



## **L'infrastructure de l'information et les besoins des utilisateurs : tout le pouvoir aux données structurées!**

**Vincent Boulet**

Bibliothèque nationale de France  
Département de l'Information bibliographique et numérique  
Paris, France

Meeting:

80 — *Inspired moments in cataloguing* — Cataloguing

### **Résumé :**

*Les bibliothèques peuvent se prévaloir d'une ancienne tradition de normalisation et de produire des données structurées en nombre. Cependant, leurs ressources, notamment numériques, sont rarement visibles en dehors des catalogues et des portails de bibliothèques numériques. Pour prendre la place qui leur revient dans les nouveaux modes de recherche induits par les pratiques du Web, les données des bibliothèques doivent être présentes là où passent les utilisateurs et mettre à disposition leurs données sans préjuger des usages possibles. Les données structurées et les normes de production des métadonnées, loin d'être un handicap, sont un précieux atout. L'utilisation des technologies du Web sémantique et de la structure des documents Web leur ouvre la porte d'utilisations nouvelles. Le projet de la BnF, [data.bnf.fr](http://data.bnf.fr), en est un exemple.*

### **Introduction**

En 1378, le garde de la bibliothèque du roi<sup>1</sup> de France Charles V<sup>2</sup>, Gilles Malet<sup>3</sup>, rédige un inventaire des manuscrits qui y sont conservés<sup>4</sup>. Il s'agit du premier « catalogue » de la Bibliothèque royale, future Bibliothèque nationale. Il répond à trois besoins : répertorier les ouvrages, indiquer leur localisation dans la bibliothèque et éviter leur dispersion qui est susceptible de se produire à l'occasion de la succession royale. En 1383, Jean Blanchet opère un récolement des ouvrages et indique dans la marge du catalogue si le livre indiqué se trouve bien dans la bibliothèque (par la mention « *il y est* ») ou s'il a été « donné ». Le catalogue est alors utilisé pour faire le point sur l'état de la bibliothèque après la mort de Charles V. Pour répondre à l'ensemble de ces besoins, il est nécessaire d'identifier les manuscrits de manière

<sup>1</sup> <http://catalogue.bnf.fr/ark:/12148/cb16221102z>

<sup>2</sup> <http://catalogue.bnf.fr/ark:/12148/cb119360912/>

<sup>3</sup> <http://catalogue.bnf.fr/ark:/12148/cb10320056q>

<sup>4</sup> BnF, Paris, Département des manuscrits, Français 2700. Consultable sur Gallica : <http://gallica.bnf.fr/ark:/12148/btv1b90590207>

univoque. Pour cela, sont établies des descriptions courtes dans lesquelles on retrouve un souci de structuration. C'est ainsi qu'on y trouve « *une Bible en un volume en françois couverte de cuir rouge a IIII fermoirs* » ou encore « *unes croniques de Godeffroy de Villon de la conquete de la terre d'oultre mer en prose bien viellez couverte de cuir a queue blanc* ». La « notice » doit donc inclure au moins un titre et une description de la reliure. Pour les ouvrages non anonymes, on y adjoint le nom de l'auteur. On peut également mentionner, de manière facultative, la langue et le nombre de volumes pour les œuvres qui sont susceptibles d'être rédigées en différentes langues et dont le manuscrit peut ne pas être en un seul volume. Il s'agit donc d'un embryon de normalisation, pour répondre à des fins précises, ici des besoins de gestion interne. La question que se posent les bibliothécaires du roi du XIV<sup>e</sup> siècle est déjà : quels sont les éléments d'information nécessaires pour identifier un manuscrit, c'est-à-dire pour le distinguer d'un autre, en fonction des besoins ?

On voit donc **que les bibliothèques peuvent se prévaloir d'une tradition ancienne de normalisation des données descriptives**. Mais ces normes, élaborées dans un contexte catalographique et culturel donné, avec des objectifs précis, sont très diverses. Elles sont aussi le produit de leur époque de rédaction et sont donc évolutives. Ce caractère évolutif se répercute naturellement dans les descriptions produites à partir de ces normes.

Pourtant, la mise en place d'Internet et la diffusion de nouveaux modes d'accès à l'information nécessite de **revoir l'infrastructure des données des bibliothèques qui restent difficiles à trouver sur le Web**. Les données descriptives des bibliothèques sont donc confrontées à un paradoxe qu'il convient de dépasser : ce sont des données riches mais invisibles.

On pourrait comparer la situation de ces données à un village au bord de l'autoroute, sans sortie d'autoroute pour y accéder. L'enjeu est d'en faire, au contraire, un lieu de passage.

## I- **Les défis et les atouts des données des catalogues pour les mettre sur la route des utilisateurs**

### ***Défis internes et défis externes posés par le Web aux données des bibliothèques***

C'est une évidence : les défis posés par le Web sont considérables. De plus, comme les pratiques du Web évoluent, les besoins des utilisateurs évoluent également.

Les défis pour le traitement des données par les bibliothèques sont de plusieurs ordres, qu'ils soient internes ou externes au monde des bibliothèques.

**Les défis externes se rapportent à la manière dont le Web s'organise et aux usages qui en sont faits**. Cela pose la question de l'exploitation des données des catalogues ou des ressources numériques par **les moteurs de recherche et plus généralement par tous les programmes informatiques**. En effet, si les données des bibliothèques sont présentes sur Internet, elles demeurent peu indexées par les moteurs de recherche et, lorsque cela est le cas, il est difficile de trier les résultats pertinents. De plus, les catalogues sont souvent construits sur le modèle des bases de données relationnelles, inexploitable pour des moteurs de recherche. Cela oblige les utilisateurs à passer par les portails des bibliothèques, dont ils ne

connaissent pas toujours l'existence. En outre, les recherches effectuées sur le Web ne se font pas avec un langage contrôlé et répondent aux besoins les plus variées de recherche d'information. L'ensemble de ces facteurs fait que les ressources liées aux données descriptives, même si elles sont librement accessibles comme les ressources numériques, sont invisibles dans la forêt du Web.

**Les défis internes sont dus à la structure des catalogues des bibliothèques, qui sont rédigés dans des formats spécifiques et prennent la forme de silos isolés.** Les normes ISBD et les formats MARC sont particulières au monde des bibliothèques parce qu'ils répondent aux besoins spécifiques de la description et de la recherche en bibliothèque. Par exemple, la première norme ISBD a été rédigée dans l'esprit de répondre aux obligations posées aux agences bibliographiques nationales pour établir les bibliographies nationales.

De plus, la mise en ligne des données concernant également des collections pour lesquelles l'usage des normes ISBD et des formats MARC n'est pas adapté. La description des fonds d'archives voire, pour le cas des bibliothèques françaises, des collections de manuscrits nécessite l'emploi de normes particulières (ISAD (G)<sup>5</sup> ou DACS<sup>6</sup> pour les fonds d'archives, ou en France DeMarch<sup>7</sup>). Pour rendre compte de la structure hiérarchisée des fonds d'archives et du principe d'héritage de l'information<sup>8</sup>, le format XML-EAD est le plus adapté.

**Les données des bibliothèques sont donc à la fois normalisées et diverses. Des normes et des formats différents s'appliquent selon la nature des documents traités.** Si l'on ajoute les ressources numérisées et d'autres services proposés par les bibliothèques, on aboutit à la constitution de silos de données différents.

Par exemple, la BnF a deux catalogues : *BnF-Catalogue général*<sup>9</sup> pour les descriptions en MARC et *BnF-Archives et manuscrits*<sup>10</sup> pour les descriptions en EAD. Il faut y ajouter la bibliothèque numérique Gallica<sup>11</sup>, les expositions virtuelles<sup>12</sup>, la base des manuscrits enluminés Mandragore<sup>13</sup>, le catalogue du Centre national de la littérature pour la jeunesse « la Joie par les livres »<sup>14</sup>, la banque d'images ou encore des bases internes Access ou Excel du département des Monnaies, Médailles et Antiques et des bases spécialisées : description des reliures conservées à la Réserve des livres rares, catalogue dit « Renouard » des éditions parisiennes du XVIe s.

Les ressources, riches, sont donc éclatées et peu visibles si l'on considère l'économie générale du Web. Pour autant, les données de bibliothèques peuvent se prévaloir d'atouts précieux.

---

<sup>5</sup> Norme générale et internationale de description archivistique, 2000 ([http://www.icacds.org.uk/fr/ISAD\(G\).pdf](http://www.icacds.org.uk/fr/ISAD(G).pdf), consulté le 2012-05-14).

<sup>6</sup> Describing Archives : a content standard, 2004. Implémentation américaine d'ISAD(G) et d'ISAAR-CPF qui prend en compte la description des collections de manuscrits et qui est compatible avec le chapitre 4 des AACR-2 qui indique quelques règles générales sur le traitement des fonds d'archives conservés en bibliothèque.

<sup>7</sup> DeMarch-Recommandation de description des fonds d'archives et des manuscrits, 2010. (<http://www.bivi.fonctions-documentaires.afnor.org/livres-blancs/demarch-recommandation-de-description-des-manuscrits-et-fonds-d-archives>, consulté le 2012-05-14). Élaborée par le groupe AFNOR chargé de la « description des manuscrits modernes et contemporains), cette recommandation s'applique aux manuscrits et fonds d'archives modernes et contemporains conservés en bibliothèque. Elle suit les règles éditées par ISAD(G).

<sup>8</sup> L'héritage de l'information est un des principes fondamentaux de la description archivistique : les éléments indiqués à un niveau donné de la description doivent être valables pour l'ensemble des niveaux inférieurs.

<sup>9</sup> <http://catalogue.bnf.fr> (consulté le 2012-05-14).

<sup>10</sup> <http://archivesetmanuscrits.bnf.fr/> (consulté le 2012-05-14).

<sup>11</sup> <http://gallica.bnf.fr/> (consulté le 2012-05-15).

<sup>12</sup> <http://expositions.bnf.fr/> (consulté le 2012-05-15).

<sup>13</sup> <http://mandragore.bnf.fr/> (consulté le 2012-05-15).

<sup>14</sup> <http://lajoieparleslivres.bnf.fr/> (consulté le 2012-05-15).

## **Les atouts : prendre parti des atouts des données structurées des bibliothèques**

Le premier atout tient au fait que **l'hétérogénéité globale des catalogues correspond en réalité à la juxtaposition de plusieurs blocs homogènes**. Par exemple, à la BnF, le « bloc » normalisé selon les normes ISBD et le « bloc » normalisé selon ISAD(G). Si l'on prend en compte les conversions rétrospectives, on peut citer le « bloc » normalisé selon les règles de description du catalogue général par ordre alphabétique d'auteur<sup>15</sup> ou les règles de description des manuscrits latins fixées au XIXe s.

De plus, **ces descriptions sont homogénéisées par les données d'autorité, dont les atouts sont nombreux**. En premier lieu, les entités décrites sont diverses : personnes, familles, collectivités, titres, noms géographiques, noms communs. Celles-ci sont pourvues **d'identifiants** pérennes. La BnF utilise des identifiants ARK<sup>16</sup>. Les données d'autorité s'appliquent à des ressources de plus en plus variées : fonds d'archives et collections de manuscrits, expositions virtuelles... Elles fournissent un ensemble de données d'identification précises, fiables et vérifiées.

Les données d'autorité sont un atout précieux des bibliothèques dans le monde du Web. Elles prennent une importance nouvelle. Le code de catalogage international **RDA** leur accorde une place centrale<sup>17</sup>. Deux des trois scénarios d'implémentation prévoient l'établissement de liens entre les entités FRBR du groupe 1<sup>18</sup> et celles du groupe 2<sup>19</sup> et du groupe 3<sup>20 21</sup>. Il ne s'agit plus seulement de contrôler la description bibliographique par des points d'accès normalisés mais de la consolider en explicitant systématiquement les personnes, les collectivités, les familles, les concepts, les lieux et les événements liés à une œuvre, une expression, une manifestation, un item ainsi que la nature du lien qui les unit. La question des **liens entre le fichier bibliographique et le fichier d'autorité**, qui se fait à la BnF, se pose dès lors comme une évidence.

Un autre atout des données d'autorité est de **pouvoir répondre aux besoins et aux pratiques variées des utilisateurs**. Par le jeu des formes rejetées des notices d'autorité, ils ont la possibilité de trouver l'entité recherchée quelle que soit la forme retenue établie selon les normes professionnelles. De plus, certains formats, comme EAC-CPF<sup>22</sup> ou InterMarc utilisé à la BnF, permettent de saisir plusieurs formes retenues pour une même notice d'autorité

---

<sup>15</sup> Entreprise réalisée à la Bibliothèque nationale à partir de 1895 qui se poursuit jusqu'en 1981 et qui totalise 231 volumes.

<sup>16</sup> Archival Resource Key

([http://www.bnf.fr/fr/professionnels/s\\_informer\\_autres\\_numeros/a\\_ark\\_autres\\_numeros.html](http://www.bnf.fr/fr/professionnels/s_informer_autres_numeros/a_ark_autres_numeros.html), consulté le 2012-05-14).

<sup>17</sup> Voir le schéma du modèle de données de RDA où la place des données d'autorités est évidente :

[http://www.rdatoolkit.org/backgroundfiles/RelationshipsOverview\\_10\\_9\\_09.pdf](http://www.rdatoolkit.org/backgroundfiles/RelationshipsOverview_10_9_09.pdf) (consulté le 2012-05-15).

<sup>18</sup> Produits d'une activité intellectuelle ou artistique qui sont nommés ou décrits dans les notices bibliographiques : *oeuvre, expression, manifestation* et *item*.

<sup>19</sup> Entités auxquelles revient la responsabilité du contenu intellectuel ou artistique, de la production matérielle et de la distribution, ou de la gestion juridique de ces produits : *personne* et *collectivité*.

<sup>20</sup> Sujet de l'activité intellectuelle ou artistique : *concept, objet, événement* et *lieu*

<sup>21</sup> Le scénario 1 prévoit l'implémentation totale du modèle de données FRBR, avec des liens établis aux niveaux œuvre, expression, manifestation et item vers les entités du groupe 2 et de l'entité œuvre vers les entités du groupe 3. Le scénario 2 prévoit des liens entre le fichier bibliographique (comprenant les niveaux expression et manifestation) et le fichier d'autorité (au sein duquel figure le niveau œuvre du groupe 1, ainsi que les entités des groupes 2 et 3).

<sup>22</sup> Encoded Archival Context - Corporate bodies, Persons, Families. Schéma XML publié en 2010 pour appliquer la norme ISAAR-CPF. <http://eac.staatsbibliothek-berlin.de/> (consulté le 2012-05-15).

répondant à des besoins différents, dans des langues différentes<sup>23</sup>. Cela permet de conférer davantage de souplesse aux données d'autorité, et donc d'élargir leurs potentialités d'utilisation et de réutilisation.

Les données des bibliothèques sont donc riches d'éléments finement structurés, vérifiés et validés, reposant sur une ancienne tradition de normalisation. Leur inconvénient principal réside dans le fait que ces données sont essentiellement cantonnées dans les catalogues de bibliothèques. Par conséquent, l'utilisateur doit faire preuve d'une démarche volontariste pour pouvoir les utiliser : non seulement entrer à l'intérieur des catalogues mais savoir comment les manier pour mettre à profit l'ensemble de leurs potentialités.

Il convient donc de mettre en contact les données structurées avec les utilisateurs, à l'endroit où ceux-ci passent.

### ***A partir des atouts, donner des éléments de réponse aux défis***

Si les données des bibliothèques sont peu visibles, rédigées dans des formats spécifiques et ayant pris la forme de silos isolés, cela justifie-t-il de se faire *tabula rasa* et faire passer par pertes et profits leurs atouts et les pratiques professionnelles qui y sont associées ? Ce serait reconnaître qu'il y a incompatibilité entre la structuration des données du monde des bibliothèques et la structuration plus faible des données sur le Web.

A la BnF, nous avons pris le parti de considérer qu'il n'y a pas contradiction mais **complémentarité**. En conséquence, nous avons fait le choix de nous **fonder sur l'existant sans réécrire ni refonder les catalogues**. Cela se justifie pour des raisons stratégiques : les données structurées sont un atout. Mais également pour des raisons économiques.

En outre, cela permet de concilier plusieurs approches parallèles. Le Web a une **pratique sélective**. Les catalogues de bibliothèques ont une visée exhaustive. Il est inutile et contre-productif d'essayer de forcer l'une de ces approches à se fondre dans l'autre, au risque de les dénaturer et de leur faire perdre de leur efficacité respective. Cela justifie une approche de sélection, de choix. Le fait d'être sur le Web, avec la publication de pages HTML, ne nécessite pas de traiter l'ensemble des données présentes dans les catalogues. Il ne s'agit pas de remplacer les catalogues mais de donner une nouvelle visibilité et de permettre une large réutilisation à des **données choisies**. Ce choix s'est porté sur les données qui permettent d'identifier les ressources de manière univoque pour accéder à la fois aux descriptions complètes des catalogues et aux ressources numérisées. Cela implique l'utilisation des langages du Web, et non les langages propres aux bibliothèques.

Mais rendre les données utilisables pour l'ensemble des besoins possibles des utilisateurs implique d'aller au-delà d'un affichage HTML. **Exposer les données dans le « linked data »** permet que les données soient aussi utiles que possible pour le grand public comme pour les professionnels. L'utilisation des standards et des technologies du Web sémantique permet d'exposer des données structurées exploitables par des machines et de répondre tant à des **enjeux internes qu'à des enjeux externes**. En interne, cela met en communication différents silos de données qui utilisent non seulement un format mais également une structure de données différentes, sans dénaturer la spécificité des silos d'origine. En externe, les

---

<sup>23</sup> Par exemple : la notice d'autorité d'Aristophane (0445 ?-0386 ?) : <http://catalogue.bnf.fr/ark:/12148/cb12262274m> (consulté le 2012-05-15). La première forme retenue, en français, est une forme d'usage courant (par exemple : un lycéen qui étudie « Les Nuées »). La deuxième forme est en latin, forme traditionnelle utilisée par les spécialistes de l'Antiquité grecque. Les troisième et quatrième forme sont en grec ; elles correspondent à la langue et à l'écriture la plus couramment associée à la personne.

technologies du Web sémantique permettent de faire dialoguer les données de bibliothèques avec celles provenant d'autres métiers, qui utilisent des modèles de données différentes : par exemple le monde des archives, dont le modèle de données est orienté pour décrire des fonds, et celui des musées, dont le modèle de données est conçu pour décrire des objets.

En outre, les technologies du Web sémantique sont les plus adéquates pour permettre une **réutilisation libre des données, dans le cadre de l'ouverture des données publiques** (« *open data* »). La récupération et l'exploitation des données la plus libre possible est un facteur essentiel pour pouvoir mettre celles-ci à disposition du plus grand nombre<sup>24</sup>.

L'application de ces principes permet de dépasser le paradoxe pointé en introduction, entre la grande richesse et l'invisibilité des données des bibliothèques. A la BnF, nous cherchons à le faire dans le cadre du projet data.bnf.fr.

## **II- Un exemple qui prend le parti des données structurées : le projet data.bnf.fr<sup>25</sup>**

Les besoins sont donc multiples : répondre aux attentes et aux usages du Web et à l'enjeu de la réutilisation libre des données. L'enjeu économique et professionnel est également essentiel : il ne s'agit pas de réinventer la poudre mais de se servir des traditions bibliographiques et des données déjà existantes sans « refonder » les catalogues.

Pour appliquer ces principes, la BnF met en œuvre le projet data.bnf.fr, dont une première version a été publiée en ligne en juillet 2011. Il s'agit de mettre à disposition directement sur le Web un accès aux ressources de la BnF, dont les ressources numériques. L'accès à ces ressources est structuré par des « pages auteurs »<sup>26</sup>, des « pages œuvres »<sup>27</sup> et des « pages thèmes »<sup>28</sup> dont les contenus sont issus respectivement des notices d'autorité personnes, familles et collectivité pour les « pages auteurs », des notices d'autorité titres textuels et musicaux pour les « pages œuvres » et des notices d'autorité matière Rameau pour les « pages thèmes ». Chacune de ces pages fournit des liens vers les différentes ressources de la BnF issus des différents silos de données.

### **1) Le périmètre contrôlé du projet**

#### **a- Les principes de fonctionnement**

On l'a vu, les bases de données de la BnF répondent à des objectifs différents et ont des formats adaptés au type de ressources qu'ils décrivent. Pour publier les pages auteurs, œuvres et thèmes, des données sont extraites de ces bases et alignées grâce aux données d'autorité qui les contrôlent. Des ressources d'autres bibliothèques (via le catalogue collectif de France), et d'autres ressources, comme Wikipedia, sont également alignées avec les données de la BnF. Les pages sont publiées dans une interface Web, selon les principes du Web de documents,

---

<sup>24</sup> Le présent article ne se concentre pas sur cet aspect très important. Voir : Gildas Illien, « Are you ready to dive in ? A case for Open data in national libraries », IFLA 2012, National Libraries Section.

<sup>25</sup> <http://data.bnf.fr/>

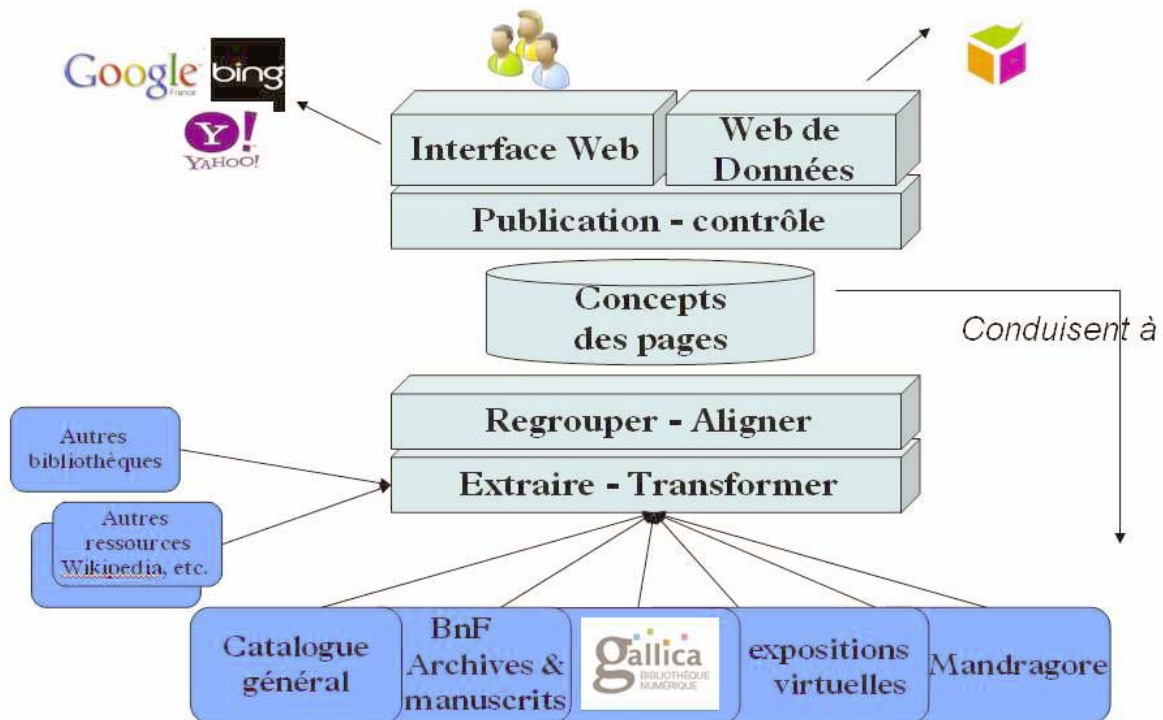
<sup>26</sup> Exemple : <http://data.bnf.fr/11887823/virgile/>

<sup>27</sup> Exemple : [http://data.bnf.fr/12008287/virgile\\_eneide/](http://data.bnf.fr/12008287/virgile_eneide/)

<sup>28</sup> Exemple : [http://data.bnf.fr/11932282/litterature\\_latine/](http://data.bnf.fr/11932282/litterature_latine/)



pour pouvoir être visibles directement par les moteurs de recherche. Elles le sont également en utilisant les technologies du Web sémantique et les bonnes pratiques des *linked data*.



## b- Un périmètre évolutif et la méthode agile

L'ensemble des données des catalogues n'ont pas nécessairement vocation à figurer dans data.bnf.fr. L'objectif du projet est de fournir un accès aux ressources de la BnF issues de ces différentes bases de données via les données structurées des notices d'autorité et non de reproduire les descriptions détaillées des catalogues. De plus, la **montée en charge est volontairement progressive** afin de garder la maîtrise du périmètre des données mis à disposition des utilisateurs. La première version se concentrait sur les œuvres classiques de la littérature française. Le périmètre s'est depuis élargi aux auteurs antiques, aux juristes anciens et aux musiciens, pour lesquels la BnF apporte des informations rares et pertinentes. Les pages consacrées aux auteurs « de la longue traîne » figurent parmi celles les plus consultées. Le projet compte en mai 2012 17 000 pages auteurs, 44 000 pages œuvres et 167 000 pages thèmes, qui regroupent l'accès à 2,5 millions de ressources. En outre, les fonctionnalités associées et le modèle de données RDF utilisé est lui-même susceptible d'évoluer. Le projet est conduit selon la **méthode agile** ce qui permet de répondre aux attentes qui se font jour progressivement autour du projet et de garder la maîtrise du périmètre du projet répondant à ces attentes ainsi que le souci de la fiabilité des données mis en ligne.

## 2) Les moyens

Pour répondre à l'ensemble de ces enjeux, différents moyens sont mis en œuvre : l'application du modèle de données FRBR, l'emploi des technologies du Web sémantique et du Web de données ainsi que l'importance accordée aux alignements sur des référentiels extérieurs.

## a- Le modèle de données FRBR

La structure des données suit le modèle FRBR en l'adaptant aux données existantes. Elle intègre également des ressources pour lesquels le modèle FRBR ne s'applique pas, comme les fonds d'archives.

### **Le groupe 1 des entités FRBR**

La notion **d'œuvre** structure le modèle de données de data.bnf.fr, à partir des notices d'autorité titres textuels et musicaux. Chaque œuvre fait l'objet d'une page particulière, à l'intérieure de laquelle se structurent les autres entités du groupe 1.

Le niveau **expression** n'est, le plus souvent, pas directement visible dans les pages HTML mais il figure dans la modélisation RDF. Il est utilisé en interne pour effectuer des regroupements de manifestations à partir des liens aux auteurs et aux contributeurs ainsi que des critères de type de document et de langue. Il apparaît dans les pages HTML dans certains cas, par exemple pour les spectacles. Un spectacle est considéré comme une expression d'une œuvre<sup>29</sup>.

Les **manifestations** apparaissent dans la page œuvre par le biais de liens avec les notices bibliographiques dans les catalogues ou avec les ressources numérisées.

Les **items** en tant que tels n'apparaissent pas dans la modélisation des données, car data.bnf.fr ne reprend pas l'ensemble du catalogue. Ils sont intégrés dans le modèle de données de data.bnf.fr comme des « manifestations singleton » du modèle FRBRoo, par exemple pour les manuscrits qui sont considérés à la fois comme la ou les premières manifestations d'une œuvre et comme des items.

Les **attributs** des entités sont gérés soit par des liens vers des littéraires, c'est-à-dire du texte libre, soit par des liens vers des référentiels internationaux<sup>30</sup>. Les éléments gérés dans le modèle de données de data.bnf.fr correspondent aux entités FRBR. Pour autant, comme les données proviennent du contenu des notices d'autorité ou des notices bibliographiques de la BnF, il peut y avoir un certain décalage entre le modèle de données de data.bnf.fr et le modèle de données FRBR. La règle est toujours de partir des données existantes.

Les éléments gérés par data.bnf.fr se répartissent ainsi :

---

<sup>29</sup> Exemple : [http://data.bnf.fr/11940749/jean\\_racine\\_phedre/](http://data.bnf.fr/11940749/jean_racine_phedre/)

<sup>30</sup> Voir ci-dessous.



<b>Entité FRBR</b>	<b>Eléments gérés par le modèle de données de data.bnf.fr</b>	<b>Attributs FRBR correspondants</b>	<b>Commentaires</b>
<b>Oeuvre</b>	Date	Date de l'œuvre	
	Titre	Titre de l'œuvre	
	Description	Forme de l'œuvre Autre caractéristique distinctive Public visé Contexte Complétude visée	Selon les données d'autorité sources
	Langue		Pas d'entité FRBR correspondante : ces éléments proviennent des données d'autorité sources
	Sujet (classe Dewey)		
	Sujet (littéral)		
<b>Expression</b>	Type	Forme de l'expression	
	Langue	Langue de l'expression	
<b>Manifestation</b>	Titre	Titre de la manifestation	
	Date	Date d'édition/diffusion	
	Description		Correspond au contenu de la notice bibliographique : peut recouper plusieurs attributs FRBR
	Editeur	Editeur/diffuseur	
	Ouvrage jeunesse		Correspond au contenu de la notice bibliographique
	ISBN	Identifiant de la manifestation	
	Cote		Pour les manuscrits, traités comme des manifestations-singleton
	Désignation de l'édition	Type de support Nombre d'unités matérielles	
	Lieu d'édition	Lieu d'édition	

## Le groupe 2 des entités FRBR et ses liens avec le groupe 1, l'application du modèle de données FRAD

Les pages auteurs regroupent des personnes, des familles et des collectivités. Les personnes peuvent avoir un rôle d'auteur d'une œuvre, ou de contributeur d'une expression (traducteur, adaptateur, illustrateur, préfacier...). Les liens vers les entités du groupe 1 se font par la gestion des codes qui indiquent la responsabilité intellectuelle de la personne sur la ressource décrite.

Les attributs des entités du groupe 2 sont gérés de la manière suivante :

Entité FRBR	Éléments gérés par le modèle de données de data.bnf.fr	Attributs FRBR correspondants	Attributs FRAD correspondants
Personne/ Collectivité	Nom	Nom / autre qualificatif	
	Nom de famille		
	Prénom		
	Naissance		
	Genre		Sexe
	Date	Date	Dates associées à la personne
	Date de naissance		Dates associées à la personne
	Lieu de naissance		Lieu de naissance
	Date de décès		Dates associées à la personne
	Lieu de décès		Lieu de mort
	Langue de la personne		Langue de la personne
	Nationalité		Pays
	Champ d'activité		Domaine d'activité
	Information biographique	Titre	Titre de la personne, Affiliation, Profession/occupation Biographie/histoire Autre information associée à la personne

Les éléments de données gérés dans le modèle de données de data.bnf.fr proviennent à nouveau des notices d'autorité de la BnF, ici les notices personnes et collectivités.

### **Le groupe 3 des entités FRBR**

L'intégration de Rameau dans data.bnf.fr permet d'amorcer l'apparition des entités du groupe 3 des FRBR. Pour le moment, il s'agit des noms communs et des lieux géographiques qui figurent dans le fichier Rameau. L'intégration du fichier d'autorité Noms géographiques, du Département des cartes et plans de la BnF, est à l'étude.

### **L'intégration de données hors du modèles FRBR**

Le modèle de données intègre également des ressources qui n'entrent pas dans le champ d'application des FRBR. C'est le cas des documents archivistiques, dont la structure hiérarchisée n'est pas conçue autour de la notion d'œuvre mais de fonds. Le choix a été fait de les considérer comme des agrégats. Nous nous écartons ici d'autres possibilités de modélisation. Le projet LOCAH<sup>31</sup>, dont l'objectif est d'utiliser les technologies du Web sémantique à partir des données de *Archives Hub* et de COPAC du Royaume-Uni, rend la hiérarchie des descriptions en EAD par les propriétés du Dublin Core<sup>32</sup> **dct:isPartOf**. Le fait de traiter la hiérarchie des fonds d'archives comme des agrégats permet de la faire entrer dans le cadre général du modèle de données FRBR, tout en maintenant sa spécificité.

### **Conclusion : comment faire naître les FRBR à partir des données existantes et des alignements automatisés ?**

L'application du modèle de données FRBR est le cadre général du modèle de données de data.bnf.fr. Les entités du groupe 2 ont des attributs plus proches de FRAD par leur richesse. Tous ces traitements ne sont possibles que grâce à la **qualité des données sources**, dans les notices d'autorité et dans les notices bibliographiques. L'application des FRBR est encadrée par la nature des données existantes dans les catalogues. A partir de ces données, il est cependant possible de créer, dans data.bnf.fr, des éléments de données supplémentaires, qui ne figurent pas explicitement dans le catalogue, par des spécifications **d'alignements automatisés**. C'est par exemple le cas de l'entité œuvre. A la BnF, les notices d'autorité titres textuels sont établies systématiquement pour les œuvres classiques et anonymes ainsi que pour les œuvres ayant fait l'objet d'une étude. Cela ne recouvre donc pas l'ensemble des œuvres mentionnées dans les catalogues. Une étude de recherche et de développement a montré qu'il est possible de créer les pages œuvres pour celles qui ne font pas l'objet d'une notice d'autorité titre, en utilisant les données bibliographiques et les liens avec les notices d'autorité personnes. Des manifestations ayant le même titre et le même auteur se rapportent à la même œuvre. C'est ainsi qu'il sera possible de créer des pages œuvres de manière plus systématique que le sont les notices d'autorité titre. C'est une manière d'exploiter les données existantes pour appliquer le modèle de données FRBR.

### **b- L'emploi des technologies du Web sémantique : la priorité donnée aux ontologies communes**

Le modèle de données FRBR est exposé en utilisant les technologies du Web sémantique. Pour assurer l'utilisation la plus large des données, data.bnf.fr applique les bonnes pratiques du *linked data* et met l'accent sur les ontologies communes, utilisées en dehors du monde des bibliothèques.

---

<sup>31</sup> <http://blogs.ukoln.ac.uk/locah/> et plus précisément <http://blogs.ukoln.ac.uk/locah/tag/modelling/>.

<sup>32</sup> Dont certaines propriétés sont utilisées par data.bnf.fr (voir ci-dessous).

### **i- Les URI et les bonnes pratiques du *linked data***

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.

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<sup>33</sup>.

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

- nommer les ressources avec des URI
- 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
- lorsqu'on déréférence une URI, renvoyer des informations utiles grâce à RDF
- se relier avec d'autres URI pour créer un réseau de liens.

### **ii- Les concepts**

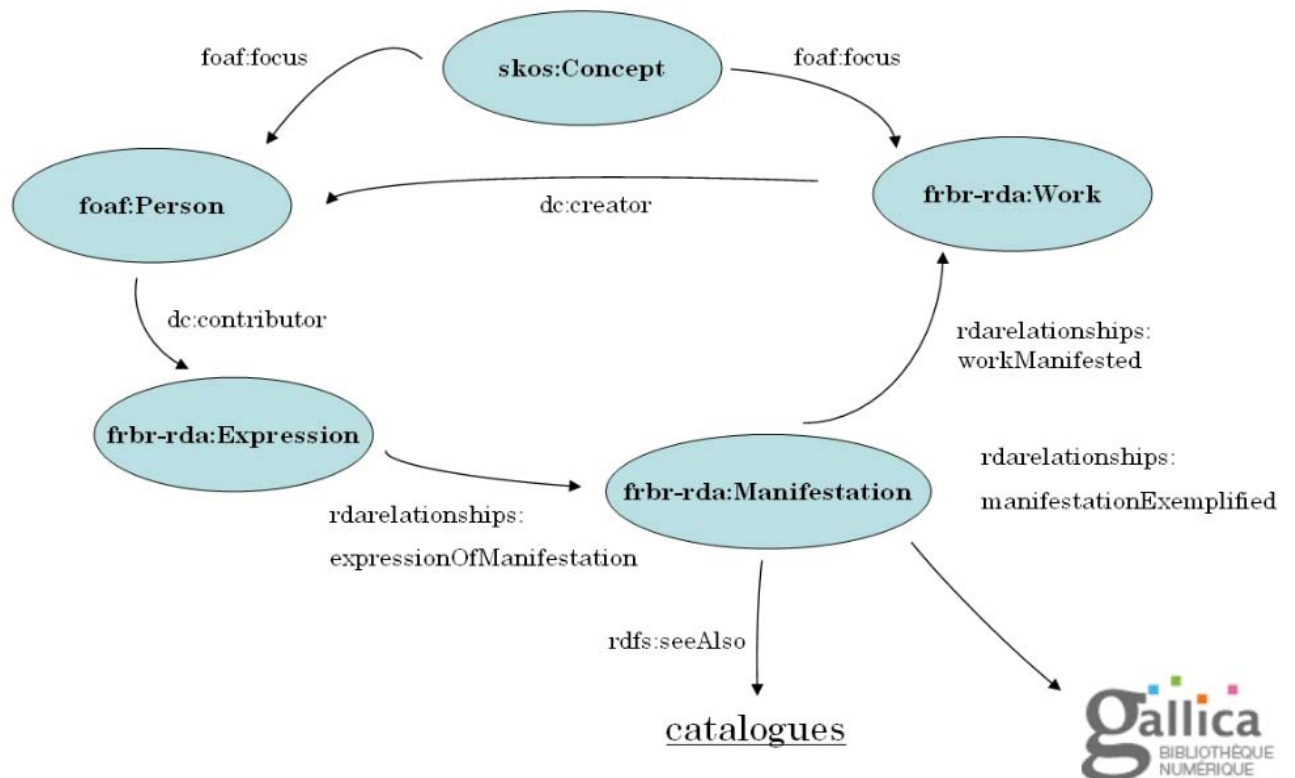
Le modèle de données distingue les **concepts**, au sens large (thèmes, mais aussi œuvres et personnes), et **les œuvres et les auteurs** (personnes ou collectivités). Ainsi, Charles Perrault est à la fois décrit comme concept (représentation de l'auteur dans un contexte donné) et comme personne (informations factuelles sur la vie de l'auteur), en utilisant les vocabulaires spécifiques et pertinents dans chaque cas.

### **iii- Utilisation de langages valables hors du monde des bibliothèques**

La priorité est donnée aux standards communs du Web, largement utilisées 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. 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 communes. Le modèle général de données RDF peut se résumer de la manière suivante :

---

<sup>33</sup> Voir ci-dessous.



### RDFS<sup>34</sup> et OWL<sup>35</sup>

Ces deux vocabulaires contiennent des propriétés et classes très génériques que l'on utilise dans de très nombreux contextes. Dans data.bnf.fr, la propriété **owl:sameAs** permet d'exprimer une équivalence exacte entre deux ressources. Elle est par exemple utilisée pour les liens entre les auteurs et identifiants VIAF qui leur correspondent :

#### Exemple pour Alexandre Dumas : lien de la page auteur de data.bnf.fr vers la notice VIAF

```

<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb119010630">
  <owl:sameAs rdf:resource="http://viaf.org/viaf/51688902" />
</rdf:Description>

```

La propriété **rdfs:seeAlso** permet de renvoyer vers une autre ressource. Elle est utilisée pour les liens entre les manifestations et les URI des notices correspondantes dans *BnF Catalogue général* et *BnF Archives et Manuscrits* ou entre les pages auteurs et les notices d'autorité personnes correspondantes :

<sup>34</sup> <http://www.w3.org/TR/rdf-schema/>

<sup>35</sup> <http://www.w3.org/TR/owl-features/>

### Exemple pour Alexandre Dumas : lien de la page auteur de data.bnf.fr vers la notice d'autorité personne

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb119010630">
<rdfs:seeAlso rdf:resource="http://catalogue.bnf.fr/ark:/12148/cb119010630" />
</rdf:Description>
```

### SKOS<sup>36</sup>

Dans data.bnf.fr, SKOS permet de décrire les concepts au sens large : les auteurs, les œuvres et les thèmes. On utilise la classe principale **skos:Concept**. Les propriétés appropriées **skos:closeMatch**, **skos:exactMatch** décrivent des résultats d'alignements automatiques ou manuels entre des concepts issus de thésaurus distincts, par exemple avec LCSH, Dewey, le Thésaurus de la série W, utilisée dans les archives départementales françaises pour décrire les archives contemporaines.

Par ailleurs, les données du langage d'indexation RAMEAU, soit environ 160 000 notices, ont été converties en SKOS. Cela concerne essentiellement des noms communs et noms géographiques.

### Exemple pour la mythologie slave : la notice Rameau en SKOS<sup>37</sup>

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb11959217b">
<skos:closeMatch rdf:resource="http://id.loc.gov/authorities/sh85089427#concept" />
<skos:closeMatch rdf:resource="http://dewey.info/class/200/" />
<skos:closeMatch rdf:resource="http://dewey.info/class/390/" />
<skos:prefLabel xml:lang="fr">Mythologie slave</skos:prefLabel>
<skos:altLabel xml:lang="fr">Mythologie des Wendes</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie bulgare</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie tchèque</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie slovaque</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie russe</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie slovène</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie biélorusse</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie polonaise</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie caucasienne</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie serbe</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie ukrainienne</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie croate</skos:altLabel>
<skos:altLabel xml:lang="fr">Mythologie sorabe</skos:altLabel>
<skos:narrower rdf:resource="http://data.bnf.fr/ark:/12148/cb121163661" />
<skos:broader rdf:resource="http://data.bnf.fr/ark:/12148/cb119420992" />
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept" />
<skos:inScheme xml:lang="fr">Noms communs (et noms de peuples, de batailles,
etc.)</skos:inScheme>
<skos:related rdf:resource="http://data.bnf.fr/ark:/12148/cb12482657w" />
<skos:related rdf:resource="http://data.bnf.fr/ark:/12148/cb12116367c" />
<skos:related rdf:resource="http://data.bnf.fr/ark:/12148/cb11967423j" />
</rdf:description>
```

En outre, dans data.bnf.fr, SKOS est utilisé pour l'expression de toutes les données d'autorité, en tant que concept, c'est-à-dire la représentation de l'auteur ou de l'œuvre dans un contexte donné. Par exemple un auteur est décrit, comme concept, en SKOS.

<sup>36</sup> <http://www.w3.org/2004/02/skos/> (consulté le 2012-05-15)

<sup>37</sup> Pour la correspondance entre les zones Interمارc et le RDF, voir le mapping donné en annexe.



### Exemple pour Alexandre Dumas : le contenu de la notice d'autorité personne en SKOS

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb119010630">
  <skos:prefLabel>Alexandre Dumas (1802-1870)</skos:prefLabel>
<skos:editorialNote>BN Cat. gén. et BN Cat. gén. 1960-1969 : Dumas (Alexandre),
  père</skos:editorialNote>
<skos:editorialNote>Le comte de Monte-Cristo / Dumas, 1981. - Memorias de José
  Garibaldi / publicadas por Alejandro Dumas, 1860. - La Jeunesse de Pierrot /
  par Aramis [Alexandre Dumas], 1954</skos:editorialNote>
<skos:editorialNote>DBF. - Laffont Bompiani, Auteurs : Dumas père, Alexandre
  (pseud. de Davy de la Pailleterie). - Thieme</skos:editorialNote>
  <skos:altLabel>Alejandro Dumas (1802-1870)</skos:altLabel>
  <skos:altLabel>Alexandre Davy de La Pailleterie (1802-1870)</skos:altLabel>
  <skos:altLabel>Aramis (1802-1870)</skos:altLabel>
  <skos:altLabel>Alexandre Dumas père (1802-1870)</skos:altLabel>
</rdf:Description>
```

### Dublin core metadata terms<sup>38</sup>

Le Dublin Core (DC) est un vocabulaire de métadonnées développé par le DCMI (Dublin Core Metadata Initiative) pour exprimer des descriptions de documents. Il reprend les 15 éléments de base du Dublin Core auxquels s'ajoutent un certain nombre de classes et de propriétés supplémentaires.

Dans le cadre de data.bnf.fr, les classes et les propriétés du DC Terms sont utilisées pour exprimer les données bibliographiques, et les relations entre les entités bibliographiques (œuvres ou expressions) et les entités auteurs et contributeurs. C'est ainsi que le lien entre une œuvre et un auteur sont exprimés par les propriétés **dc:creator** et **dc:contributor**. Les liens entre une expression et un contributeur le sont par la propriété **dc:contributor**.

Les éléments fondamentaux de la description des œuvres sont également exprimés en DC : **dc:date** (date de composition), **dc:language** (langue originale), **dc:subject** (grande classe Dewey à laquelle l'œuvre se rapporte), **dc:date**, **dc:title** et **dc:description**.

Les métadonnées principales des manifestations sont aussi décrites en DC : **dc:title** (pour le titre), **dc:date** (date de publication), **dc:description** (caractéristiques matérielles) et **dc:publisher** (éditeur). Les données proviennent des notices bibliographiques.

### Exemple : description d'une des manifestations de *A la recherche du temps perdu* de Marcel Proust

```
<rdf:Description rdf:about="http://catalogue.bnf.fr/ark:/12148/cb331448270">
  <dc:publisher>1969 [Paris] Gallimard</dc:publisher>
  <dc:date>1969</dc:date>
  <dc:title>À la recherche du temps perdu</dc:title>
  <dc:description>395 p.</dc:description>
</rdf>
```

### FOAF<sup>39</sup>

FOAF est un vocabulaire RDF destiné à la description des personnes et des relations qu'elles entretiennent entre elles. Dans data.bnf.fr, il est utilisé pour exprimer les informations traitant

<sup>38</sup> <http://dublincore.org/documents/dcmi-terms/> (consulté le 2012-05-16).

<sup>39</sup> <http://www.foaf-project.org/> (consulté le 2012-05-15)

des personnes et des collectivités, à partir des données des notices d'autorité noms de personnes et de collectivités.

Les propriétés employées sont **foaf:name** (nom des collectivités et nom complet pour une personne), **foaf:familyName**, **foaf:givenName**, **foaf:birthday**, **foaf:gender**, **foaf:nationality**, **foaf:homepage** (pour les sites personnels). **Foaf:depiction** sert à établir un lien vers une illustration. Dans l'exemple ci-dessous, les vignettes se trouvent dans la bibliothèque numérique de la BnF Gallica et dans Wikimedia.

### Exemple : Jean Racine décrit en FOAF à partir des données d'autorité

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb120076761#foaf:Person">
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k71902t.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k50524c.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k2208884.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k2984335.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k298434j.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k200141f.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/btv1b8407318z.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k1039753.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k50608d.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k207953z.thumbnail" />
  <foaf:depiction
    rdf:resource="http://upload.wikimedia.org/wikipedia/commons/thumb/5/5f/Jean-
    Baptiste_Racine.PNG/200px-Jean-Baptiste_Racine.PNG" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k54291902.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/btv1b8427272z.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/btv1b8407317j.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k5832776x.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/bpt6k6931h.thumbnail" />
  <foaf:depiction rdf:resource="http://gallica.bnf.fr/ark:/12148/btv1b8407319c.thumbnail" />
  <xfoaf:nationality rdf:resource="http://id.loc.gov/vocabulary/countries/fr" />
  <foaf:page rdf:resource="http://data.bnf.fr/12007676/jean_racine/" />
  <foaf:name>Jean Racine</foaf:name>
  <foaf:familyName>Racine</foaf:familyName>
  <foaf:givenName>Jean</foaf:givenName>
  <foaf:gender>male</foaf:gender>
</rdf:Description>
```

### OAI ORE

ORE (pour *Object Reuse and Exchange*<sup>40</sup>) est un vocabulaire développé par la communauté OAI (Open Archive Initiative) pour décrire des agrégations d'objets complexes. Il permet de représenter des relations entre différentes entités, d'exprimer leur structure et d'associer les métadonnées au bon niveau de description.

Nous utilisons ce standard dans data.bnf.fr pour décrire les manifestations complexes telles que la hiérarchie de l'EAD qui décrivent des fonds d'archives ou les notices bibliographiques imbriquées à plusieurs niveaux.

<sup>40</sup><http://www.openarchives.org/ore/>

On emploie pour cela les propriétés **ore:aggregates** et **ore:isAggregatedBy**

**Exemple : pour exprimer que « A la recherche du temps perdu » de Marcel Proust contient *Du côté de chez Swann, A l'ombre des jeunes filles en fleurs, Le côté de Guermantes, Sodome et Gomorrhe, La prisonnière, Albertine disparue et Le temps retrouvé***

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb11940457r#frbr:Work">
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb119662421#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb12574850g#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb13604863p#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb15539123f#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb12010842s#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb15000146f#frbr:Work" />
  <ore:aggregates rdf:resource="http://data.bnf.fr/ark:/12148/cb14518899t#frbr:Work" />
  [...]
</rdf:Description>
```

Dans l'exemple ci-dessous, l'URI de la propriété **rdf:Description** est celle de *A la recherche du temps perdu*. Les propriétés **ore:aggregates** sont les URI des différentes œuvres composant de *A la recherche du temps perdu* et renvoient aux différentes pages œuvres qui leur sont consacrées.

#### iv- Les entités des groupes 1 et 2 décrites avec l'ontologie RDA

Dans [data.bnf.fr](http://data.bnf.fr), l'ontologie RDA est utilisée pour décrire les entités bibliographiques (œuvre, expression, manifestation, item) ainsi que les relations entre elles et les relations avec les entités du groupe 2<sup>41</sup>.

Les propriétés utilisées pour les entités FRBR du groupe 1 et du groupe 2 complètent ainsi respectivement celles de DC et de FOAF.

La description des manifestations est complétée par les propriétés : **rdagroup1elements:publishersName**, **rdagroup1elements:designationOfEdition** et **rdagroup1elements:placeOfPublication**.

**Exemple complet de manifestation à partir des données de la notice bibliographique : les *Fables de Phèdre***

```
<rdf:Description rdf:about="http://catalogue.bnf.fr/ark:/12148/cb31099668t">
  <dc:description>Gr. in-8° , XLVI-188 p.</dc:description>
  <dc:title>Les Fables de Phèdre, édition paléographique, publiée, d'après le manuscrit
  Rosanbo, par Ulysse Robert,... - Appendice : de monstres, belluis et serpentibus
  liber</dc:title>
  <rdagroup1elements:placeOfPublication>Paris</rdagroup1elements:placeOfPublication>
  <rdf:type rdf:resource="http://rdvocab.info/uri/schema/FRBRentitiesRDA/Manifestation" />
  <rdagroup1elements:designationOfEdition />
  <dc:publisher>1893 Paris Impr. nationale</dc:publisher>
  <dc:date>1893</dc:date>
  <rdarels:workManifested
    rdf:resource="http://data.bnf.fr/ark:/12148/cb13320754c#frbr:Work" />
  <rdagroup1elements:publishersName>Impr. nationale</rdagroup1elements:publishersName>
  <rdfs:seeAlso rdf:resource="http://catalogue.bnf.fr/ark:/12148/cb31099668t" />
</rdf:Description>
```

<sup>41</sup> <http://metadataregistry.org/rdabrowse.htm> (consulté le 2012-05-15).

La description de l'entité FRBR du groupe 2 collectivité est précisée par les propriétés :  
**rdagroup2elements:periodOfActivityOfTheCorporateBody,**  
**rdagroup2elements:dataAssociatedWithTheCorporateBody,**  
**rdagroup2elements:fileOfActivityOfTheCorporateBody** et  
**rdagroup2elements:corporateHistory.**

**Exemple complet de collectivité à partir des données d'autorité de la collectivité :  
l'abbaye de Corbie**

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb12191770t#foaf:Organization">
  <dc:date>657</dc:date>
  <dc:title xml:lang="fr">Abbaye de Corbie</dc:title>
  <rdagroup2elements:dateAssociatedWithTheCorporateBody>657</rdagroup2elements:dateAssociated
    WithTheCorporateBody>
  <rdagroup2elements:dateAssociatedWithTheCorporateBody>08-
    1790</rdagroup2elements:dateAssociatedWithTheCorporateBody>
  <xfoaf:nationality rdf:resource="http://id.loc.gov/vocabulary/countries/fr" />
  <foaf:page>http://data.bnf.fr/12191770/abbaye_de_corbie/</foaf:page>
  <rdagroup2elements:corporateHistory xml:lang="fr">Fortement diminuée par la guerre de Cent
    Ans et les guerres du XVIe siècle, elle connut un renouveau religieux et intellectuel après
    son adhésion à Saint-Maur en 1619. - Fermée fin août 1790 et détruite. Il reste une
    partie de l'abbatiale Saint-Pierre et de l'église Saint-
    Etienne</rdagroup2elements:corporateHistory>
  <rdagroup2elements:corporateHistory>Fortement diminuée par la guerre de Cent Ans et les
    guerres du XVIe siècle, elle connut un renouveau religieux et intellectuel après son
    adhésion à Saint-Maur en 1619. - Fermée fin août 1790 et détruite. Il reste une partie de
    l'abbatiale Saint-Pierre et de l'église Saint-Etienne</rdagroup2elements:corporateHistory>
  <rdagroup2elements:corporateHistory>Abbaye fondée en 657 (ou 662) par sainte Bathilde pour
    des Bénédictins venus de Luxeuil. Elle joua un grand rôle dès l'époque carolingienne. Dès
    le IXe siècle elle disposait d'un atelier de copistes renommé qui diffusa l'écriture dite "de
    Corbie". Au Xe siècle elle put battre monnaie et devint d'ailleurs une puissance féodale
    libérée de toute tutelle dont le domaine s'étendit sur 200
    paroisses</rdagroup2elements:corporateHistory>
  <rdagroup2elements:corporateHistory xml:lang="fr">Abbaye fondée en 657 (ou 662) par sainte
    Bathilde pour des Bénédictins venus de Luxeuil. Elle joua un grand rôle dès l'époque
    carolingienne. Dès le IXe siècle elle disposait d'un atelier de copistes renommé qui
    diffusa l'écriture dite "de Corbie". Au Xe siècle elle put battre monnaie et devint
    d'ailleurs une puissance féodale libérée de toute tutelle dont le domaine s'étendit sur
    200 paroisses</rdagroup2elements:corporateHistory>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization" />
</rdf:Description>
```

La description de l'entité du groupe 2 personne est précisée par les propriétés :  
**rdagroup2elements:dateOfBirth,** **rdagroup2elements:dateOfDeath,**  
**rdagroup2elements:placeOfBirth,** **rdagroup2elements:placeOfDeath,**  
**rdagroup2elements:fieldOfActivityOfThePerson** et  
**rdagroup2elements:biographicalInformation.**

## Exemple complet d'une personne à partir des données d'autorité de la collectivité : Phèdre

```
<rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb11985800x#foaf:Person">
  <rdagroup2elements:dateOfDeath>54</rdagroup2elements:dateOfDeath>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person" />
  <xfoaf:nationality rdf:resource="http://id.loc.gov/vocabulary/countries/xx" />
  <dc:title xml:lang="fr">Phèdre</dc:title>
  <rdagroup2elements:dateOfBirth>- 15</rdagroup2elements:dateOfBirth>
  <dc:date>0015? av. J.-C.-0054</dc:date>
  <rdagroup2elements:biographicalInformation>Fabuliste, auteur du premier recueil de fables de
    la littérature latine. - Affranchi d'Auguste</rdagroup2elements:biographicalInformation>
  <foaf:gender>male</foaf:gender>
  <foaf:page rdf:resource="http://data.bnf.fr/11985800/phedre/" />
  <foaf:familyName>Phèdre</foaf:familyName>
  <rdagroup2elements:languageOfThePerson rdf:resource="http://id.loc.gov/vocabulary/iso639-
    2/lat" />
</rdf:Description>
```

### v- Ontologie propre à la BnF

Pour certains éléments, nous avons recouru à des propriétés propres à la BnF. Cela concerne en mai 2012 :

- le marquage des ouvrages propres à la jeunesse<sup>42</sup>
- les cotes utilisées pour identifier un document d'archive ou un manuscrit<sup>43</sup>
- le numéro d'identification EAN<sup>44</sup>
- l'identifiant ISBN<sup>45</sup>
- l'identifiant ISMN pour la musique notée<sup>46</sup>
- l'identifiant ISSN<sup>47</sup>
- les codes de rôle correspondant à une responsabilité intellectuelle sur un document<sup>48</sup>
- les rôles correspondant à une responsabilité intellectuelle<sup>49</sup>
- l'URI d'une exposition virtuelle<sup>50</sup>
- la vignette d'une image choisie manuellement par un opérateur<sup>51</sup>
- du texte alternatif pour des images<sup>52</sup>

Ces propriétés particulières s'inscrivent cependant dans des ontologies communes : DCMI<sup>53</sup>, DC<sup>54</sup> et FOAF<sup>55</sup>. Ce sont en fait des sous-propriétés de propriétés existantes dans ces ontologies communes. De plus, elles sont exprimés en RDFS et OWL. Ainsi, l'emploi limité

<sup>42</sup> URI : <http://data.bnf.fr/ontologie/ouvrageJeunesse>

<sup>43</sup> URI : <http://data.bnf.fr/ontologie/cote>

<sup>44</sup> URI : <http://data.bnf.fr/ontologie/ean>

<sup>45</sup> URI : <http://data.bnf.fr/ontologie/isbn>

<sup>46</sup> URI : <http://data.bnf.fr/ontologie/ismn>

<sup>47</sup> URI : <http://data.bnf.fr/ontologie/issn>

<sup>48</sup> URI : [http://data.bnf.fr/ontologie/code\\_role](http://data.bnf.fr/ontologie/code_role)

<sup>49</sup> URI : <http://data.bnf.fr/ontologie/role>

<sup>50</sup> URI : <http://data.bnf.fr/ontologie/expositionVirtuelle>

<sup>51</sup> URI : <http://data.bnf.fr/ontologie/depiction>

<sup>52</sup> URI : <http://data.bnf.fr/ontologie/imgAlt>

<sup>53</sup> bnf-onto:ouvrageJeunesse sous-propriété de dcmitype:text (<http://purl.org/dc/dcmitype/Text>)

<sup>54</sup> bnf-onto:cote, bnf-onto:ean, bnf-onto:isbn, bnf-onto:ismn et bnf-onto:issn sous-propriétés de dc:identifier ; bnf-onto:role sous-propriété de dc:contributor ;

<sup>55</sup> bnf-onto:code\_role sous-propriété de foaf:agent ; bnf-onto:expositionVirtuelle sous-propriété de foaf:document ; bnf-onto:depiction et bnf-onto:imgAlt sous-propriétés de foaf:depiction

de propriétés propres ne gêne pas la réutilisation libre des données. Les réutilisateurs peuvent reprendre les propriétés des ontologies communes si un niveau de granularité moins fin leur suffit.

### **c- Pour être utilisable par les moteurs de recherche et les navigateurs : utiliser la structuration des documents Web**

Les pages du Web de documents, exprimées en HTML ou XHTML, contiennent des informations qui ne sont structurées qu'en vue de leur présentation, ce qui limite la possibilité de les exploiter et les interpréter automatiquement par les moteurs de recherche. Afin d'ajouter des données structurées intelligibles aux machines, qui seront interprétées par les navigateurs et d'autres programmes, différents formats de données « embarquées » dans le code HTML peuvent être utilisés : microformats, microdonnées, RDFa.

Certains grands acteurs du Web ont défini des applications particulières de ces formats qui ont été implémentés en priorité dans data.bnf.fr afin de s'inscrire dans cet écosystème : c'est le cas de Schema.org et de Facebook Open Graph Protocol.

Schema.org<sup>56</sup>, est une initiative des moteurs de recherche Google, Bing, Yahoo et Yandex permettant d'ajouter de l'information simple, sous forme de microdonnées, dans le contenu des pages HTML, pour en améliorer le référencement. Schema.org est utilisé dans data.bnf.fr pour décrire simplement, mais de façon structurée dans le code HTML, certaines entités que sont les personnes, les organisations et les livres, correspondant aux « types » proposés par schema.org<sup>57</sup>.

De même, le réseau social Facebook a mis en place le standard Open Graph Protocol<sup>58</sup>, un vocabulaire RDF très simple pour encoder en RDFa quelques métadonnées très simples qui seront récupérées lorsque l'utilisateur ajoutera la ressource à son profil Facebook. L'Open Graph Protocol est lui aussi implémenté dans les pages data.bnf.fr.

L'emploi de ces différents formats rend possible différents types d'utilisations des données de la BnF.

### **d- Data.bnf.fr comme interface entre l'interne et l'externe**

Data.bnf.fr se conçoit comme une interface active : une porte d'entrée vers les ressources de la BnF et une correspondance vers des référentiels externes.

Data.bnf.fr donne accès aux ressources de la BnF, par des liens vers les catalogues, vers la bibliothèque numérique Gallica et vers les expositions virtuelles en utilisant les mécanismes suivants :

---

<sup>56</sup> <http://schema.org/> (consulté le 2012-05-15)

<sup>57</sup> <http://schema.org/Person> , <http://schema.org/Book> , <http://schema.org/Organization> (consultés le 2012-05-15).

<sup>58</sup> Voir <http://opengraphprotocol.org/>



Entité data.bnf.fr	Propriété utilisée	Cible
concept	rdfs:seeAlso	Notice d'autorité personne, collectivité, titre ou Rameau correspondante au concept
manifestation	rdfs:seeAlso	BnF-Catalogue général
manifestation	rdfs:seeAlso	BnF-Archives et manuscrits
manifestation	relationships:electronicReproduction	Reproduction numérique dans Gallica
oeuvre	foaf:depiction	Illustration de la page oeuvre issue de Gallica
auteur	foaf:depiction	Illustration de la page auteur issue de Gallica

En outre, data.bnf.fr se documente par l'utilisation de référentiels existants autant que cela est possible. C'est actuellement le cas pour les ressources cibles suivantes :

Entité data.bnf.fr	Propriété utilisée	Cible
concept	owl:sameAs	Stitch Rameau <sup>59</sup>
concept	skos:exactMatch	Thesaurus W
concept	skos:closeMatch	LCSH
concept		DNB
concept		Dewey-info
concept		
oeuvre	dc:language	id.loc.gov languages
oeuvre	dc:subject	Dewey.info
expression	dc:type	dcmi: type
expression	dc:language	id.loc.gov languages
auteur	owl:sameAs	VIAF
auteur	rdagroup2elements:fieldOfActivityOfThePerson	Dewey.info
auteur	xfoaf:nationality	id.loc.gov countries
auteur	rdagroup2elements:languageOfThePerson	id.loc.gov languages

Les références à des sources extérieures permettent de consolider les données en dépassant l'utilisation de référentiels internes dans les notices BnF.

Ces liens permettent de faire entrer data.bnf.fr dans le **nuage international du Web sémantique** et de mettre les ressources de la BnF en communication avec d'autres ressources. Par exemple, VIAF est utilisé comme rebond vers DBPedia. Comme data.bnf.fr n'est pas un produit figé, d'autres référentiels internationaux peuvent à l'avenir être utilisés. C'est le cas de l'identifiant international pour les identités personnes ISNI par exemple<sup>60</sup>.

En outre, la place de data.bnf.fr et, plus généralement, des données des bibliothèques dans le nuage international du Web sémantique n'est pas statique. Pour le moment, DBPedia sert de nœud central à ce nuage. Les données des bibliothèques présentes dans le nuage (VIAF, LC, data.bnf.fr) ont tous les atouts pour servir également de pivot et ne pas laisser ce privilège à d'autres.

<sup>59</sup> Projet expérimental mené par TELplus pour exprimer Rameau en SKOS (<http://www.cs.vu.nl/STITCH/rameau/>).

<sup>60</sup> <http://www.isni.org/> (consulté le 2012-05-15).

### 3) *Data.bnf.fr* et ses multiples utilisations possibles

La modélisation des données en FRBR, l'utilisation des technologies du Web sémantique alliée à celle des technologies du Web de documents et l'application d'une licence de données ouvertes donnent des possibilités illimitées d'utilisation et de réutilisation, que ce soit comme mise en valeur et moyen d'accès aux ressources des bibliothèques ou comme moyen de réutiliser les données de bibliothèques bien au-delà du monde des bibliothèques.

Les données structurées des notices d'autorité et des notices bibliographiques en sont le ressort. Grâce à leur utilisation, *Data.bnf.fr* donne un accès direct des moteurs de recherche aux ressources de la BnF, en premier lieu aux ressources numériques. De plus, la bonne structuration des données des catalogues permet de donner de l'information précise et exacte sur ce que l'internaute va trouver effectivement.

Prenons l'exemple de Gégroire de Tours<sup>61</sup>, historien des premiers rois mérovingiens. Un amateur ou un lycéen cherchant la forme « Grégoire de Tours Histoire des Francs » dans un moteur de recherche trouvera l'accès à des ressources numériques de Gallica sur la première page de résultats. Un spécialiste d'historiographie mérovingienne utilisant la forme latine « Gregorius Turonensis » y trouvera également des ressources de la BnF.

On voit donc que ce n'est donc plus les usagers qui viennent à la bibliothèque mais la bibliothèque qui se met sur la route des usagers, sans préjuger de leurs usages. L'objectif est que ceux-ci utilisent même sans le savoir les données structurées des bibliothèques pour leur recherche.

Plus largement, les données de *data.bnf.fr* sont librement réutilisables à la seule condition d'indiquer la source des données. Le dump RDF est librement accessible. Des programmeurs peuvent ainsi développer de petites applications utilisant les données de la BnF. *Data.bnf.fr* offre une vue JSON<sup>62</sup>, pour faciliter le développement d'applications simples utilisant un ensemble de données choisies et simplifiées<sup>63</sup>.

## Conclusion

*Data.bnf.fr* n'a pas été conçu comme une application destinée à remplacer les catalogues mais à donner à ces derniers un large accès, même pour les usagers qui n'ont pas *a priori* l'intention de les utiliser et encore moins d'utiliser les métadonnées descriptives ou les données d'autorité pour quelque usage que ce soit. Les données structurées à la source, selon les normes et les pratiques professionnelles, vérifiées et contrôlées sont un atout précieux. C'est sur elle qu'est fondé l'ensemble du mécanisme de *data.bnf.fr*, qui permet aux ressources de la BnF de figurer à un haut niveau dans les pages de résultats des moteurs de recherche. Le processus d'alignement utilisé, le rendu du résultat en RDF, en JSON ou en HTML ouvre de larges possibilités de réutilisation, et donne libre cours à des usages que les bibliothèques ne soupçonnent guère.

---

<sup>61</sup> <http://catalogue.bnf.fr/ark:/12148/cb11905825m/>

<sup>62</sup> <http://www.json.org/>. Ce type de vue est aussi employé dans d'autres sites tels qu'Amazon ou Openlibrary (site collaboratif autour du livre).

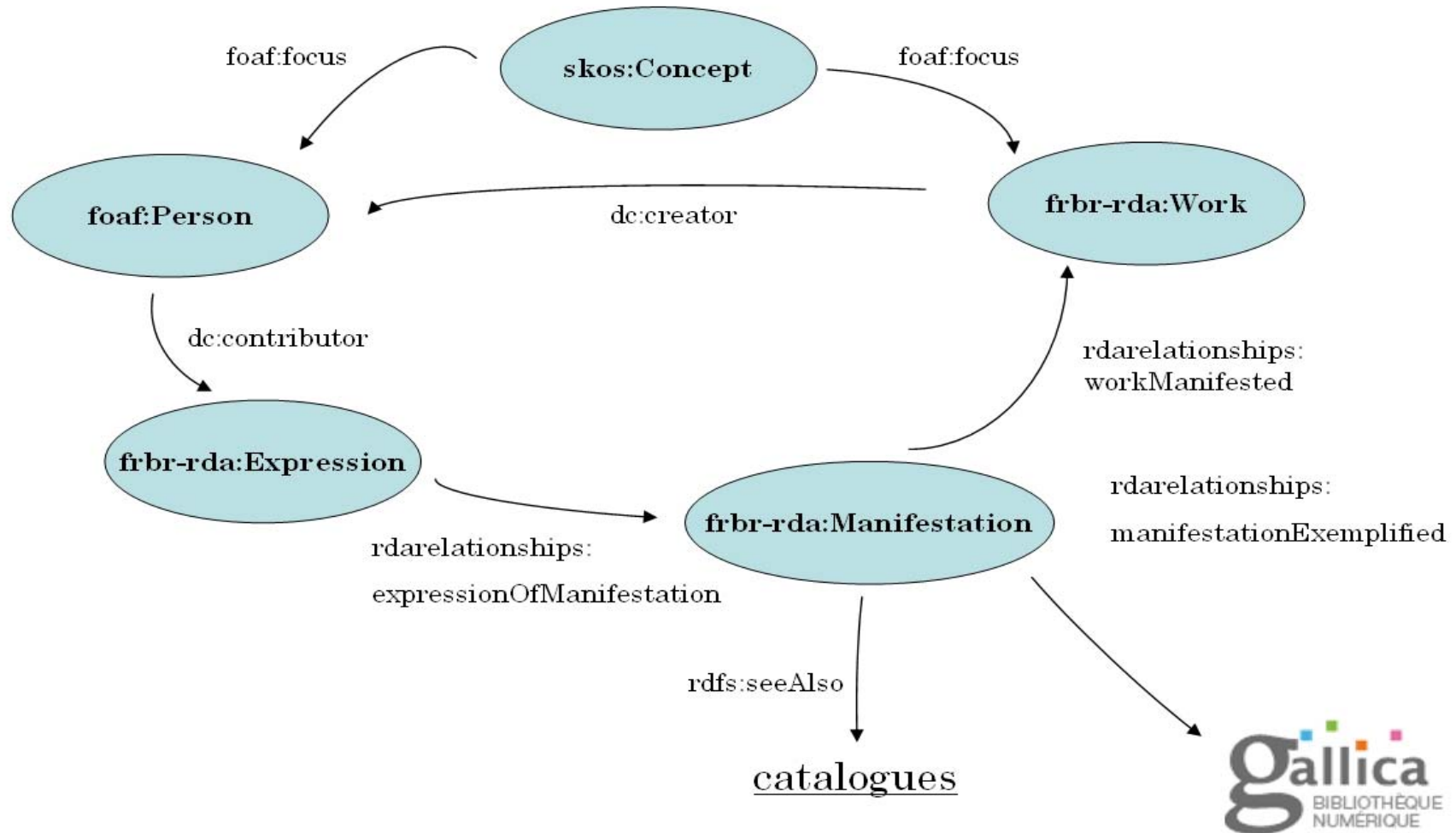
<sup>63</sup> La vue JSON présente ainsi des informations succinctes sur les œuvres et auteurs présentés sur les pages *data.bnf.fr* (nom, titre, etc.), mais aussi les vignettes issues de *Gallica* et *Wikipedia*, les liens vers les documents numérisés, la liste des identifiants ISBN des éditions d'une œuvre. Ces données ne sont donc pas aussi exhaustives que les données en RDF, mais elles sont plus faciles à manipuler. Il est par exemple possible de récupérer les ISBN des monographies par ce biais.

Par conséquent, ce sont les règles de catalogage et les données structurées elles-mêmes qui sont le moyen de sortir les données des catalogues, de les mettre sur la route des utilisateurs et de les mettre à disposition d'usages multiples, liées à la documentation et aux missions traditionnelles des bibliothèques ou non. Les missions traditionnelles des bibliothèques et les règles employées pour les appliquer ne sont pas obsolètes en soi. Les technologies du Web permettent de réutiliser les données qu'elles produisent bien au-delà de leur objectif premier sans les dénaturer.

**Pour en savoir plus :**

Consulter les pages « A propos » du site data.bnf.fr : <http://data.bnf.fr/about>

**Schéma simplifié de l'ontologie utilisée**



## Mapping entre le format InterMarc et le RDF fourni

Personnes	Correspondance RDF	Zone intermarc (notices PEP)
forme	skos:prefLabel @in_lang	100 400
autre forme	skos:altLabel foaf:familyName foaf:givenName dc:date	
nationalité	foaf:nationality	008 position 12-13
langue	RDAGroup2elements:languageOfThePerson	008 position 14 16
sexe	foaf:gender	008 position 17
naissance	RDAGroup2elements:dateOfBirth	008 position 27-36
mort	RDAGroup2elements:dateOfDeath	008 position 37-46
lieu de naissance	RDAGroup2elements:placeOfBirth	603 \$a
lieu de mort	RDAGroup2elements:placeOfDeath	603 \$b
début d'activité	RDAGroup2elements:periodOfActivityOfThePerson	008 position 47 à 51
fin d'activité	RDAGroup2elements:periodOfActivityOfThePerson	008 position 52-55
sources (note sur les sources de la notice)	skos:editorialNote	610
résumé, note	RDAGroup2elements:biographicalInformation	600
domaines	RDAGroup2elements:fieldOfActivityOfThePerson	624
liens vers la ressource Dbpedia	Owl:sameAs	
fonction de contributeur	marcrel:[code de fontion de la bibliothèque du Congrès, <a href="http://id.loc.gov">http://id.loc.gov</a> ]	
vignette Gallica représentant l'auteur	foaf:depiction	
Collectivités	Correspondance RDF	Zone Intermarc (notices ORG)
forme	skos:prefLabel @in_lang	100 400
nationalité	foaf:nationality	008 position 12-13
langue	RDAGroup2elements:languageOfThePerson	008 position 14 16
date de début	RDAGroup2Elements:dateAssociatedWithTheCorporateBody	008 pos 27 à 36
date de fin	RDAGroup2Elements:dateAssociatedWithTheCorporateBody	008 pos 37 à 46
début activité	dc:date	008 pos 47 à 51

fin activité	RDAGroup2elements:periodOfActivityOfTheCorporateBody	008 pos 52 à 55
site internet	foaf:homepage	606
sources	skos:editorialNote	610
résumé/note	RDAGroup2elements:corporateHistory	600
domaine	RDAGroup2elements:fieldOfActivityOfTheCorporateBody	624
liens vers la ressource Dbpedia	owl:sameAs	
<b>Sujets RAMEAU</b>		
<b>Correspondance RDF</b>		
titre original	skos:prefLabel	16X 46X
autre forme	skos:altLabel	16X 46X
origine (thésaurus Rameau)	skos:inScheme	
source (note sur les sources de la notice)	skos:editorialNote	610-612
note (note descriptive)	skos:scopeNote	600
termes plus larges	Skos:broader	3XX, 5XX
termes plus précis	skos:narrower	3XX, 5XX
termes reliés	skos:related	3XX, 5XX
alignement thésaurus extérieurs	skos:closematch	620
alignement thésaurus extérieurs	skos:exactmatch	
<b>Œuvre</b>		
<b>Correspondance RDF</b>		
<b>Zone InterMarc (notices TIC, TUT, TUM)</b>		
forme (titre principal)	dc:title skos:prefLabel, rdfs:label @in_lang	145 415
autre forme	skos:altLabel @in_lang	
langue	dc:language	008 position 14 16
date oeuvre	dc:date	008 position 27 à 36
source	skos:editorialNote	610
résumé/note	dc:description	600
domaine	dc:subject	624
lien vers la notice d'autorité du catalogue	owl:sameAs	
Ccntenu dans	dc:isPartOf	



Relations		
auteur principal	dc:creator	100 101 110 110
auteur secondaire	dc:contributor bnf_onto:coderole	711/702/700/701/710/712
auteur code libre	dc:contributor bnf_onto:coderole	code libre 321 322
vignette Gallica de l'œuvre numérisée	foaf:depiction	
<b>Manifestation</b>	<b>Correspondance RDF</b>	<b>Zone Intermarc (notice bibliographiques)</b>
manifestation de l'œuvre	rdarelations:workManifested	
titre	dc:title	245
contient	dc:hasPart	
édition (date d'édition)	dc:date	260
édition (lieu d'édition)	rdvocab:placeOfPublication	250
édition (nom de l'éditeur)	rdvocab:publishersName	260
description matérielle	dc:description	
ISBN	bnf-onto:ISBN	20
type de document	dc:type	
langue	dc:language	41
adaptation jeunesse	bnf-onto:ouvrageJeunesse	
<b>Expression</b>	<b>Correspondance RDF</b>	
rôle de contribution	marcrel: [code de fonction de la Bibliothèque du Congrès]	
rôle de contribution (chiffre)	bnf-onto:coderole	sous-zone \$4
Rôle de contribution (valeur littérale)	bnf-onto:role	
type de document	dc:type	

### Auteur

**Vincent Boulet** est archiviste paléographe et docteur en histoire. Conservateur de bibliothèque à la BnF, il est expert en données d'autorités et coordonne l'application de l'EAD à la BnF. Il est membre du comité permanent de la section de bibliographie.