dans les index de la grande anthologie classique de la littérature française, Le « Lagarde et Michard »²². L'extension aux thèmes, à partir de notices RAMEAU est venue ultérieurement. Data.bnf.fr couvre aujourd'hui (juillet 2013) environ 20% des catalogues : 200 000 auteurs et 20 000 œuvres.

On peut représenter le modèle de données de manière synthétique ainsi :



Ainsi, à travers les notices d'autorité, les données en MARC constituent une des bases fondamentales de data.bnf.fr.

²¹ <http://expositions.bnf.fr/>

²² Les grands auteurs français du programme: anthologie et histoire littéraire / André Lagarde, Laurent Michard. Paris : Bordas.

III- Du Marc au RDF

1) Prioriser les ontologies communes

La priorité est donnée aux standards communs du Web, largement utilisés hors du monde des bibliothèques. A défaut, sont utilisées des ontologies communes du monde de la documentation, au sens large comme Dublin Core Metadata terms, ou davantage centrées sur le monde de bibliothèques comme RDA. L'utilisation de l'ontologie RDA, et non des ontologies de l'IFLA (FRBR, FRAD ou ISBD) s'explique par la date du projet, qui a été conçu alors que les référentiels de l'IFLA n'étaient pas encore publiés en RDF. Les alignements entre les espaces de nom de l'IFLA et l'ontologie RDA sont nécessaires pour accroître l'interopérabilité entre eux et les potentialités de réutilisation de données ; Il est en effet difficile de changer le modèle de données existant au profit d'autres ontologies, ce qui risquerait de perturber les réutilisateurs actuels de l'application data.bnf.fr. En outre, quelques propriétés propres à la BnF sont employées à titre exceptionnel. Pour que les données soient le plus largement accessibles et réutilisables, il y a besoin d'ontologies communes²³.

2) L'exploitation de la structuration des données des notices d'autorité

La structuration des données en MARC est exploitée. Le choix a été fait, cependant, de ne pas rendre l'ensemble des données d'une notice bibliographique ou d'une notice d'autorité en MARC dans le RDF fourni, ni l'ensemble de sa structuration. Data.bnf.fr cherche en effet à se positionner comme un service offrant des possibilités de large récupération des données, par l'utilisation de la licence libre gouvernementale française, Etalab. Cette licence permet une libre récupération des données avec mention d'attribution de source. Elle est donc comparable à une licence CC-By ou ODC-BY.

On aura par exemple dans la page de data.bnf.fr sur le Groupe des six²⁴ les données suivantes venant de la notice d'autorité collectivité correspondante²⁵.

²³ Les vocabulaires utilisés sont les suivants :

rdf : <http://www.w3.org/1999/02/22-rdf-syntax-ns/>;
rdfs : <http://www.w3.org/2000/01/rdf-schema/>;
skos : <http://www.w3.org/2004/02/skos/core/>;
dc : <http://purl.org/dc/terms/>;
foaf : <http://xmlns.com/foaf/0.1/>;
RDAGroup2elements : <http://RDVocab.info/ElementsGr2/> ;
RDARelationships : <http://rdvocab.info/RDARelationshipsWEMI/>
RDAGroup1elements : <http://RDVocab.info/Elements/>
owl : <http://www.w3.org/2002/07/owl>
ore : <http://www.openarchives.org/ore/1.0/rdfxml>

²⁴ <http://data.bnf.fr/11985800/phedre/>

²⁵ <http://catalogue.bnf.fr/ark:/12148/cb11985800x/>

MARC	RDF
<p>Zone 008 position 12-13 Nationalité →</p>	<pre><rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb13195155c#foaf:Organization"> <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/> <foaf:page rdf:resource="http://data.bnf.fr/13195155/groupe_des_six/"> <rdagroup2elements:placeAssociatedWithTheCorporateBody rdf:resource="http://id.loc.gov/vocabulary/countries/fr"/></pre>
<p>Zone 624 Note de regroupement par domaines →</p>	<pre><rdagroup2elements:fieldOfActivityOfTheCorporateBody</pre>
<p>Zone 008 position 14-16 Langue →</p>	<pre>rdf:resource="http://dewey.info/class/780/"> <dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-2/fr"/> <dc:subject rdf:resource="http://data.bnf.fr/13195155/groupe_des_six/studies/"></pre>
<p>Zone 600 Note d'information →</p>	<pre><rdagroup2elements:corporateHistory xml:lang="fr">Dénomination inventée par Henri Collet dans un article de Comoedia intitulé " Un livre de Rimsky et un livre de Cocteau, les Cinq russes, les Six français et Eric Satie" (16 janvier 1920), réunissant un groupe de six musiciens</rdagroup2elements:corporateHistory> </rdf:Description></pre>
<p>Zone 110 Forme retenue des collectivités →</p>	<pre><rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb13195155c"> <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/> <skos:prefLabel xml:lang="fr">Groupe des six</skos:prefLabel> <bnf-onto:FRBNF rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">13195155</bnf-onto:FRBNF> <rdfs:seeAlso rdf:resource="http://catalogue.bnf.fr/ark:/12148/cb13195155c"/></pre>
<p>Zone 610 Source →</p>	<pre><owl:sameAs rdf:resource="http://viaf.org/viaf/121931049"/> <skos:editorialNote xml:lang="fr">Grove 6</skos:editorialNote></pre>

	<pre><foaf:focus rdf:resource="http://data.bnf.fr/ark:/12148/cb13195 155c#foaf:Organization"/> </rdf:Description></pre>
--	---

3) L'exploitation des liens entre notices bibliographiques et notices d'autorité

Les liens entre notices bibliographiques et notices d'autorité sont exploités pour représenter les relations entre œuvres et manifestations dans data.bnf.fr. Cette relation est exprimée par la classe de l'ontologie RDA `rdarerelationships:workManifested`. Les liens MARC sont donc exploités pour FRBRiser les données.

Par exemple, les relations entre l'œuvre musicale Les Mariés de la Tour Eiffel et les manifestations se rapportant à cette œuvre apparaissent de cette manière. Les zones de la notice bibliographique 144 (point d'accès titre uniforme musical) et 744 (point d'accès secondaire titre uniforme musical) sont exploités.

<pre><rdf:RDF> <rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb38171863f#frbr:Expression"><rdarerelationships:workMan ifested rdf:resource="http://data.bnf.fr/ark:/12148/cb139637319#frbr:Work"/></rdf:Description> <rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb39492787g#frbr:Expression"><rdarerelationships:workMa nifested rdf:resource="http://data.bnf.fr/ark:/12148/cb139637319#frbr:Work"/></rdf:Description> <rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb394927864#frbr:Expression"><rdarerelationships:workMa nifested rdf:resource="http://data.bnf.fr/ark:/12148/cb139637319#frbr:Work"/></rdf:Description> <rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb32952811k#frbr:Expression"><rdarerelationships:workMa nifested rdf:resource="http://data.bnf.fr/ark:/12148/cb139637319#frbr:Work"/></rdf:Description> Source : http://data.bnf.fr/13963731/groupe_des_six_les_maries_de_la_tour_eiffel/rdf.xml</pre>
--

L'URI <http://data.bnf.fr/ark:/12148/cb139637319> correspond à l'œuvre « Les mariés de la Tour Eiffel » du Groupe des Six. Les manifestations liées dans l'extrait RDF sont :

- a) Les mariés de la Tour Eiffel [Enregistrement sonore] / Louis Durey [et le groupe des Six], comp. ; Orchestre National de l'ORTF ; Darius Milhaud, dir. . Les soirées de Valfère, op. 96 / Louis Durey, comp. ; Quintette à vent de Paris. - Paris : Adès ; Maurepas : distrib. Adès, 1989 (DL). - 1 disque compact (1 h 02 min 43 s) : AAD + 1 brochure (8 p.) ; 12 cm.²⁶

²⁶ <http://catalogue.bnf.fr/ark:/12148/cb38171863f/>

- b) Les Mariés de la Tour Eiffel [Spectacle] / Jean Cocteau. Saint-Omer, 1982²⁷
- c) Les Mariés de la Tour Eiffel [Spectacle] : comédie-ballet / Jean Cocteau ; spectacle de A.C.T.E.S., 1981.²⁸
- d) Les mariés de la Tour Eiffel... [Texte imprimé] : [Paris, Théâtre des Champs-Élysées, 18 juin 1921] ; Suivi de : les Chevaliers de la Table ronde : pièce en 3 actes / Jean Cocteau. - Paris : "l'Avant-scène", 1966 (Le Mans : impr. commerciale). - 86 p. : ill. en noir et en coul., pl., couv. ill. ; 26 cm.²⁹

4) L'exploitation des liens entre notices d'autorité

Les liens entre notices d'autorité sont également exploités, par exemple entre une page sur une œuvre et une page sur une partie d'œuvre, par la propriété ore:aggregates.

Par exemple, « Les mariés de la Tour Eiffel » est une œuvre composée de plusieurs parties :

- a) [La baigneuse de Trouville et Discours du général. FP 23] par Francis Poulenc³⁰, qui elle-même comprend :
 - a. [Discours du général. FP 23, no 1]³¹
 - b. [La baigneuse de Trouville. FP 23, no 2]³²
- b) Les mariés de la Tour Eiffel. Ritournelle] par Georges Auric³³

Ce qui donne en RDF :

```

<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb139637319#frbr:Work">
<ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb13917063x#frbr:Work"/>
<ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb140194397#frbr:Work"/>
<ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb15511395h#frbr:Work"/>
<ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb13917044n#frbr:Work"/>
<ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb155113976#frbr:Work"/>
[...]</rdf:Description>

```

IV L'enrichissement des données et la création de contenu

Les fonctionnalités de data.bnf.fr permettent en outre d'enrichir les données existantes dans le catalogue, de réutiliser les données en MARC afin d'une part de créer de nouvelles données à partir

²⁷ <http://catalogue.bnf.fr/ark:/12148/cb39492787q> Il s'agit d'une notice de spectacle qui, à la BnF, n'est ni une notice d'autorité, ni une notice bibliographique. Elle sert à regrouper l'ensemble des documents se rapportant à un même spectacle décrits dans BnF catalogue général.

²⁸ <http://catalogue.bnf.fr/ark:/12148/cb394927864>. Il s'agit également d'une notice de spectacle.

²⁹ <http://catalogue.bnf.fr/ark:/12148/cb32952811k>

³⁰ <http://catalogue.bnf.fr/ark:/12148/cb140194397>

³¹ <http://catalogue.bnf.fr/ark:/12148/cb13917063x>

³² <http://catalogue.bnf.fr/ark:/12148/cb13917044n>

³³ <http://catalogue.bnf.fr/ark:/12148/cb15511395h>

de croisement de données existantes, soit par alignement avec d'autres ressources. Data.bnf.fr joue donc un rôle important de création de contenu et est un outil de création de connaissance.

1) Le croisement de données existantes par alignement

Ce croisement de données existantes sert à la création de « nouvelles » données, qui sont en fait implicites dans les catalogues car non mises en relation entre elles.

C'est le cas par exemple pour l'alignement de notices bibliographiques avec les notices d'autorité titres pour ajouter dans data.bnf.fr de nouvelles manifestations liées aux pages œuvres. Les catalogues de la BnF ne sont pas FRBRisés. Les pratiques professionnelles actuelles de la BnF conditionnent la création de notices d'autorité titres textuels aux règles suivantes : liens systématiques uniquement pour des cas de figure précis (par exemple : classiques anonymes, textes fondamentaux des religions, œuvres liturgiques, œuvres de l'Antiquité...pour les œuvres textuelles,) ; musique classique pour les œuvres musicales). Concernant les autres types d'œuvres (par exemple : la littérature moderne et contemporaine), la création d'une notice d'autorité titre est conditionnée par l'existence de variantes de titre dans la langue originale de l'œuvre ou à celle d'études sur l'œuvre. Dans ce dernier cas de figure, seules les notices bibliographiques d'études sur les œuvres sont liées à la notice d'autorité, et non les éditions de l'œuvre en question. Un nombre important de notices bibliographiques n'est donc pas rattaché à une notice d'autorité titre. Il existe en outre un nombre important de notices bibliographiques anciennes, issues de conversions rétrospectives et non mises à jour, qui ne sont pas rattachées aux notices d'autorité. Un programme d'alignement par comparaison de chaîne de caractères et indices de probabilité a permis d'accrocher dans data.bnf.fr ces notices bibliographiques aux notices d'autorité.

C'est ainsi que la notice bibliographique citée ci-dessus

Les mariés de la Tour Eiffel... [Texte imprimé] : [Paris, Théâtre des Champs-Élysées, 18 juin 1921] ; Suivi de : les Chevaliers de la Table ronde : pièce en 3 actes / Jean Cocteau. - Paris : "l'Avant-scène", 1966 (Le Mans : impr. commerciale). - 86 p. : ill. en noir et en coul., pl., couv. ill. ; 26 cm. »³⁴

n'est pas rattachée à la notice d'autorité œuvre « Les mariés de la Tour Eiffel ». Le programme d'alignement de data.bnf.fr a permis qu'elle figure néanmoins dans la page œuvre.

On peut également mentionner la **création de pages sur les dates**, regroupant les auteurs et collectivités nés ou disparus au cours d'une année donnée, les œuvres créées ou les éditions publiées cette même année³⁵. Les zones MARC suivantes sont exploitées :

Libellé	Zone INTERMARC exploitée
Auteurs nés en ...	Zone 008, position 28-31 (année de naissance des noms de personne)
Auteurs morts en ...	Zone 008, position 38-41 (année de décès des noms de personne)
Début d'activité en ...	Zone 008, position 47-51 (pour les noms de personne et de collectivité)

³⁴ La notice fut créée par la conversion rétrospective du Catalogue général auteurs pour les acquisitions faites en 1960 et 1969.

³⁵ Voir par exemple la page sur l'année 1981 <http://data.bnf.fr/what-happened/date-1981>.

Fin d'activité en...	Zone 008, position 52-56 (pour les noms de personne et de collectivité)
Œuvres créées en...	Zone 008, position 27-36 (date de composition d'une œuvre dans les notices d'autorité titres)
Œuvres terminées en...	Zone 008, position 37-16 (date de fin de composition de l'œuvre dans les notices d'autorité titres)
Enregistrements en...	Zone 008, position 07-11 et 245 \$d enregistrement sonore
Représentations théâtrales en...	Zone 008, position 07-11 et 245 \$d spectacle
Publications en...	Zone 008, position 07-11 et 245 \$d texte imprimé

Le suivi des liens entre différentes notices d'autorité permet de représenter les relations, les collaborations ou les héritages intellectuels des auteurs. **Le champ des pages auteurs « auteurs en relation »³⁶** est fondé sur le suivi des liens entre notices d'autorités et notices bibliographiques en exploitant les codes de fonctions.

Par exemple, se dessine le réseau suivant autour d'une notice bibliographique en Marc :

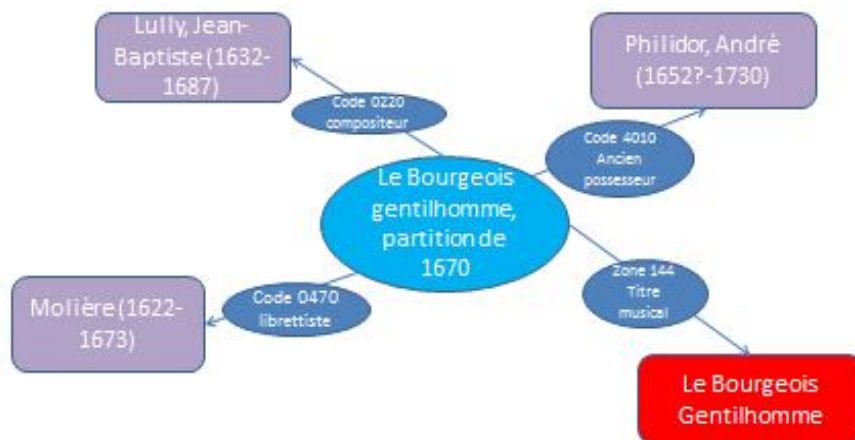
```

100 $313896861 $w 0 b $a Lully $m Jean-Baptiste $d 1632-1687 $4 0220
144 0 $313915750 $w....b.fre. $a Le | bourgeois gentilhomme $k LWW 43
245 1 $a Le // |Bourgeois [sic] // Gentil-homme // Comédie-Ballet // Donné par Le Roy a toutte sa
Cour // dans le chasteau de Chambord au Mois // D'octobre 1670 // fait par Monsieur de Lully Sur
Intendant // de la Musique du Roy et par Le Sieur // Molliere $d Musique manuscrite

700 $312452843 $w.0 .b.....$a Philidor $m André $d 1652?-1730 $4 0360 $4 0270 $4 4010
700 $311916418 $w 0 2b.....$a Molière $d 1622-1673 $4 0470

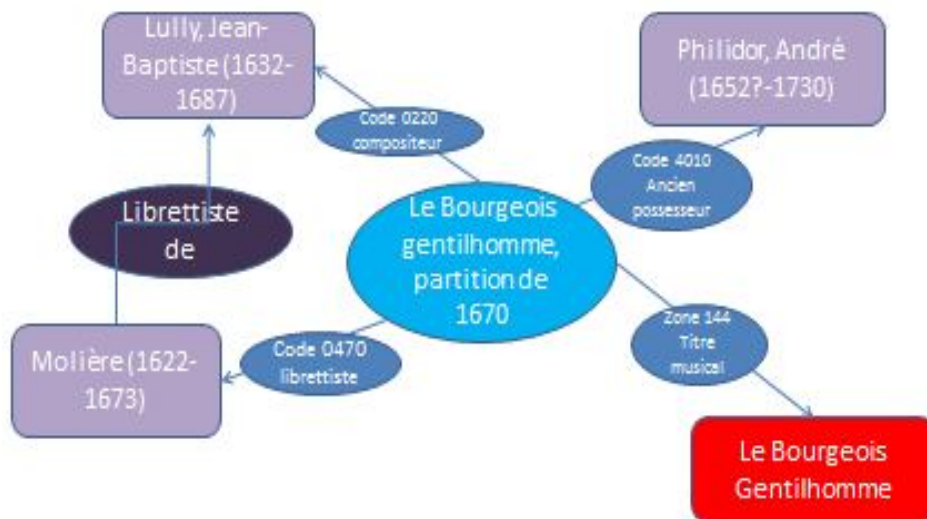
```

L'exploitation des relations entre la notice bibliographique et les notices d'autorité personnes (zones 100 et 700) et œuvre musicale (zone 144) ainsi que des codes de fonction pour les personnes (\$4 des zones 100 et 700) permet de dessiner le réseau suivant :



³⁶ Voir par exemple la page auteur de Jean-Baptiste Lully http://data.bnf.fr/13896861/jean-baptiste_lully/

Les technologies de data.bnf.fr permettent de déduire une relation entre Jean-Baptiste Lully et Molière :



Data.bnf.fr permet donc d'enrichir par déduction les données en MARC en exploitant l'ensemble des données.

2) L'alignement sur d'autres ressources

Data.bnf.fr permet en outre d'ouvrir les données à des ressources de la BnF, hors catalogues. C'est le cas, par exemple, du dépôt légal du Web³⁷. Par exemple les notices d'autorité collectivités contiennent un champ INTERMARC 606 « site internet de la collectivité ». Data.bnf.fr permet, à partir de cette indication, de consulter les pages archivées depuis une date déterminée par le dépôt légal du Web.

Parc naturel régional des Ballons des Vosges

	Pays :	France
	Langue :	français
	Création :	1989
	Activité commerciale :	Éditeur
	Site(s) Web :	http://www.parc-ballons-vosges.fr/ (Site archivé par la BnF depuis le 01/01/2001)
	Adresse :	1 cour de l'Abbaye, 68140, Munster. office: 03.89.77.90.34, fax: 03.89.77.90.30.
		+ Plus de détails

Les technologies du Web sémantique, alliées à une licence de données ouvertes, permettent de pousser plus loin les données, de les ouvrir largement pour les mettre dans le Web. C'est ainsi que,

³⁷ Sur le dépôt légal du Web, voir : « La situation du dépôt légal de l'Internet en France : retour sur cette nouvelle législation, sur sa mise en pratique depuis cinq ans, et perspectives pour le futur » Peter Stirling and Gildas Illien (Auteurs principaux) Pascal Sanz and Sophie Sepetjan (Autres auteurs, intervenants à la conférence de l'IFLA), <http://conference.ifla.org/past/ifla77/193-stirling-fr.pdf>

par exemple, les données de data.bnf.fr sont disponibles dans data.gouv.fr, plateforme rassemblant les jeux de données ouvertes de collectivités publiques françaises³⁸.

Conclusion

La structuration et les liens entre les données permis par les formats MARC sont facilement exploitables pour pousser les données des catalogues de bibliothèques dans le Web de données. Cela permet non seulement de sortir les données des catalogues, mais aussi de les FRBRiser sans réécrire les catalogues. Certaines fonctionnalités d'INTERMARC ont servi de base aux mises à jour d'UNIMARC dans l'optique de la FRBRisation des données. C'est par exemple le cas de la récupération de la zone auteur pour les notices d'autorité auteur-titre. De nouvelles zones ont été créées dans UNIMARC pour rendre compte de ces besoins. Les données peuvent être enrichies par croisement au sein d'un même catalogue. Cela permet d'explicitier les données implicites et de créer de nouvelles données. Elles peuvent également être croisées avec des ressources Web extérieures aux catalogues et déposées dans des puits de données ouvertes pour permettre leur récupération la plus large. Cela permet de mettre les données réellement dans le Web.

Les formats MARC sont donc un tremplin pour relier les outils bibliographiques et les outils numériques, qui ne dialoguent pas naturellement. Le MARC est utilisé comme base sur laquelle les technologies du Web permettent ont permis d'allier ce qui est naturellement sur le Web et métadonnées descriptives. Il s'agit donc d'un dépassement, mais comme tous les dépassements, il n'implique pas un reniement de la situation antérieure. Au contraire, cela souligne plus que jamais l'importance du travail sur les données, de la structuration des données et des liens entre les données primaires. Cela permettra aux publics les plus variés d'utiliser les données produites à l'origine dans les catalogues et ouvertes à d'autres communautés, d'autres usages, parfois très lointains de ceux initialement prévus. Par exemple, environ 20% des visiteurs de l'exposition virtuelle de la BnF *L'art d'aimer au Moyen-âge : le Roman de la rose*³⁹ cherchaient des modèles pour des déguisements médiévaux. On ne peut plus séparer de manière étanche l'usage bibliographique des usages extra-bibliographiques. Le réinvestissement des données bibliographiques et des données d'autorité en utilisant les technologies du Web sémantique permet de les adapter à ces différents usages qui ne sont pas exclusifs les uns des autres.

³⁸ <http://www.data.gouv.fr/>

³⁹ <http://expositions.bnf.fr/aimer/>