

# **Rapport d'orientation pour le Comité stratégique bibliographique**

**préparé par  
l'Agence bibliographique de l'enseignement  
supérieur et  
la Bibliothèque nationale de France**

**Rapport d'orientation**  
**pour le Comité stratégique bibliographique**  
  
**préparé par**  
**l'Agence bibliographique de l'enseignement**  
**supérieur et la Bibliothèque nationale de France**

## **Plan**

1- Préambule : une vision pour le futur des catalogues .....	page 3
2- Etat des lieux : les expérimentations en cours .....	page 7
3- Adoption de RDA en France : les orientations .....	page 10
4- La question des licences .....	page 12
5- L'enjeu du signalement de la documentation électronique .....	page 14
6- Formation, communication, accompagnement au changement.....	page 16
7- L'évolution des SIGB et la structuration des métadonnées publiques nationales : quelles perspectives ? .....	page 18
8- Les études à conduire : deux propositions .....	page 19
Glossaire .....	page 21
Compte-rendu de la réunion du Comité stratégique bibliographique du 13/11/2011.....	page 25

# 1- Préambule :

## une vision pour le futur des catalogues

### Préambule

La BnF et l'ABES partagent plusieurs constats et préconisations préalables relatifs à l'évolution des usages, des collections, des technologies et des organisations dans le domaine de l'information bibliographique et numérique. Elles s'engagent dans une démarche commune de redéfinition à dix ans des modalités de structuration, de production et de diffusion des métadonnées de bibliothèque à l'échelon de la France. A ce stade de leur réflexion, il leur est encore difficile de proposer un calendrier ou un plan d'action concerté pour mettre en oeuvre les évolutions envisagées. Ce document vise avant tout à exposer et partager les principes et les motivations du changement et les premières réalisations et expérimentations qui permettent d'en comprendre et d'en mesurer la portée.

### 1) Usages : valoriser les données bibliographiques selon les usages du Web

Il faut désormais considérer que le principal mode d'accès à l'information est le Web. Si l'on doit se garder d'ignorer les pratiques ancrées dans les modalités traditionnelles de recherche bibliographique (nécessité de maintenir des catalogues classiques là où ils sont utilisés), une vision prospective à dix ans impose de se projeter dans un horizon résolument numérique. L'utilisateur, qu'il fréquente ou non le lieu de la bibliothèque, utilise ou utilisera au quotidien le Web (depuis des terminaux fixes ou mobiles) pour identifier des ressources documentaires et les exploiter. Cette nouvelle donne implique de reformuler l'offre bibliographique publique selon les usages et l'environnement des internautes plutôt qu'en fonction des contraintes et exigences de production des bibliothécaires, encore très prégnantes dans la conception des catalogues.

- **Simplicité.** Les modalités de recherche devront être plus simples, plus intuitives et ne plus impliquer que l'utilisateur maîtrise les référentiels des bibliothèques ni leurs logiques de gestion pour trouver ce qu'il cherche. Les ressources des bibliothèques devront favoriser la sérendipité, c'est-à-dire pouvoir être proposées et découvertes et pas seulement trouvées. Les références aux différentes manifestations d'une même œuvre devront être signalées de manière fédérée et transparente indépendamment des modalités de leur traitement et des silos de production dont elles sont issues. Les mécaniques nécessaires aux traitements bibliographiques devront, par défaut, rester invisibles pour l'utilisateur qui ne doit bénéficier que de leur résultat, tout en restant intelligibles pour des recherches plus approfondies.
- **Référencement par les moteurs.** Les données bibliographiques devront être référencées et accessibles en bonne place dans le classement des résultats, directement depuis les moteurs de recherche. Les interfaces des catalogues sont en effet aujourd'hui accessibles en ligne mais leurs données restent le plus souvent invisibles pour les moteurs car elles sont enfouies dans des bases de données fermées.
- **Des granularités variables.** Les futurs systèmes devront donner accès non seulement à des données produites dans des formats et des systèmes différents mais également à des niveaux de signalement différents. La recherche requiert un traitement à un niveau de granularité toujours plus fin, en particulier pour le signalement des ressources continues au niveau de l'article, mais aussi pour l'exploration des livres numériques au niveau des sommaires, des chapitres, voire des pages, des mots. Pour autant, toutes les ressources ne se prêtent pas à des traitements (dépouillements, indexation...) aussi fins. Il sera tout autant nécessaire de se repérer et de naviguer dans de très grandes masses de données brutes, fonds, recueils ou corpus numériques. Entre l'infiniment petit et l'infiniment grand, se dégage un besoin global de modularité : les processus de signalement devront être capables de combiner des recherches à partir de métadonnées structurées issues d'une analyse humaine et des processus d'indexation automatique massifs.

- **Accès direct au document.** En lien avec les politiques de numérisation et d'acquisition de ressources électroniques conduites au niveau national, les références bibliographiques devront être directement liées aux documents primaires lorsque ceux-ci sont techniquement et juridiquement accessibles sous une forme numérique. L'internaute attend généralement un accès direct au document, qui peut le conduire à délaisser une référence sans lien au texte intégral. Il ne faut pas pour autant minorer la valeur du signalement de ressources non accessibles en ligne (soit qu'elles n'existent pas sous forme numérique, soit que leur consultation fasse l'objet de restrictions juridiques). Ce service demeure en effet essentiel pour identifier et localiser les collections.
- **Des données liées.** Dans la logique hypertextuelle du média Internet, les données bibliographiques devront être fortement liées : entre elles (afin d'exploiter toute la richesse de la structuration interne des bases), mais aussi avec des ressources externes (ressources institutionnelles ou commerciales, culturelles, scientifiques, géographiques...).
- **Interaction et réutilisation.** Au-delà des services de base attendus (recherche, lecture, téléchargement...), les bibliothèques devront faciliter de nouveaux services à partir de leurs métadonnées, services qui sont devenus courants sur le Web (et le Web 2.0 en particulier) : les métadonnées devront pouvoir être citées, partagées, enrichies de contenus produits par les usagers eux-mêmes (annotations...) ; des traitements automatisés d'exploitation des données (data mining, calculs bibliométriques...) devront également être proposés.

La nécessaire insertion des données des bibliothèques dans l'écosystème de l'internaute conduira ainsi à considérer aussi bien des utilisateurs humains que des machines chargées de retraiter les métadonnées des bibliothèques afin de les agréger à d'autres ressources. L'ABES et la BnF souhaitent ainsi inscrire leur trajectoire dans le mouvement d'ouverture des données publiques.

## 2) Collections : des documents aux données, élargissement de l'offre documentaire

Le développement des collections numériques (édition nativement numérique, produits de la numérisation, captations massives de données brutes issues de la recherche ou directement de l'Internet...) élargit le champ de l'offre documentaire susceptible d'être acquise et valorisée par les bibliothèques. Cette transformation majeure impacte la problématique du signalement d'au moins deux façons :

- **Des collections plus volumineuses et durablement hybrides :** la quantité globale de ressources à signaler croît de manière exponentielle, rendant économiquement irréaliste un signalement unitaire par les professionnels. Le signalement conjoint de la documentation analogique (jusqu'ici maîtrisé, mais dont les volumes ne baissent pas comme le montrent les statistiques du dépôt légal) et d'une documentation électronique en expansion (essentielle pour l'IST mais aussi dans le champ patrimonial alors que se construit un patrimoine né numérique) constitue un objectif crucial compte tenu des attentes des usagers mais aussi des investissements publics consentis pour produire, conserver ou acquérir des collections numériques.
- **Des collections dont la structure évolue :** en lien avec l'évolution des usages et des technologies, les collections documentaires ne sont plus seulement des regroupements de documents, ils s'apparentent plus largement à des jeux de données (datasets). Ce nouveau paradigme, déjà investi sur le Web par d'autres secteurs que les bibliothèques, repose sur une atomisation du document au profit de la multiplication des liens et des possibilités d'agrégation entre éléments de données. De la même façon, la notice bibliographique traditionnelle peut être décomposée en éléments et en relations grâce auxquels s'ouvrent des possibilités nouvelles d'enrichissement de

l'expérience de recherche documentaire : c'est le principe du Web de données ou Web sémantique, qui structure toute la réflexion prospective engagée par l'ABES et la BnF.

### **3) Technologies : Adopter les standards du Web**

L'ABES et la BnF considèrent que le Web de données est la destination à privilégier pour les métadonnées des bibliothèques. Ce nouveau sous-ensemble du Web, dont la création est portée et normalisée par le consortium W3C, constitue en effet un espace documentaire unifié, ouvert et global, capable de tisser des liens entre données à un niveau très fin de description et de désenclaver les données de leurs silos de production enfouis dans le Web invisible.

Dans cet espace, où les données sont structurées aussi bien pour une lecture par les navigateurs humains que pour un référencement et des traitements par des machines, les métadonnées déjà très liées et très structurées des bibliothèques pourront gagner une audience et une visibilité inégalées. Elles pourront également être liées à d'autres ressources, que celles-ci proviennent d'autres bibliothèques (possibilités de mutualisation du catalogage à une échelle globale) ou d'autres secteurs d'activité (démultiplication des possibilités d'exploitation de ces données publiques) du fait d'une interopérabilité accrue.

Travailler à transformer les données des bibliothèques françaises pour les faire entrer dans le Web de données permettrait de trouver des réponses aux problématiques économiques que soulève l'accroissement exponentiel des volumes à signaler puisque les moyens humains pour les décrire n'augmenteront pas.

Atteindre cette cible implique toutefois des changements conceptuels, normatifs et technologiques majeurs pour le monde des bibliothèques. Alors que les usagers et les collections elles-mêmes fonctionnent de plus en plus dans l'environnement technologique du Web, les bibliothèques s'appuient toujours sur un ensemble de normes, de formats (MARC...) et de protocoles (Z39-50...) qui sont propres à leur communauté professionnelle et méconnues des autres. Cette spécificité constitue un handicap majeur pour leur insertion dans le Web. Le principal défi technique à relever consiste donc à revoir radicalement un dispositif vieux de près de 40 ans sans pour autant perdre la richesse et la cohérence des pratiques de description qui font la valeur ajoutée des données de bibliothèques. Les nécessaires évolutions identifiées à ce jour afin de réussir cette transformation concernent :

- L'application du modèle de données FRBR (Functional Requirements for Bibliographic Records), modèle conceptuel qui a jeté les bases de catalogues innovants ;
- Le code de catalogage : RDA (Resource Description and Access), sous réserve d'adaptations ;
- Le format : RDF (Resource Description Framework), qui permet la gestion de données liées ;
- L'utilisation d'identifiants (URIs) pérennes et univoques ;
- Le développement et la maintenance de référentiels d'autorité riches et structurés.

L'ensemble de ces standards se veulent agnostiques du point de vue des systèmes informatiques ; la collaboration entre ABES et BnF (et d'autres organisations) n'implique donc pas qu'elles utilisent des systèmes identiques. C'est au niveau de la structuration et du statut des métadonnées (formats, identifiants, liens, licences d'utilisation) et directement sur le Web (et non plus dans des applications) que peut s'établir l'interopérabilité et la collaboration.

### **4) Organisation : Parier sur la mutualisation et la coopération**

L'accroissement de l'offre éditoriale est tel qu'on ne peut plus attendre des bibliothèques qu'elles décrivent par elles-mêmes tous les documents qu'elles gèrent. Priorité devra donc être donnée à la coopération avec les producteurs d'information et en particulier les acteurs de l'édition afin de récupérer des métadonnées préexistantes, comme cela se pratique déjà dans d'autres pays comme le Royaume-Uni. Pour l'ABES, l'enjeu est d'abord celui du signalement des périodiques électroniques. Le dispositif des licences nationales constitue à cet égard un important levier de négociation afin d'obtenir des éditeurs la fourniture et le droit de réutilisation de leurs métadonnées. Pour la BnF, l'accent est à placer en priorité sur la production éditoriale nationale à travers le dépôt légal, qui s'étend au numérique depuis la loi

Dadsvi de 2006. Des complémentarités se dessinent de ce point de vue entre les deux établissements. Pour chacun, une démarche de coopération et de négociation est à conduire, avec les représentants des éditeurs, qui pourrait bénéficier d'un appui politique. A un niveau plus technique, il leur faudra travailler à une meilleure interopérabilité des formats de métadonnées et des référentiels utilisés par les éditeurs et distributeurs en production (ONIX...) avec les métadonnées de bibliothèque. Cette collaboration souhaitée avec les éditeurs complétera la démarche de mutualisation déjà bien engagée auprès d'autres bibliothèques étrangères qui exposent leurs données et les partagent, à travers des initiatives dont celles coordonnées par OCLC : WorldCat, VIAF pour les données d'autorité.

Parallèlement, il conviendra de renforcer les logiques de mutualisation et de répartition de la production bibliographique au sein des bibliothèques françaises. Il faut viser à la déduplication des efforts en recherchant une optimisation du rôle des agences. Celles-ci doivent s'affirmer comme les opérateurs de réservoirs de métadonnées de référence dans lesquels chaque établissement viendra puiser au lieu d'en dupliquer la production localement. La mise à disposition de ces métadonnées, coordonnée et mutualisée entre l'ABES et la BnF, et le développement de services en favorisant la réutilisation devraient faire profondément évoluer le travail de catalogage dans les établissements dans le sens d'un allègement ou d'une transformation des charges. Si les données de référence sont mutualisées et rendues librement accessibles, alors le travail de signalement dans les établissements pourra se concentrer sur le traitement des données locales, la description de fonds spécialisés et l'enrichissement des données par la création de liens.

Au-delà d'une volonté commune, on constate toutefois aujourd'hui un contraste important entre le réseau structuré de catalogage partagé de l'enseignement supérieur et de la recherche et tous les autres établissements, aux pratiques hétérogènes et qui ne sont pas structurés selon une logique de production bibliographique concertée. Dans le secteur de la lecture publique, la demande des établissements offre un spectre extrêmement large quant au type et au niveau de description des données attendus, la fraîcheur de l'information bibliographique constituant une priorité pour tous.

Ce double effort de mutualisation (entre bibliothèques et éditeurs, entre établissements publics) sera pensé à un niveau international et pas seulement français. Aujourd'hui, les grandes collections numériques, les catalogues, les réseaux et les technologies s'inscrivent dans un contexte global et irriguent une coopération internationale accrue. L'échelon européen notamment constitue un niveau de collaboration régional qui présente des opportunités politiques, scientifiques et économiques. A travers leur participation aux travaux de l'IFLA, de LIBER et d'autres organisations de ce type, mais aussi de collaborations bilatérales avec des bibliothèques et des réseaux innovants et engagés (British Library, Library of Congress, Deutsche Nationalbibliothek, RERO...), l'ABES et la BnF continueront de participer à l'effort international de veille, de normalisation, d'innovation et de coopération. Il faut en effet que la stratégie française tire pleinement parti des avancées des autres, mais aussi que les initiatives nationales soient valorisées dans ce cadre au bénéfice d'une meilleure visibilité des collections françaises.

L'ABES avec son projet de hub des métadonnées de la recherche, et la BnF avec celui de FRBRisation et d'exposition des métadonnées de référence qu'elle produit pour la collectivité nationale travailleront à rendre interopérables leurs bases afin de satisfaire aux ambitions de mutualisation entre les deux établissements.

## 2- Etat des lieux : les expérimentations en cours

### Etat des lieux et des expérimentations à la BnF

#### **Une stratégie prudente en matière de production bibliographique**

La démarche d'expérimentation engagée par la BnF se veut prudente, pragmatique et phasée au vu des contraintes budgétaires et des incertitudes qui pèsent sur l'évolution rapide des marchés et des pratiques dans le domaine bibliographique et numérique.

Les priorités de développement concernent d'abord la rationalisation et la modernisation des applications les plus fondamentales qui montrent des signes d'essoufflement : refonte de l'interface du catalogue général, évolution du logiciel utilisé pour Archives et manuscrits, intégration du catalogue du Centre national de la Littérature pour la jeunesse. Une collaboration accrue avec les éditeurs est par ailleurs recherchée afin de favoriser la récupération de métadonnées à la source (via l'Extranet du dépôt légal) tandis que la dérivation de notices pour les acquisitions étrangères (via WorldCat) s'accélère. Ce premier axe de travail concerne principalement les fonctions de production bibliographique, qu'il vise à consolider afin de pallier des obsolescences technologiques ou de réaliser de gains de productivité. Il ne remet toutefois pas en cause fondamentalement l'organisation bibliographique actuelle, qui dispose d'atouts certains par rapport aux évolutions envisagées avec RDA et FRBR, du fait, en particulier, de la richesse des liens déjà existants et constamment enrichis entre notices bibliographiques et d'autorité.

#### **Des expérimentations innovantes pour la diffusion des données**

Le second axe concerne la diffusion, l'exposition et la réutilisation des données. C'est dans ce domaine que la BnF se veut particulièrement innovante et teste en situation réelle les possibilités et les effets de l'intégration de son patrimoine bibliographique au Web de données. Elle a lancé en juillet 2011 l'application pilote, *data.bnf.fr*, qui a pour objectifs principaux :

- de regrouper dans des pages sur les auteurs, les œuvres ou les thèmes une sélection de données bibliographiques issues de différentes bases (catalogue général, archives et manuscrits, Gallica...).
- d'exposer ces pages aux moteurs de recherche;
- de structurer les données en RDF et conformément aux standards du W3C, de façon à pouvoir les lier, à un niveau de granularité très fin, à d'autres jeux de données du web sémantique;
- de permettre, techniquement et juridiquement, la réutilisation par des tiers de ces données, qui ont été placées fin 2011 sous licence ouverte de l'Etat et intégrées au portail gouvernemental *data.gouv.fr*.

Un an après le lancement de *data.bnf.fr*, le bilan qu'on peut en établir est convaincant:

- le seuil critique atteint par *data.bnf.fr*, qui au travers de ses 17 000 pages 'auteurs', 44 300 pages 'œuvres', 16 7200 pages 'thèmes', est relié à plus de 2,5 millions de notices du catalogue général, soit environ 20 % de ses notices. Cette montée en charge est assurée de manière entièrement automatique, sans intervention humaine, grâce à la qualité des données produites à la source par les catalogueurs ;
- l'amélioration de l'accès aux ressources due au référencement : d'une part, plus de 80 % des visites dans *data.bnf.fr* proviennent des moteurs de recherche, d'autre part *data.bnf.fr* permet d'optimiser le référencement des ressources des bases auxquelles ses pages sont liées : *data.bnf.fr* contribue ainsi à améliorer l'audience de Gallica comme du catalogue général avec un taux de conversion vers ces applications de plus de 70%;
- l'amélioration de l'identification de données issues de bases hétérogènes, qui résulte de l'intégration de ces données dans un modèle FRBR : *data.bnf.fr* joue un rôle de fédérateur ou de facilitateur pour naviguer depuis le Web entre des bases relativement étanches pour l'utilisateur humain ou machine;

- la reconnaissance de la BnF dans des réseaux d'innovation du web, grâce à la qualité des données et à la montée en compétences de ses équipes sur ce projet.

Après une année d'exploitation, *data.bnf.fr* confirme donc ses promesses et la BnF souhaite dépasser l'expérimentation pour viser une montée en charge dans le cadre d'un marché de développement qui vient d'être renouvelé pour 3 ans.

Il s'agit d'abord d'accroître progressivement les volumes exposés de façon à ce que soient signalées à terme toutes les ressources que possède la BnF. Dans le même temps, la multiplication de liens vers davantage de ressources extérieures, en commençant par IDref et VIAF, permettra d'assurer des logiques de consolidation, de partage et d'appui réciproques. *data.bnf.fr* sera enfin exploité au profit des bases de données de production, pour lesquelles seront réutilisés les algorithmes de regroupement mis au point pour ce projet. La BnF envisage d'expérimenter ainsi l'intégration a posteriori du modèle FRBR dans ses catalogues en attaquant la problématique d'évolution de la production bibliographique selon un mode plus rétroactif que proactif.

La possibilité pour des tiers de se lier aux données de *data.bnf.fr* et de les réutiliser constitue par ailleurs pour la BnF une nouvelle offre de service et un élargissement du champ de ses publics potentiels. Parmi les projets déjà lancés dans le cadre de la réflexion sur l'évolution des produits et services bibliographiques, on peut citer OpenCat, porté par la bibliothèque municipale de Fresnes et soutenu par le ministère de la Culture et de la communication dans le cadre d'un appel à projets sur les services numériques innovants. Ce projet permet d'expérimenter la possibilité pour les bibliothèques publiques de réutiliser les données de *data.bnf.fr* comme socle de leur catalogue. Un prototype sera présenté en mars 2013. D'autres cas de réutilisation hors du monde des bibliothèques sont identifiés et encouragés, que ce soit par des administrations (IFVERSO<sup>1</sup>, portail de traductions multilingues de l'édition française récemment initié par l'Institut français pour le ministère des Affaires étrangères), des communautés de chercheurs<sup>2</sup>, des initiatives internationales promotrices de l'Open Data comme l'Open Knowledge Foundation<sup>3</sup> ou encore des développeurs d'applications grand public pour smart phones<sup>4</sup>.

Au-delà de *data.bnf.fr*, qui doit rester un produit applicatif cohérent et lisible, la BnF envisage l'exploitation dans d'autres cadres des technologies et du savoir-faire capitalisés pour ce projet. Dès 2013, une base de reliures remarquables sera ainsi publiée sur le web selon les mêmes principes. Dans les années qui viennent, l'expérience acquise avec *data.bnf.fr* devrait éclairer de manière déterminante la réflexion et la stratégie de la BnF pour l'évolution de ses catalogues.

### **ABES : FRBRisation du Sudoc**

FRBRiser le Sudoc : cela peut s'entendre soit au niveau de la production des métadonnées (le catalogue), soit à celui de leur exposition sur le Web, soit aux deux – étant entendu que des données FRBRisées dès leur production le seraient ipso facto une fois converties dans des formats de type RDF.

### **L'option « FRBRisation de la production des métadonnées »**

Dans un premier temps, l'ABES a retenu d'explorer la faisabilité de cette dernière option. La capacité reconnue de CBS à représenter tous types de liens entre différentes entités bibliographiques, de même que ses grandes capacités de paramétrage, rendaient cette option crédible.

Le projet consistait donc à expérimenter sur un serveur de test un paramétrage de la base de production du Sudoc de manière à ce qu'il soit possible d'y cataloguer selon le modèle FRBR «vrai», à raison d'une notice par type d'entité (œuvre, expression, manifestation, item). Dans la mesure où il rejoignait en partie celui du réseau des bibliothèques publiques néerlandaises,

<sup>1</sup> <http://ifverso.com>

<sup>2</sup> <http://plindenbaum.blogspot.fr/2011/07/drawing-svg-timeline-with-httpdatabnf.html>

<sup>3</sup> CKAN : <http://thedatahub.org/dataset/data-bnf-fr>

<sup>4</sup> <https://sites.google.com/site/catbnf/>



OCLC a manifesté un certain intérêt pour ce projet suite à sa présentation lors de la réunion annuelle des « partenaires CBS » (utilisateurs et fournisseurs) en septembre 2011 à Berlin. Les représentants de la DNB (Deutsche Nationalbibliothek) se sont également déclarés intéressés à l'expérience (la DNB était alors sur le point d'annoncer officiellement son choix d'appliquer RDA).

En décembre 2011 l'ABES a donc décrit le projet de manière à le faire évaluer et chiffrer par OCLC. Devant les zones d'imprécision de la réponse d'OCLC et les coûts annoncés d'une part, et la DNB n'ayant pas confirmé son intérêt pour le projet d'autre part, l'ABES n'a pas procédé à l'expérimentation projetée.

Le projet initial, déjà modifié à la suite des discussions avec OCLC (pas de notices pour l'entité expression des FRBR), doit de toute façon être revu à la lumière du projet de SGB mutualisé porté par l'ABES. La FRBRisation ne se fera vraisemblablement pas dans CBS. Deux options sont à présent sur la table : se placer dans la perspective de solutions apportées par le fournisseur du SGB mutualisé, ou bien FRBRiser indépendamment de tout outil comme cela a été fait pour les thèses.

### **Quid des données ?**

Quel que soit le devenir technique du Sudoc, les métadonnées qu'il renferme, et dont la production se poursuit, doivent sans attendre être adaptées aux standards émergents.

L'ABES a consenti un effort d'information important sur les changements en cours au niveau international (mise en œuvre des modèles FRBR et FRAD dans des codes de catalogage tels que REICAT et RDA), sur la nécessité de placer les métadonnées des bibliothèques dans le web de données, sur le renouveau des catalogues. L'agence a sollicité ses réseaux pour participer aux travaux du groupe technique « RDA en France » de même qu'au groupe de formateurs mis en place au niveau national, groupes dans lesquels les bibliothèques de l'enseignement supérieur sont bien représentées. Des recommandations ont été données afin de préparer la FRBRisation des données.

Autant de raisons pour entreprendre rapidement un travail concret sur les métadonnées du Sudoc.

L'ABES prévoit donc de mettre en place dès que possible un jeu de consignes de catalogage précises, avec pour objectif l'identification systématique des œuvres, et autant que possible de leurs expressions, dans les métadonnées produites.

Des modifications du paramétrage de la base CBS devront probablement être pratiquées (mise en œuvre des zones ad hoc du format UNIMARC), mais elles seront toutefois limitées par rapport au projet initial décrit ci-dessus.

Parallèlement, l'ABES mettra à l'étude des procédures de traitement automatique rétrospectif, de manière à aligner au mieux les données déjà présentes dans le Sudoc sur celles produites dans le nouvel environnement.

Ces travaux seront menés en relation étroite avec le chantier d'enrichissement de la sortie RDF/XML des notices bibliographiques du Sudoc, qui se donne pour objectif d'exprimer de plus en plus de notions encodées en MARC, en employant des vocabulaires RDF du métier comme RDA ou ISBD. Ce chantier démarre en septembre 2012.

## 3- Adoption de RDA en France : les orientations

### Qu'est-ce que RDA ?

RDA (Ressources : Description et Accès) est le code de catalogage développé par la communauté des bibliothèques anglo-américaines pour adapter les catalogues de bibliothèques à leur nouvel environnement – le Web, ses technologies et ses pratiques. S'il remplace les règles de catalogage anglo-américaines (AACR2), il a été conçu comme un code à vocation internationale, à une époque où les échanges et la dérivation de notices sont des pratiques de plus en plus répandues.

RDA présente un intérêt majeur pour l'évolution des catalogues de bibliothèques. En effet, le code a été développé pour permettre aux données des bibliothèques d'être présentes sur le Web et incluses dans la circulation des données sur le Web : sur le chemin des internautes, en relation avec des données produites par d'autres acteurs et réutilisables. Pour atteindre cet objectif, le code est fondé sur les modèles FRBR et FRAD : produire des données FRBRisées permet de répondre de manière plus pertinente aux besoins des utilisateurs, d'aller vers une nouvelle structure des catalogues (bases de données relationnelles ou orientées-objet, proches de la structure en réseau du Web) et de cesser de produire des données dans un format propre au monde des bibliothèques (MARC) pour adopter des formats largement partagés, notamment ceux du web (RDF en particulier).

### RDA en France : groupe technique et groupe stratégique

Suite à la publication en juin 2010 de RDA, le GC46 de l'AFNOR a décidé de créer deux groupes de travail pour examiner l'opportunité d'adopter ce nouveau code de catalogage en France : un groupe technique chargé d'instruire les différents aspects de l'adoption de RDA en France<sup>5</sup> et un comité stratégique qui définit les orientations et prend les décisions.

L'analyse des règles de RDA menée par le Groupe technique a montré que plusieurs problèmes font obstacle à une adoption immédiate de RDA en France. En effet, l'ambition novatrice du code est contrecarrée par la volonté de la communauté des bibliothèques anglo-américaines de maintenir une compatibilité avec les notices existantes, créées selon les AACR2, ce qui a pour conséquences :

- une interprétation particulière et contestable du modèle FRBR qui réduit considérablement l'intérêt de son adoption, pour la recherche<sup>6</sup> comme pour le catalogage<sup>7</sup> ;
- le maintien de règles archaïques qui semblent ignorer les possibilités offertes par les progrès des technologies informatiques, pour la recherche<sup>8</sup> comme pour le catalogage<sup>9</sup> ;
- un point de vue anglo-américain prédominant qui ne tient compte ni des réalités historiques ou culturelles des autres pays<sup>10</sup> ni de leurs besoins<sup>11</sup>.

En outre, RDA prévoit plusieurs scénarios possibles d'implémentation qui autorisent une utilisation du code aussi bien dans des catalogues traditionnels de notices MARC, même sans liens entre elles, qu'en allant vers une réelle mise en œuvre du modèle FRBR, avec des données liées décrivant les différentes entités du modèle (et encodées selon les standards du Web, RDF en particulier). Seul ce dernier scénario (désigné comme le scénario 1 d'implémentation de RDA) est réellement novateur et constitue un progrès pour les catalogues des bibliothèques françaises.

---

<sup>5</sup> Site : <http://rda-en-france.enssib.fr/>

<sup>6</sup> Création d'« œuvres » ou de versions de celles-ci uniquement pour refléter le contenu de chaque publication, ce qui nuit à la fédération des données autour de l'œuvre et de ses différentes déclinaisons

<sup>7</sup> Multiplication de notices

<sup>8</sup> Recherche par mots, navigation en suivant les liens entre notices

<sup>9</sup> Uniformisation et correction automatique des points d'accès grâce aux liens entre notices bibliographiques et d'autorité

<sup>10</sup> Par exemple : forme des noms, tradition juridique

<sup>11</sup> Identification précise des noms géographiques

Lors de la première réunion du Groupe stratégique sur l'adoption de RDA en France, le 25 mars 2011<sup>12</sup>, il a été décidé de viser l'objectif de rejoindre le scénario 1 d'implémentation de RDA mais de ne pas adopter RDA en l'état car cela signifierait une régression par rapport à la situation actuelle des catalogues en France et de « préparer l'avenir », c'est-à-dire :

- avoir pour objectif l'adoption de RDA en France afin de ne pas isoler les bibliothèques françaises ;
- travailler à rendre possible cette adoption en s'appuyant sur une concertation européenne pour faire évoluer le code sur les points qui constituent des obstacles majeurs ;
- travailler à la FRBRisation des catalogues des agences bibliographiques nationales, car la FRBRisation est essentielle pour exposer les catalogues de bibliothèques sur le Web de données ;
- mettre à jour les normes françaises de catalogage par rapport aux textes internationaux de référence (ISBD intégré) pour disposer dans la période de transition vers RDA de règles adaptées à la nouvelle organisation, FRBRisée, de l'information bibliographique.

Un important travail a donc été mené au niveau européen, avec la participation française au Groupe d'intérêt européen sur RDA (European RDA Interest Group, EURIG<sup>13</sup>, dont la BnF assure la vice-présidence), créé pour identifier et défendre des intérêts communs et avoir plus de force vis-à-vis du JSC (Joint Steering Committee for development of RDA) qui maintient le code. C'est le canal qui a été choisi pour présenter les demandes d'évolution de RDA jugées nécessaires par la France. Quatre propositions d'évolution de RDA ont d'ores et déjà été transmises à l'été 2012 au JSC par l'intermédiaire d'EURIG ; elles visent à maintenir le même niveau de précision pour l'identification des ressources que celle qui existe actuellement dans les catalogues français. Des travaux sont en cours sur le traitement des agrégats (illustrations, préfaces et autres contributions additionnelles à l'œuvre publiée, compilations d'œuvres).

Toutefois, le JSC n'a qu'une réunion par an et il ne faut pas s'attendre à une acceptation immédiate de toutes les propositions d'évolution présentées par EURIG. L'accueil réservé par le JSC aux propositions envoyées cette année constituera un test quant à son ouverture aux besoins européens. Mais certaines propositions qui impliquent une révision radicale des choix faits par RDA pourraient prendre plusieurs années avant d'aboutir. Par ailleurs, l'adoption de RDA par la France impliquerait non seulement une conformité du code aux attentes françaises mais aussi un engagement de mise en œuvre opérationnel par les établissements impliquant un accord de leurs tutelles et une plus large adhésion de la communauté professionnelle.

Pour éviter toute régression des catalogues français, l'évolution de RDA sur un certain nombre de points majeurs constitue un préalable à son adoption en France. En attendant, la révision des normes françaises de catalogage s'effectue dans la double perspective de permettre la production de données FRBRisées et de préparer la transition vers RDA, avec une convergence des règles toutes les fois où cela est possible.

Au-delà des règles de catalogage, il faut aussi que les conditions de mise en œuvre du scénario 1 d'implémentation soient réunies pour pouvoir produire en catalogage courant des données FRBRisées et diffusables en RDF, et ainsi tirer tout le profit attendu du remplacement des normes françaises de catalogage par RDA.

## 4- La question des licences

### Une licence pour les données et les services de l'ABES.

---

<sup>12</sup> Compte-rendu accessible en ligne : <http://rda-en-france.enssib.fr/sites/rda-en-france.enssib.fr/files/CR%20RDA%20Comité%20stratégique%2025%20mars%202011.pdf>

<sup>13</sup> Site : <http://www.slainte.org.uk/eurig/rda.htm>

## Contexte

Le Sudoc est composé de deux types de notices :

- les notices originales produites par les catalogueurs des bibliothèques membres du Sudoc
- celles dérivées de bases externes (BnF, WorldCat, ISSN, DNB, etc.)

Si l'ABES a la maîtrise des notices produites par son réseau, ce n'est pas le cas des notices dérivées de bases externes auxquelles s'applique la licence du producteur de la base.

Les conditions de réutilisation des notices du Sudoc sont régies par les dispositions d'une convention signée entre l'ABES et les bibliothèques Sudoc. Pendant longtemps très restrictives, ces dispositions ont été libéralisées en 2008 après que l'ABES a négocié un assouplissement des licences avec la plupart de ses fournisseurs de bases externes.

Ces dispositions restent toutefois en deçà des attentes de ses membres et ne concernent que les seules bibliothèques Sudoc alors que l'ABES a ouvert depuis 2010 des services et applications qui s'adressent à un cercle plus large : référentiel d'autorités IdRef, STAR, STEP, thèses.fr, exposition du Sudoc et Calames sur le web des données.

Conformément à une politique d'ouverture des données et des services consacrée dans son projet d'établissement, ce processus s'est accéléré. L'ABES souhaite encourager non seulement l'accès à ses données et à ses services mais également leur réutilisation par ses réseaux aussi bien que par des utilisateurs extérieurs à ses réseaux. Aujourd'hui, ces données et ces services sont mis à disposition sans qu'aucune condition d'accès et de réutilisation ne leur soit attachée. L'ABES entend désormais rendre ces conditions explicites.

## Principes et contraintes

L'ABES souhaite attacher aux données et services qu'elle rend accessibles sur le Web une licence la plus libérale possible, y compris en autorisant des réutilisations commerciales pour des utilisations innovantes par des prestataires privés. Etant donné que certaines de ses données ont un autre propriétaire que l'ABES, il est probable que toutes les données exposées par l'ABES ne pourront pas être soumises à la même licence.

L'ABES souhaite imposer la mention de l'ABES comme titulaire (attribution).

Les modes d'accès permis par l'ABES seraient les suivants : service Z3950, Web service SRU, service Web d'accès aux notices bibliographiques incomplètes en RDF/XML, Dump, OAI-PMH.

## Licences étudiées

Depuis quelques années, la vague de l'open data touche également les données bibliographiques. L'ABES a étudié avec le concours d'un cabinet juridique spécialisé en propriété intellectuelle plusieurs types de licences :

- Creative Commons Attribution (CC-BY)
- Creative Commons Public Domain (CC0)
- Creative Commons Attribution 2.0 (CC-BY 2.0)
- Open Data Commons Attribution License (ODC-BY)
- Licence Ouverte / Open Licence de la mission Etalab (France)

## Recommandation

A l'issue de ces études, l'ABES envisage de retenir la Licence Ouverte, recommandée par la mission Etalab, chargée sous l'autorité du Premier ministre d'ouvrir le plus grand nombre de données publiques des administrations de l'Etat et de ses établissements publics.

Cette Licence Ouverte a été conçue pour faciliter la réutilisation libre et gratuite des informations publiques, telles que définies par l'article 10 de la loi n°78-753 du 17 juillet 1978, sous réserve de mentionner la source de l'information. Ses dispositions sont conformes à l'article 12 de l'ordonnance de 2005 qui prévoit que la réutilisation des informations publiques est soumise à la condition que ces dernières ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées.

La licence Etalab autorise les utilisateurs à réutiliser les informations, à les reproduire, les redistribuer, et les exploiter à titre commercial notamment en les combinant ou en les incorporant avec d'autres informations ou produits.

Cette licence est compatible avec toute licence libre qui exige a minima la mention de paternité : licences « Open Government Licence » (OGL) du Royaume-Uni, « Creative Commons Attribution 2.0 » (CC-BY 2.0) de Creative Commons et « Open Data Commons Attribution » (ODC-BY) de l'Open Knowledge Foundation, qui vient en outre d'être adoptée par OCLC.

Toutefois, s'agissant des droits de propriété intellectuelle détenus par des tiers, il est précisé que « le producteur garantit que l'Information ne contient pas de droits de propriété intellectuelle appartenant à des tiers ». En conséquence, si l'ABES n'a pas acquis les droits de propriété intellectuelle relatifs à certaines données, la licence Etalab ne pourra être reprise en l'état. En effet, dans cette hypothèse il conviendra à minima d'adapter la licence afin de tenir compte des droits de propriété intellectuelle détenus par des tiers, dans la mesure où l'ABES ne peut accorder aux utilisateurs plus de droits qu'elle n'en possède.

L'ABES s'attachera toutefois à obtenir l'autorisation de ces tiers pour la libre réutilisation de leurs données, dans le cadre de la Licence Ouverte Etalab.

## Décision

Le choix de la licence relève du Conseil d'administration de l'ABES qui en délibérera en novembre 2012 après avis de son Conseil scientifique.

## Licences des métadonnées : politique et perspectives à la BnF

La BnF est directement productrice de plus de 98% des notices bibliographiques présentes dans ses catalogues et n'est donc pas soumise à d'importantes dépendances juridiques concernant le statut juridique de ses métadonnées. Si elle entend développer à l'avenir la dérivation de notices depuis des réservoirs extérieurs et en collaboration avec les éditeurs français, la situation actuelle lui permet de définir sa politique de licences de manière assez autonome.

Comme d'autres grandes bibliothèques nationales (British Library, Deutsche Nationalbibliothek, Bibliothèque Royale de Suède...) ou universitaires (Harvard) et de grands acteurs commerciaux ou collaboratifs comme OCLC ou la fondation DBpedia, la BnF souhaite résolument accompagner sa stratégie d'insertion dans le Web de données par une politique d'ouverture des licences associées à ses métadonnées. Cette évolution lui paraît à la fois cohérente et nécessaire:

- **du point de vue de ses missions** : Selon son décret fondateur, la BnF a pour mission « d'assurer l'accès du plus grand nombre aux collections. ». L'objectif d'exposition et de réutilisation maximale des métadonnées de la BnF sur le Web est conforme à cette mission fondamentale et permet, de fait, de la prolonger et de l'amplifier.
- **du point de vue de la politique gouvernementale** : L'inscription de la BnF dans la démarche d'ouverture des données publiques préconisée par le gouvernement permet de se conformer aux bonnes pratiques de l'administration. La licence d'Etat établie par la mission gouvernementale Etalab pour encourager la réutilisation des données publiques avec le maintien de leur mention de source (attribution) lui apparaît de fait, comme à l'ABES, l'option juridique la plus pertinente.
- **du point de vue des usages et des coopérations** : La trajectoire de long terme des catalogues de la BnF repose sur le développement de l'exposition de leurs métadonnées dans le Linked Open Data. La présence de la BnF dans ce réseau ne sera possible que si elle adopte un régime juridique compatible avec celui des autres acteurs du Web sémantique. Au niveau national, il est indispensable que l'ABES et la BnF fassent des choix de licences compatibles. Au niveau international, l'interopérabilité juridique avec des organismes comme OCLC (qui recommande la licence ODC-BY pour

les données de WorldCat) est également nécessaire pour permettre la dissémination et l'échange.

- **du point de vue du marché** : l'économie se juge sous deux angles :
  - En s'inscrivant dans le Web de données, la BnF constituera non seulement un réservoir de données aisément réutilisables par d'autres organismes publics (dont des bibliothèques, et l'ABES au premier chef) mais pourra aussi récupérer elle-même des données produites par d'autres et jouer ainsi pleinement la carte de la complémentarité et de la coopération dans le but de mutualiser les coûts.
  - Les grands acteurs du Web ont fait la démonstration d'un modèle économique où les ressources propres découlent indirectement de l'audience : c'est la libre diffusion de métadonnées qui génère l'affluence, laquelle permet en retour de développer d'éventuels services payants.

La BnF a pris ou s'apprête à prendre des décisions allant dans ce sens :

- En 2011, elle a adopté la licence d'Etat préconisée par Etalab pour toutes les données au format RDF exposées dans *data.bnf.fr*.
- Son président s'est engagé au nom de la Conférence des Directeurs de Bibliothèques nationales européennes (CENL) en faveur de l'adoption du Data Exchange Agreement proposé par Europeana, qui implique la libération totale des métadonnées (licence creative commons CC-0). La BnF a signé cet accord, qui a libéré les métadonnées en Dublin Core associées aux collections exposées dans Gallica et numérisées par l'établissement. Il en sera de même pour fonds numérisés par les bibliothèques françaises partenaires de Gallica qui auront pris la même décision.
- La BnF souhaiterait à présent faire de la licence d'attribution recommandée par Etalab le régime de référence pour l'ensemble de ses métadonnées quel qu'en soit le format et le protocole de récupération, à l'exception des métadonnées de collections de Gallica, bénéficiant d'un statut d'ouverture totale du fait des accords spécifiques avec Europeana. Ceci implique la révision des tarifs et conditions d'utilisation de ses produits et services bibliographiques.

## 5- L'enjeu du signalement de la documentation électronique

### Signalement des documents électroniques : état des lieux

Depuis plusieurs années, les ressources des bibliothèques ne sont plus signalées dans un seul catalogue comme du temps de la prépondérance de l'imprimé. Cette situation se vérifie aussi bien à la BnF et dans le Sudoc que dans les autres catalogues collectifs.

Les catalogues de bibliothèques ont en effet été développés à une époque où n'existaient que les ressources analogiques. Ils recensent des objets physiques alors que la large dissémination des ressources électroniques, souvent instables, dont un nombre croissant n'ont pas d'équivalent papier, et la multiplication des bibliothèques numériques ont engendré une situation hybride.

La plupart des établissements ont opté pour un signalement de leurs ressources électroniques (en premier lieu les périodiques) sur des outils s'appuyant sur des bases de connaissances commerciales (ExLibris, EBSCO, Serials Solutions,...), ces ressources restant largement ignorées des catalogues locaux comme du Sudoc. La BnF signale certaines ressources numériques dans son catalogue général (en particulier les documents numérisés dans Gallica) mais elle a développé ou acquis par ailleurs des applications propres au signalement de la documentation électronique : répertoire de signets, archives de l'Internet, solution Discovery d'EBSCO qui a récemment succédé à A to Z pour le signalement de ses acquisitions électroniques.

Cet état de fait pose plusieurs problèmes :

- séparation des solutions de signalement pour les ressources analogiques et numériques, qui ne permet pas une véritable recherche fédérée ;
- qualité des données : les outils commerciaux se focalisent sur l'accès à un instant T à une ressource en ligne, font passer la qualité bibliographique au second plan et ne s'intéressent pas – pour le moment en tout cas – aux potentialités du Linked Data ;
- disponibilité des données : certains fournisseurs de bases de connaissance sont également fournisseurs de contenus (EBSCO, Serials Solutions via Proquest) et se livrent à une concurrence acharnée en refusant de livrer à leurs concurrents les métadonnées les concernant.

Jusqu'à présent, les tentatives de l'ABES pour apporter une réponse à ces problèmes se sont toutes appuyées sur des outils calibrés pour la gestion de stock et se sont soldées par un échec. L'orientation vers des outils dits de nouvelle génération, traitant au même niveau les documents numériques et les documents analogiques devrait permettre de résoudre certains de ces problèmes. Par ailleurs, d'autres leviers juridiques (dépôt légal, pour la BnF) et financiers (licences nationales, pour l'ABES) peuvent renforcer le dispositif d'un signalement de qualité des ressources électroniques.

### Les nouveaux leviers

Depuis 2006 (loi DADVSI) la collecte des sites Internet du domaine français est réalisée par la BnF au titre du dépôt légal. Ces collectes sont réalisées à l'aide de robots moissonneurs qui copient pages, images, animations, fichiers audios et vidéos, etc. Compte tenu de la masse à traiter, tous les sites ne peuvent pas être archivés, et ceux qui le sont ne sont pas décrits à l'unité. Dans une application qui leur est propre, les archives de l'Internet font essentiellement l'objet d'une indexation par adresse URL, et d'une indexation plein texte sur des corpus. Leur valorisation via *data.bnf.fr* est par ailleurs engagée.

Le dépôt légal des publications numériques françaises commercialisées sur le Web, autorisé depuis peu (décret du 19/12/2011), n'est en revanche pas encore effectif. La priorité sera vraisemblablement donnée à la collecte et au signalement des livres numériques, afin de remédier au plus vite à l'absence de ces ressources émergentes dans les collections de la BnF et dans les recensements de référence dont elle a la charge. Le dépôt des monographies numériques pourrait être effectué par les éditeurs ou, par délégation de ceux-ci, par les e-distributeurs. Les livres numériques français ainsi déposés pourraient faire l'objet d'un signalement à l'unité au catalogue général et d'une annonce dans Nouveautés Editeurs, par exemple, et dans la Bibliographie nationale française. Le dialogue entre la BnF et les représentants des éditeurs français pour amorcer la mise en œuvre de ce dispositif s'est engagé à l'automne 2012. Il devrait pouvoir s'appuyer notamment sur les échanges déjà initiés en matière de normalisation et d'échange de métadonnées entre éditeurs et bibliothèques dans le cadre des travaux de la Commission de Liaison interprofessionnelle du livre (CLIL)<sup>14</sup>. Pour les périodiques en ligne, et la presse nationale et régionale en particulier, la BnF envisage plutôt de s'appuyer sur les processus de moissonnage du Web. Une expérimentation est en cours avec le journal Ouest France, qui permet la récupération quotidienne des fichiers PDF. Les solutions d'accès et de référencement de ces fichiers sont en cours d'instruction.

Le signalement de périodiques ou de livres électroniques publiés à l'étranger et ne ressortant donc pas de l'obligation de dépôt légal peut quant à lui bénéficier du programme d'acquisitions de documentation scientifique en licences nationales dans le cadre d'ISTEX. Ce projet comprend en effet l'acquisition des métadonnées produites par chaque éditeur dans le cadre de la constitution de sa plateforme. Cela permet de s'affranchir du recours à des intermédiaires pour obtenir ces métadonnées et ainsi de les redistribuer plus librement. La BnF souhaite s'inscrire dans ce dispositif que l'ABES est chargée de piloter au niveau national.

### Collaboration et travail en réseau

L'offre éditoriale, analogique et numérique, est en accroissement constant.

<sup>14</sup> Un travail de normalisation des données indispensables à la commercialisation des livres numériques a été engagé et une première version du Guide des bonnes pratiques de l'ONIX 3.0 a été publié début 2012. Voir : <http://www.clil.org/information/detailDoc.html?docId=33>

La BnF ne peut plus décrire seule tous les documents du patrimoine français qu'elle doit conserver. Pour absorber le travail supplémentaire généré par la prise en compte de nouveaux types de ressources tels que les livres numériques, la mutualisation et la coopération s'imposent dans la production des métadonnées de référence. Les bibliothèques universitaires ont déjà une habitude éprouvée de travail en réseau à l'échelle nationale, mais le contexte du numérique rend les frontières peu pertinentes.

Côté BnF, la collaboration se ferait directement avec les éditeurs, via le dépôt légal. Les outils nécessaires au dépôt légal des livres numériques existent déjà pour les livres imprimés, mais doivent être développés : il s'agit de l'application Extranet du dépôt légal<sup>15</sup>, mise en place en 2009 pour gérer en ligne les déclarations de dépôt légal des documents analogiques et leurs visuels. Des éléments de la déclaration de dépôt (notamment les données descriptives fournies par les éditeurs), sont d'ores et déjà chargés dans le catalogue général où ils constituent une pré-notice enrichie ensuite selon les normes par les catalogueurs de l'établissement. De la même façon, les notices descriptives de référence des livres numériques déposés pourraient être élaborées de façon collaborative par les éditeurs et les bibliothécaires de la BnF : données descriptives, éléments d'identification des contributeurs, image des couvertures, table des matières, résumé du contenu, indexation sujet, public visé, etc. Le fichier numérique déposé par l'éditeur inclurait alors tout ou partie de ces métadonnées, à convertir et importer ensuite dans le catalogue général avant un éventuel enrichissement par les catalogueurs.

Les enrichissements des métadonnées obtenues dans le cadre des licences nationales (lien aux autorités,...) se feraient quant à eux de manière automatique grâce à une série de traitements effectués par l'ABES. Ils pourront bénéficier aux bibliothèques et à leurs usagers via l'amélioration des bases de connaissances commerciales ou, idéalement, via la participation à des projets internationaux d'ouverture et de mise à disposition de métadonnées de ressources électroniques (GoKB)<sup>16</sup>.

## **6- Formation, communication, accompagnement au changement**

L'évolution du paysage bibliographique aura des incidences majeures sur le catalogage, qui constitue un aspect important du métier et de l'identité professionnelle des bibliothécaires, mais aussi sur les catalogues eux-mêmes, dont les systèmes informatisés représentent une part importante dans les budgets des établissements et de l'Etat. Ces perspectives génèrent de ce fait une réelle appréhension aux différents niveaux de la profession, tant chez les catalogueurs que chez les directeurs d'établissement.

Le groupe technique sur l'adoption de RDA en France a déjà engagé, par la publication d'un site d'information dédié, d'une plaquette et de supports de formation, un important travail de vulgarisation technique en regard des enjeux liés aux évolutions des technologies de l'information sur le Web. Certains organismes de formation, tels les CRFCB, sont mobilisés et proposent déjà des mises à niveau sur les FRBR par exemple. Pour aller plus loin sur ces sujets qui requièrent une vision nationale, il est aujourd'hui nécessaire de porter la communication à un plus haut niveau stratégique, afin de dissiper les malentendus et de lever les appréhensions mais aussi de trouver des moyens auprès des organismes publics capables de relayer et de soutenir les besoins de communication et de formation induits par les évolutions à venir.

### **Nécessité d'un plan de communication national**

Les actions de formation et d'information professionnelles ne pourront ainsi être véritablement efficaces que précédées d'un plan de communication porté par les tutelles et cohérent avec les décisions stratégiques que ce rapport souhaite initier. La plupart des établissements relevant de l'Enseignement supérieur appartiennent à un réseau structuré porté par l'ABES, dont les relais permettent une bonne concertation et diffusion de l'information. La situation est plus

---

<sup>15</sup> <http://depotlegal.bnf.fr/>

<sup>16</sup> Global Open KnowledgeBase. <http://gokb.org/>



compliquée pour les établissements de lecture publique qui ne sont pas insérés dans un réseau national aussi formel et relèvent de collectivités locales très diverses. Leurs problématiques se sont par ailleurs beaucoup éloignées des questions de catalogage, ce qui est une bonne chose lorsqu'il s'est agi de développer la récupération de notices des agences bibliographiques, mais ce qui rend plus difficile aujourd'hui l'adoption de bonnes pratiques sans compréhension des nouveaux enjeux. En collaboration avec le groupe technique sur l'adoption de RDA en France, le CNFPT a programmé au premier semestre 2013 quatre journées nationales à destination des directeurs de bibliothèques. Elles pourraient être l'occasion, pour le ministère de la Culture, d'aborder directement au niveau politique des décideurs la stratégie nationale définie par le comité stratégique bibliographique. L'implication des associations professionnelles apparaît comme un autre élément important de la concertation et de la communication sur ces sujets.

### **Sensibiliser et former aux pré-requis**

Un premier niveau d'information professionnelle de qualité est aujourd'hui assuré notamment sur les sites web ou blogs hébergés entre autres par l'ABES, la BnF ou l'ENSSIB. Y sont exposés non seulement les enjeux du Web sur l'évolution des catalogues de bibliothèques, mais également les informations techniques permettant de faire le lien avec les pratiques bibliothéconomiques : modélisation FRBR des données bibliographiques, évolution des normes et formats de métadonnées, technologies mises en œuvre avec le Web sémantique, etc. Le groupe technique RDA s'est par ailleurs associé au Comité français Unimarc et à la FULBI pour organiser des rencontres avec les éditeurs de systèmes informatisés de gestion de bibliothèques en vue de l'évolution de leurs systèmes. Le service du Livre et de la lecture publie enfin des fiches pratiques pour accompagner les bibliothèques dans leurs projets numériques, dont l'une porte sur l'enjeu de la récupération des données bibliographiques. Outre les perspectives d'évolutions du métier à moyen terme et indépendamment du calendrier des différentes étapes à mener, il s'agit en effet de s'assurer que les bibliothèques respectent un ensemble de bonnes pratiques sans lesquelles il leur sera impossible de faire évoluer leurs catalogues. Avant même la fourniture de données FRBRisées par les agences bibliographiques et l'évolution des systèmes eux-mêmes, il est en effet déjà possible aux bibliothèques de préparer la FRBRisation future de leurs catalogues en respectant un certain nombre de consignes, comme le maintien des identifiants et des liens indispensables à l'interopérabilité des données, la transcription exacte des titres originaux, le signalement systématique des codes de langues ou encore l'attention accordée au choix des accès auteurs. L'information professionnelle sur ces sujets devra être soutenue, amplifiée, coordonnée et si possible essaimer davantage vers les réseaux d'information et publications professionnelles.

Pour ce qui relève de la formation proprement dite, il n'apparaît pas utile d'enseigner dès aujourd'hui de nouvelles règles de catalogage alors que les futurs choix de structuration ne sont pas entièrement définis.

Des formations doivent cependant être dispensées de manière urgente et prioritaire pour permettre à la profession non seulement d'être sensibilisée aux enjeux sur les limites des catalogues actuels, mais également de se mettre à niveau sur le contexte normatif, les possibilités de la modélisation (FRBR, FRAD, FRSAD), les technologies du Web sémantique. C'est l'objet du support de référence élaboré par le groupe technique sur l'adoption de RDA en France à destination du réseau national de formateurs récemment constitué. Cette formation de mise à niveau est intitulée « Pré-requis à l'évolution des catalogues et des règles de catalogage », son support a été conçu pour être décliné en fonction du public et du temps de formation. 25 formateurs répartis équitablement sur le territoire national (y compris les Antilles-Guyane) ont été identifiés et formés ces derniers mois. Ils sont à la disposition des organismes de formation qui les solliciteront pour assurer cette formation aux pré-requis depuis septembre 2012. Leur contribution (temps de travail, frais de déplacements...) nécessite l'accord et l'appui des directions d'établissement et pourrait peut-être bénéficier d'appuis complémentaires au niveau des tutelles ou des organismes de formation.

En effet, si l'évolution du métier du catalogage vers la gestion de métadonnées devrait amener un nombre moindre de professionnels à cataloguer et un moins grand nombre de descriptions à la source du fait de la dérivation de notices, il sera essentiel de former les bibliothécaires travaillant au référencement des ressources à un travail d'analyse et d'indexation signalétique

qui permette de lier la description aux référentiels d'autorité, sur lesquels il leur faudra développer leur expertise.

### **L'adoption effective de RDA induira un chantier et des moyens conséquents**

La décision d'adopter RDA impliquera la mise en place d'un dispositif de formation plus ambitieux et plus coûteux. Il s'agira en effet d'assurer cette fois la transition effective et opérationnelle des règles françaises à RDA et à la gestion de catalogues FRBRisés. Il conviendra alors de créer de nouveaux supports de formation, des exercices pratiques, des didacticiels, et sûrement étoffer le cercle des formateurs. Les plateformes de formation à distance (celles de l'ABES et de l'Essib par exemple) pourraient également être utilisées. Pour cette campagne de formation de grande ampleur, un dispositif ad hoc sera à bâtir avec l'appui des organismes de formation compétents.

## **7- L'évolution des SIGB et la structuration des métadonnées publiques nationales : quelles perspectives ?**

### **L'évolution des SIGB : tendances du marché international**

Le système intégré de gestion de bibliothèque (SIGB) reste depuis trente ans le principal système informatique local de la plupart des bibliothèques tant publiques qu'universitaires. L'intégration signifiait l'appui de toutes les fonctions de la bibliothèque sur un catalogue unique de toutes ses collections. L'usage par le SIGB d'un format MARC pour les données du catalogue (en France le plus souvent Unimarc) est au fondement du catalogage partagé entre bibliothèques et du réseau SUDOC administré par l'ABES pour les établissements d'enseignement supérieur et de recherche. A la BnF, le système d'information développé pour l'ouverture de l'établissement repose sur les mêmes principes, nombre de fonctions vitales du système (communication, conservation, numérisation...) s'appuyant sur le catalogue (en Intermarc).

Depuis une bonne dizaine d'années ce modèle s'essouffle. Jour après jour les collections des bibliothèques et surtout l'usage de leur public sont de moins en moins simples à enfermer dans un catalogue local et s'immergent de plus en plus dans un continuum avec des bases de données distantes, publiques ou éditoriales, nationales ou internationales, et plus largement dans le Web. Catalogues locaux et nationaux connaissent à des degrés divers une baisse d'usage au profit de moteurs de recherche plus ouverts et leur modèle traditionnel de production se révèle incapable de fournir un accès unifié à l'ensemble des ressources imprimées et en ligne utilisées par le public des bibliothèques. Comme indiqué précédemment, celles-ci ont complété les SIGB par d'autres systèmes locaux dédiés aux ressources en ligne, mais elles peinent à gérer et à mettre en cohérence les différents dispositifs.

Certains fournisseurs proposent depuis plusieurs années de remplacer le catalogue public des bibliothèques par des systèmes de découverte hébergés qui agrègent les métadonnées des bibliothèques et des éditeurs pour offrir un accès unifié à toutes les ressources locales ou distantes. Quelques-uns de ces fournisseurs proposent aujourd'hui de remplacer les SIGB et la plupart des systèmes locaux de gestion et de diffusion des ressources en ligne par un service Web mutualisé de gestion de la bibliothèque complémentaire du système de découverte pour le public. Ces systèmes de gestion de bibliothèque, dits de nouvelle génération, constituent des réseaux internationaux qui concurrenceront sérieusement les réseaux nationaux comme le SUDOC dans les années qui viennent.

### **Le projet de l'ABES : un SIGB mutualisé pour l'enseignement supérieur**

L'ABES est très attentive à l'émergence de ces systèmes et à la demande de mutualisation des SIGB formulée par les représentants des établissements au conseil d'administration et inscrite dans son projet d'établissement 2012-2015. Elle a donc décidé en 2012 de collaborer avec les

établissements intéressés du réseau SUDOC pour l'étude d'impact, le choix et la mise en œuvre d'un système de gestion de bibliothèque de nouvelle génération qui remplacerait à terme le SUDOC. Il est vraisemblable que la période de transition entre le SUDOC actuel et le nouveau système partagé de nouvelle génération s'étendra sur une dizaine d'années.

Deux systèmes de ce type sont actuellement en production hors de France et sont en cours d'évaluation par l'ABES. Ils se tiennent régulièrement à la pointe des technologies, y compris en matière de traitement des métadonnées, et proposeront dès 2013 à leurs utilisateurs un catalogage en RDA codé en Marc 21. Une de leurs caractéristiques par rapport au SIGB traditionnels est leur capacité à agréger des métadonnées de formats variés. L'ABES, dans la mesure de ses moyens, diffusera le plus largement possible l'expertise du nouveau réseau, au moins par le Web, au bénéfice de l'ensemble de la communauté française des bibliothèques et de leurs fournisseurs.

### **La situation de la BnF**

Contrairement à l'ABES, la BnF n'assume pas un rôle de pilotage et de coordination d'un réseau national de catalogage partagé et n'a pas de responsabilité de coordination en matière de SIGB parmi les établissements de lecture publique. La question de l'évolution technologique se pose d'abord pour elle du point de vue de son propre système d'information. A cet égard, comme détaillé au chapitre 2, elle adopte aujourd'hui une position prudente, conforme à ses moyens, qui n'envisage pas de modifier profondément son système ni son format de production avant quelques années, l'accent étant mis sur la stratégie de diffusion.

Toutefois le souhait de la BnF d'amplifier et de simplifier la diffusion de ses données bibliographiques et d'en faciliter la réutilisation non plus seulement par l'importation de notices bibliographiques en MARC mais en utilisant les technologies du Web sémantique implique une bonne prise en compte de l'évolution générale de l'environnement informatique des bibliothèques de lecture publique dans les prochaines années. Ceci implique une concertation forte de la BnF avec sa tutelle afin que ce que développe la BnF réponde effectivement aux attentes et contraintes des bibliothèques et aux potentialités du marché, ce qui implique un dialogue prospectif aussi bien avec les associations professionnelles et les acteurs du marché des logiciels documentaires.

## **8- Les études à conduire : deux propositions**

**1) Une étude d'impact analysant les conditions et les coûts d'évolution des systèmes des bibliothèques publiques** vers des SIG de nouvelle génération permettant une adaptation au nouveau modèle de l'information bibliographique s'avère nécessaire. Cette étude serait comparable à ce que l'ABES a engagé pour les établissements de l'enseignement supérieur et constituerait un préalable très utile pour tous les acteurs impliqués (Etat, collectivités, BnF, fournisseurs). Deux enquêtes ont été conduites en 2012 qui pourraient apporter des éléments d'analyse préliminaires sur l'état des lieux (systèmes, compétences, ressources) et les attentes des établissements, et également servir à la préparation de cette étude :

- l'enquête conduite par le sous-groupe sur les systèmes d'information du groupe technique RDA ;
- l'étude réalisée par la BnF sur la récupération de notices bibliographiques.

Les conclusions de ces deux études seront diffusées avant la fin de l'année.

**2) Une étude sur la structuration des données publiques nationales en matière de documentation** afin de favoriser l'interopérabilité et la complémentarité des données issues de l'Enseignement supérieur et de la recherche et de la BnF.

Les évolutions constatées rendront à terme la question des systèmes informatiques secondaire par rapport à celle des données. La communauté des bibliothèques françaises, des bibliothèques de recherche en particulier, ne doit pas hésiter à se raccrocher à des systèmes de documentation internationaux. Elle doit en revanche se mobiliser pour que les données de la recherche et des éditeurs français soient présentes et valorisées dans les systèmes internationaux. Ce point mérite une étude spécifique ou pour le moins un approfondissement de la réflexion stratégique entre les deux tutelles. Cette proposition serait conforme au projet d'établissement de l'ABES appelant à la définition d'une stratégie conjointe entre les opérateurs pour la production, l'acquisition et le moissonnage des métadonnées et à une répartition des responsabilités de chacun. Elle répondrait également aux souhaits de la BnF à la fois de sécuriser un espace public de données nationales de référence dans le Web sémantique et une bonne articulation des rôles avec l'ABES pour la fourniture et la maintenance de ces données.

## **GLOSSAIRE**

**AACR2** : Anglo-American Cataloguing Rules, Règles de catalogage anglo-américaines. Code de catalogage en usage dans les pays anglophones, traduit dans plus de 30 langues depuis 1967. La deuxième édition, datant de 1978 et ayant subi plusieurs révisions, s'appelle AACR2.

**ABES** : Agence bibliographique de l'enseignement supérieur

**BnF** : Bibliothèque nationale de France

**CALAMES** : Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur.

**CC-BY** : Creative Commons Attribution. Licence permettant la libre réutilisation des données, y compris à des fins commerciales, mais requérant la mention d'attribution de source.

**CC0** : Creative Commons Public Domain. Licence permettant la libre réutilisation des données, y compris à des fins commerciales, sans obligation d'attribution de la source.

**CLIL** : Commission de Liaison interprofessionnelle du Livre

**CNFPT** : Centre national de la fonction publique territoriale

**CRFCB** : Centres régionaux de formation aux carrières des bibliothèques

**DADVSI** : Droits d'auteur et droits voisins dans la société de l'information (Loi n° 2006-961 du 1<sup>er</sup> août 2006)

**DNB** : Deutsche Nationalbibliothek, Bibliothèque nationale d'Allemagne

**FRAD** : Functional Requirements for authority Records, Spécifications fonctionnelles des notices d'autorité. Modèle conceptuel pour les données d'autorité qui servent à contrôler les points accès dans les bases bibliographiques des bibliothèques, des archives et des musées. FRAD vient en complément de la modélisation des notices bibliographiques proposée par les FRBR. Il traite essentiellement des entités Personne, Collectivité, Famille et Œuvre, en étudiant les attributs de chacune de ces entités et les relations qu'elles entretiennent entre elles et avec les notices bibliographiques auxquelles elles donnent accès.

**FRBR** : Functional Requirements for Bibliographic Records, Spécifications fonctionnelles des notices bibliographiques. Modèle conceptuel de données bibliographiques, validé en 1997 par l'IFLA. Ce modèle définit les entités sur lesquelles porte l'information catalographique, ainsi que les relations qui existent entre ces entités. Il analyse la notice catalographique comme un ensemble de données reflétant essentiellement les attributs et les relations de quatre entités distinctes : Œuvre, Expression, Manifestation et Item.

**ENSSIB** : Ecole nationale des sciences de l'information et des bibliothèques

**EURIG** : European RDA Interest groupe, Groupe d'intérêt européen sur RDA

**IFLA** : International Federation of Library Associations, Fédération internationale des associations de bibliothécaires et d'institutions. L'IFLA est la principale organisation internationale représentant les intérêts des bibliothèques, des services d'information et de leurs usagers. Créée en 1927, elle rassemble 1700 membres de 150 pays.

**IdRef – Autorités Sudoc** : Identifiants et Référentiels – Autorités Sudoc est une application Web développée et maintenue par l'ABES qui permet à des utilisateurs et à des applications clientes d'interroger, de consulter, de créer et d'enrichir des autorités Sudoc.

**ISBD : International Standard for Bibliographic Description**, Description bibliographique internationale normalisée. Elaborée par l'IFLA, elle constitue un ensemble normatif de règles validées au niveau international, pour la description bibliographique de toute ressource publiée existant dans les bibliothèques, quel qu'en soit le support.

**ISSN : International Standard Serial Number, Numéro international normalisé des publications en série**. Code normalisé de 8 chiffres créé en 1974 qui identifie de façon brève, unique et univoque une publication en série. Il n'a aucune signification intrinsèque. Il se présente sous la forme de deux groupes de 4 caractères séparés par un tiret et précédés du préfixe ISSN.

**IST** : Information scientifique et technique

**ISTEX** : Initiative d'excellence de l'Information Scientifique et Technique. Le projet ISTEX s'inscrit dans le programme « Investissements d'Avenir », initié par le Ministère de l'Enseignement Supérieur et de la Recherche, dont l'ambition est de renforcer la recherche et l'enseignement supérieur français sur la scène mondiale. Ce projet est porté par quatre partenaires : le Centre National de la Recherche Scientifique (CNRS), l'Agence Bibliographique de l'Enseignement Supérieur (ABES), le Consortium Universitaire de Publications Numériques (Couperin), et l'Université de Lorraine agissant pour le compte de la Conférence des Présidents d'Université (CPU). Signée le 19 avril 2012 par l'ANR (Agence Nationale pour la Recherche), l'Etat et le CNRS, la convention d'attribution d'aide au projet s'élève à 60 millions d'euros sur trois ans dont 55 millions d'euros consacrés aux acquisitions des ressources documentaires et 5 millions d'euros alloués à la création de la plateforme destinée à héberger les données.

**JSC** : Joint Steering Committee for development of RDA. Réunissant des représentants des associations professionnelles principalement du monde anglo-américain et des bibliothèques nationales telles que la Library of Congress et la British Library, cet organisme assure le développement et la maintenance du code de catalogage RDA. Il examine les demandes d'évolution du code.

**LIBER** : Ligue des bibliothèques européennes de recherche

**MARC** : MACHine-Readable Cataloguing, Catalogage lisible en machine. Format le plus courant de structuration des données bibliographiques pour une exploitation informatique du catalogage. Le premier format MARC est apparu en 1965. Aujourd'hui il existe 2 formats MARC d'utilisation internationale : MARC21 géré conjointement par la Bibliothèque du Congrès à Washington et la Bibliothèque et archives nationales au Canada, et UNIMARC géré par l'IFLA. À la BnF, les notices sont saisies en INTERMARC et exportées en UNIMARC ou en INTERMARC au choix du client.

**OAI** : Open Archives Initiative, Initiative pour les archives ouvertes. L'Open Archive Initiative est une organisation internationale dont l'objectif est de créer des standards libres et ouverts pour faciliter la visibilité sur Internet des archives numériques. Elle est à l'origine du protocole technique OAI-PMH (Protocol for Metadata Harvesting, protocole pour la collecte des métadonnées) qui permet d'échanger des métadonnées descriptives de ressources numériques, afin d'en multiplier les points d'accès sur Internet et d'en améliorer le référencement par les moteurs de recherche généralistes. Seules les métadonnées sont échangées, les documents restent accessibles uniquement sur le site d'origine, via une URL contenue dans la notice OAI. Une bibliothèque, fournisseur de données, peut ainsi décider d'exposer ses métadonnées sur un serveur Internet (appelé "entrepôt") afin qu'elles soient automatiquement récupérées (ou moissonnées) par un ou plusieurs fournisseurs de services (autres bibliothèques, sites portails, moteurs de recherche, etc.). Le protocole OAI-PMH utilise XML comme langage de transport des métadonnées, HTTP

comme langage de communication entre serveurs et le Dublin Core simple comme format de métadonnées descriptives.

**OCLC** : Online Computer Library Center. Organisation mondiale à but non lucratif fondée en 1967 au service des bibliothèques. 72 000 bibliothèques dans le monde utilisent les services d'OCLC, notamment WorldCat, le catalogue collectif en ligne qu'il gère.

**RDA** : Resource Description and Access, Ressources : description et accès. Nouveau code de catalogage, à vocation internationale, ayant pour but d'adapter les catalogues à l'environnement numérique et aux nouveaux modes d'accès à l'information. Il a été publié en juin 2010, sous la forme d'un site Web payant, RDA Toolkit. C'est une norme de contenu où les règles de catalogage sont indépendantes de la présentation des données comme du format d'encodage. Conforme aux Principes internationaux de catalogage (2009), RDA repose sur les modèles conceptuels de l'IFLA, FRBR pour les données bibliographiques et FRAD pour les données d'autorité, et permet à terme l'évolution des catalogues et leur ouverture vers le Web sémantique.

**RDF** : Resource Description Framework, Cadre de description des ressources. Modèle pour l'échange de données sur le Web. C'est un des éléments constitutifs du Web sémantique, qui décrit les données, ainsi que les relations qui existent entre elles, selon la syntaxe : sujet – prédicat – objet.

**RERO** : Réseau des bibliothèques de Suisse occidentale

**SIGB** : Système intégré de gestion de bibliothèque. Un SIGB est un progiciel conçu pour automatiser les tâches en bibliothèque et notamment le catalogage en ligne (accès d'une bibliothèque à un réservoir de notices produites par une bibliothèque centrale ou mises en commun par plusieurs bibliothèques) ; la gestion de la circulation des documents ; les acquisitions.

**SNE** : syndicat national de l'Édition

**SRU/SRW** : SRU (Search/Retrieve via URL) et SRW (Search/Retrieve Web service) sont deux protocoles permettant d'interroger simplement, via le protocole HTTP du Web, des bases de données jusqu'alors cachées ou difficilement accessibles. Maintenus par la Bibliothèque du Congrès, ils sont fondés sur le langage CQL (Common Query Language, standard de syntaxe de requête) et succèdent à Z 39.50 en intégrant les avancées technologiques du Web, notamment les Web Services.

**SUDOC** : Le catalogue du Système Universitaire de Documentation est le catalogue collectif français réalisé par les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche et maintenu par l'ABES. Il comprend plus de 10 millions de notices bibliographiques qui décrivent tous types de documents (livres, thèses, revues, ressources électroniques, documents audiovisuels, microformes, cartes, partitions, manuscrits et livres anciens...).

**STAR** : application développée et maintenue par l'ABES qui constitue le guichet unique par lequel l'établissement de soutenance peut donner à ses thèses de nouveaux débouchés, qu'il s'agisse de débouchés incontournables (archivage, signalement) ou optionnels (serveurs de diffusion extérieurs à l'établissement, indexation du texte intégral).

**STEP** : Signalement des thèses en préparation. Application développée et maintenue par l'ABES, Step est l'interface professionnelle dédiée aux gestionnaires des thèses en préparation dans les établissements. Les gestionnaires sont à l'origine de la création de la thèse en préparation et valident notamment les données relatives à l'affiliation du doctorant. Les doctorants disposent également d'un accès sur authentification, à l'application, qui leur permet de prendre la main pour saisir et compléter certaines données.

**URI** : Uniform resource Identifier. Chaîne de caractères permettant d'identifier une ressource abstraite ou physique.

**VIAF** : Virtual International Authority File, Fichier d'autorité international virtuel. C'est un programme commun à plus d'une trentaine de partenaires à travers le monde, mis en œuvre et hébergé par OCLC. VIAF apparie et crée des liens entre les notices d'autorité établies pour une même entité (personne, collectivité, etc.) chez les partenaires. VIAF participe au Web sémantique et favorise l'utilisation des données d'autorité au-delà du monde des bibliothèques.

**W3C** : World Wide Web Consortium. Organisme qui définit les standards du Web.

**Web de données** : Extension du Web, permettant de relier non pas des documents (pages HTML) mais les données elles-mêmes, et de les rendre exploitables par des machines.

### **Web sémantique**

Ensemble de technologies permettant une description approfondie et structurée des données du Web en vue d'un accès performant et global à ces données. Le Web sémantique repose notamment sur le modèle RDF (Resource Description Framework) destiné à décrire les données et leurs relations entre elles selon la syntaxe suivante : sujet - prédicat - objet. L'intérêt du Web sémantique pour les bibliothèques réside dans la capacité à créer davantage de liens entre les données des catalogues, dans la même conception que les modèles de données tels que les FRBR, et à les rendre plus visibles sur le Web.

**WorldCat** : réseau mondial de services et de contenus de bibliothèques qui utilise le Web. Il est géré par OCLC. C'est le catalogue le plus volumineux qui soit avec plus de 250 millions de notices bibliographiques fin 2011. 34 bibliothèques nationales et la plupart des bibliothèques Sudoc y participent parmi les 72 000 bibliothèques partenaires.



# Compte rendu de la réunion du Comité stratégique bibliographique du 13/11/2012

Participants :

**Pour la BnF :** Jacqueline Sanson, Arnaud Beaufort, Denis Bruckmann, Catherine Dhérent, Gildas Illien, Véronique Falconnet (représentant Aline Girard)

**Pour le SLL :** Fabien Plazannet, Patricia Le Galèze

**Pour l'ABES :** Raymond Bérard

**Pour la MISTRD :** Alain Colas (représentant Michel Marian), Julie Ladant

Excusés : Nicolas Georges, Michel Marian, Aline Girard

## **Ordre du jour**

**1°) Présentation du rapport préparé conjointement par l'ABES et la BnF**

**2°) Discussion**

**3°) Actions en découlant**

### **Rappel du contexte et des objectifs**

Le comité « SU » (Système universitaire) – « SI » (Système d'information de la BnF) – « CCFR » (Catalogue collectif de France) a été renommé en 2011. La création du « CSB » (Comité Stratégique Bibliographique) a permis d'acter :

- la nécessité d'organiser la coordination bibliographique nationale au-delà du seul périmètre des catalogues collectifs dans un contexte de profond bouleversement des techniques et des usages, qui appelle au renforcement d'un pilotage concerté;

- la prise en compte du caractère de plus en plus stratégique des enjeux relatifs aux métadonnées et au traitement de l'information bibliographique, car les métadonnées commandent non seulement la recherche documentaire mais aussi les processus de numérisation, de gestion de droits, de conservation.

Lors de sa réunion du 15/02/2012, le CSB a commandé à l'ABES et à la BnF un document d'orientation et de prospective devant servir de cadre de discussion entre les deux établissements et leurs tutelles respectives dans le but d'élaborer un nouveau projet national dans ce domaine.

Les équipes de la BnF et de l'ABES ont rédigé conjointement ce document d'orientation dans les mois qui ont suivi et consulté le 13/09/2012 le groupe stratégique RDA dont les avis ont pu être intégrés dans le rapport avant présentation au CSB. La réunion visait à discuter des préconisations de ce rapport et à élaborer un plan d'action.

### **Relevé de décisions**

En préambule, le CSB remercie R. Bérard et G. Illien pour la qualité du document présenté, se félicitant que la BnF et l'ABES n'abordent pas la question du futur des catalogues en ordre dispersé.

Le CSB donne par ailleurs son accord pour la publication du rapport présenté, enrichi des conclusions de cette réunion.

#### **1°) Le Web sémantique : un horizon commun**

Le CSB adhère à la vision de long terme proposée par l'ABES et la BnF sur l'horizon du web sémantique pour les catalogues de demain et à la nécessité d'organiser la coopération bibliographique entre les deux établissements autour de ce nouveau paradigme.

Ce nouvel environnement technologique est à la fois adapté aux nouveaux usages numériques (gains de visibilité pour les bibliothèques) et aux contraintes budgétaires qui pèsent sur les établissements (gains de productivité) dans la mesure où il favorise davantage les échanges et la réutilisation de métadonnées, recentrant le travail de catalogage sur la valorisation de ressources locales et la création de liens.

Il permet aussi de s'affranchir des questions de compatibilité entre applications de production ou de diffusion spécifiques comme ce fut souvent le cas jusqu'à présent, pour penser la coopération et la mutualisation nationales au niveau des métadonnées elles-mêmes.

Cela implique de travailler à l'interopérabilité technique et juridique des données d'une part et à la complémentarité des collections signalées par les deux établissements d'autre part.

Le web de données constituant le point de rencontre des métadonnées produites par l'ABES et la BnF, leur objectif commun peut donc être formulé comme la co-maintenance d'un ensemble de métadonnées françaises publiques de référence qui devront demeurer ouvertes à une large réutilisation sur le Web.

Cet horizon est un chantier qui s'ouvre, mais en attendant sa mise en œuvre à grande échelle, le CSB souligne :

- que l'ABES comme la BnF devront continuer de servir leurs réseaux respectifs depuis les outils et formats existants, une assez longue période de transition devant nécessairement être envisagée. A moyen terme, les coopérations engagées (au travers du CCFR en particulier, et dans le cadre de la convention ABES-BnF de 2011 encadrant la transmission de notices par la BnF à l'ABES) restent donc pertinentes.
- que l'utilisateur doit rester au centre des préconisations. Pour les étudiants et les chercheurs en particulier, le CSB recommande qu'ils soient bien pris en compte et associés à cette orientation vers le web sémantique.

## **2°) Poursuite de la coopération en matière de normalisation et de modélisation**

L'interopérabilité technique implique l'adoption de standards communs et une coopération nationale autour de leur implémentation. Cette coopération est déjà bien engagée avec la coopération sur l'adoption de RDA en France, sous l'égide de l'AFNOR et dans le cadre du groupe technique et du groupe stratégique RDA. Les travaux du groupe technique et de ses différents sous-groupes ont été tout particulièrement salués. Ils ont de fait dépassé la seule problématique de l'adoption d'un nouveau code de catalogage pour amorcer des études et des initiatives utiles dans les domaines de l'évolution des systèmes d'information et de la formation professionnelle notamment.

La recherche de solutions compatibles avec les grandes orientations internationales qui se dessinent est confirmée afin d'éviter un isolement français. Cela ne doit toutefois pas se faire au prix de régressions majeures. La démarche engagée permet de veiller à ces deux objectifs. Le CSB rappelle que la fixation d'un objectif temporel (date prévisionnelle d'adoption de RDA en production) sera le principal signal véritablement déclencheur du changement à l'échelle nationale. A ce stade, ni l'ABES (engagée dans le projet de SIGB mutualisé) ni la BnF (qui n'a pas les moyens de changer son système ni ses pratiques de production bibliographique avant plusieurs années) ne sont en mesure d'arrêter une date précise. Leurs stratégies respectives privilégient le RDF comme format de sortie ou d'exposition des données, les formats de production étant encore peu impactés.

La poursuite de la collaboration entre BnF et ABES sur RDA (et son prolongement dans le cadre du groupe d'intérêt européen d'EURIG) est encouragée par le CSB, qui prend acte de son développement opérationnel en termes d'implémentation : l'ABES et la BnF s'engageront début 2013 dans un travail commun de modélisation pour une mise en œuvre concertée de la FRBRisation de leurs catalogues respectifs.

### **3°) Licences : des convergences et des divergences**

Les analyses juridiques et économiques conduites par l'ABES et la BnF dans le contexte grandissant du *Linked Open Data* les ont conduites aux mêmes conclusions : c'est la licence d'Etat préconisée en 2011 par l'organisme Etalab (désormais rattaché aux services du Premier ministre) qui semble la plus adaptée à une ouverture des données à la fois cohérente avec les principes du droit français et au maintien de l'obligation d'attribution, auquel le CSB confirme son attachement compte tenu de l'exigence de traçabilité des données produites par les pouvoirs publics.

La BnF a d'ores et déjà adopté cette licence pour les données au format RDF qu'elle diffuse dans le cadre du projet data.bnf.fr. L'utilisation de cette licence n'a pu toutefois être étendue aux autres formats, en particulier aux formats MARC diffusés en tant que produits bibliographiques, qui font l'objet d'une licence et d'une facturation en cas de réutilisation commerciale. Sur ce point, le ministère de la Culture et de la communication privilégie aujourd'hui une position conforme à l'exigence de générer des ressources propres et qui soit commune à tous les établissements culturels (Archives, musées, bibliothèques...).

L'ABES et sa tutelle défendent le choix de la licence d'Etat issue des travaux d'Etalab pour la totalité des données (qu'il s'agisse du RDF ou du MARC), choix qui sera prochainement présenté au conseil scientifique et au conseil d'administration de l'ABES. Toutefois l'ABES restera largement tributaire des décisions juridiques des réservoirs de données dont le Sudoc s'alimente, la part de production originale par le réseau Sudoc restant limitée en volume.

La question de l'interopérabilité juridique des licences adoptées par la BnF et l'ABES est donc centrale pour l'avenir de leur collaboration. En l'état, cette collaboration ne peut porter que sur les données en RDF. Pour les données en MARC, la BnF apportera des arguments scientifiques complémentaires à sa tutelle afin d'éviter un blocage de long terme qui constituerait un frein réel à la collaboration interministérielle dans ce domaine. De son côté, l'ABES, qui n'exclut pas l'acquisition d'un SIGB mutualisé reposant sur une base de connaissances commerciale, veillera à conserver la maîtrise des données du Sudoc qui devront demeurer ouvertes à une large réutilisation.

### **4°) Signalement de la documentation électronique : vers un partage des rôles**

Le partage des tâches préconisé dans le rapport reçoit le plein assentiment du CSB :

- à la BnF le signalement du patrimonial national né numérique en collaboration avec les éditeurs français (en utilisant le levier juridique du dépôt légal, qui va notamment s'étendre aux livres numériques),
- à l'ABES le signalement de la documentation électronique scientifique, notamment étrangère (en utilisant le levier des licences nationales et d'ISTEX).

La mise à disposition par la BnF des métadonnées issues des éditeurs français dans le cadre du dépôt légal numérique constituerait un apport utile pour l'ABES.

Par sa participation au dispositif des licences nationales, la BnF pourrait de son côté profiter de solutions de signalement des ressources électroniques organisées par l'ABES au niveau national.

Ce partage de principe ne couvre pas la totalité des besoins identifiés mais le CSB convient qu'il permet d'ores et déjà de poser des bases saines pour la mutualisation des futurs efforts de signalement de la documentation électronique.

## 5°) Formation, communication, évolution des SIGB et accompagnement au changement

Le CSB souligne la qualité des efforts initiés en matière de communication et de formation professionnelles, tant par la BnF et l'ABES que par les organismes (CNFPT et CRFCB en particulier) associés au groupe technique RDA, qui a su constituer un premier vivier de formateurs couvrant toutes les régions françaises

Il encourage la poursuite de ces efforts et la concertation entre acteurs de la formation et de l'information professionnelle en France afin de poursuivre le travail de sensibilisation engagé.

A cet égard, il recommande :

- de poursuivre le programme de sensibilisation engagé avec les organismes de formation initiale et permanente ;
- de veiller à aborder régulièrement ces sujets lors des congrès professionnels ;
- d'associer davantage les associations professionnelles ;
- d'associer davantage les clubs utilisateurs de SIGB et les éditeurs de logiciels de bibliothèque de même que des entreprises nouvelles spécialistes du web sémantique.

Le CSB reconnaît toutefois que l'exposé des changements envisagés au niveau national se heurte aux questions qui mobilisent légitimement tant les établissements que les fournisseurs de logiciels : il est difficile de répondre concrètement aux questions impactant les réinformatisations ou la réorganisation du travail des catalogueurs tant qu'on ne disposera pas d'un planning prévisionnel sur l'adoption de nouveaux formats de production et l'arrêt de la production au format MARC.

### Calendrier et cadres de travail

A l'issue de la première analyse conduite par l'ABES et la BnF en 2012, les deux établissements approfondiront en 2013 les pistes validées par le CSB afin d'aboutir à un plan d'action qui pourra être formalisé dans le cadre de la convention ABES-BnF qui doit être renouvelée début 2014.

Les axes retenus sont les suivants :

- **interopérabilité technique** : poursuite du travail sur RDA, modélisation RDF et concertation autour de la FRBRisation des catalogues, identifiants (ISSN, ISNI...), convergence des fichiers et référentiels d'autorité. L'objectif général est de dessiner les contours et les conditions d'interopérabilité technique du futur ensemble de métadonnées publiques de référence dont la BnF et l'ABES auront à assurer la co-maintenance dans le futur. Ce travail sera conduit dans le cadre des réunions opérationnelles ABES-BnF, dont la prochaine réunion est programmée le 11/01/2013 à la BnF. Cette collaboration continuera dans un cadre international, en lien avec la participation conjointe à EURIG, à VIAF, à l'IFLA et des discussions conduites de part et d'autre avec OCLC. Elle pourra également trouver écho au niveau national dans le prolongement des réunions des groupes techniques et stratégique RDA (dont le champ pourrait être élargi au-delà de la seule question du code de catalogage) et à travers la participation conjointe de la BnF et de l'ABES aux travaux de la Bibliothèque scientifique numérique (groupe BSN3 sur le signalement en particulier).

- **interopérabilité juridique** : ce dossier, plus politique, sera suivi directement avec les tutelles. Suite aux consultations du CS et du CA de l'ABES, il reviendra à la BnF de revenir en 2013 vers sa propre tutelle afin de poursuivre la discussion sur les licences pour les données au format MARC. L'interopérabilité juridique des données au format RDF devrait être rapidement effective compte tenu des décisions déjà prises par la BnF et des orientations retenues par l'ABES.

- **récupération de métadonnées issues des éditeurs** : ce dossier pourra être suivi au rythme d'avancement des instructions pilotées respectivement par l'ABES (chantier des licences nationales dans le cadre de la BSN et d'ISTEX) et la BnF (développement du dépôt légal numérique et de l'Extranet du dépôt légal pour les livres électroniques).

- **concertation et sensibilisation, étude d'impact** : l'ABES, à travers son projet de SIGB mutualisé, dispose d'ores et déjà d'un cadre de concertation pour mesurer l'impact de ses projets de changement de système au sein du réseau de l'enseignement supérieur et de la recherche. Pour les établissements qui ne relèvent pas de ce périmètre, une étude d'impact relevant plutôt du ministère de la Culture et de la communication pourrait être lancée en complément des études déjà conduites en 2012 par la BnF et le groupe technique RDA sur la récupération de notices d'une part et l'informatisation des bibliothèques d'autre part. La BnF et sa tutelle étudieront ensemble les conditions de réalisation d'une étude d'impact s'intéressant plus particulièrement à l'évolution des systèmes des établissements concernés en 2013 ou 2014.