

Boîte à outils du numérique en bibliothèque

Fiche pratique n°1

La récupération des données pour le catalogue

Groupe de travail *Boîte à Outils du Numérique en bibliothèque*

Pascal Allard (DRAC Nord-Pas-de-Calais)

Anne Balavoine (Association pour le Développement des Documents Numériques en Bibliothèques)

Catherine Dhérent (Bibliothèque nationale de France)

Philippe Gauchet (Médiathèque départementale du Pas-de-Calais)

Patricia Le Galèze (Ministère de la Culture et de la Communication / Direction générale des médias et des industries culturelles / Service du Livre et de la Lecture) – Coordinatrice du groupe de travail

Sophie Perrusson (Médiathèque de Levallois)

Experts invités :

Caroline Demessence (Bibliothèque nationale de France)

Françoise Leresche (Bibliothèque nationale de France)

Sommaire

<i>Introduction</i>	3
<i>1 Des normes et recommandations, piliers de la réussite</i>	4
1.1 Formats de structuration de données.....	4
1.2 Normes et protocoles d'échange de données.....	4
1.3 Encodage des caractères.....	5
1.4 Recommandations pour l'échange de données d'exemplaire en format UNIMARC.....	6
<i>2 Procéder à la récupération des données</i>	7
2.1 La structuration de la base de données.....	7
2.2 Quelles données, pourquoi et quand les récupérer ?.....	11
2.3 Comment préparer la récupération des données ?.....	11
2.4 Exemples de procédures de récupération conformes au format.....	14
<i>3 Choisir des réservoirs de notices</i>	19
3.1 Les critères de choix.....	19
3.2 Les différents fournisseurs de données.....	19
3.3 Le droit des données.....	20
<i>4 Perspectives d'évolution : vers l'interopérabilité et la visibilité sur le web</i>	21
4.1 Des solutions en cours.....	21
4.2 Un environnement en mutation.....	21
<i>Recommandations</i>	23
<i>Pour en savoir plus</i>	24

Introduction

Le catalogue – reflet des collections pour le public – constitue le fondement d'une bibliothèque de lecture publique. En revanche, tel n'est plus le cas aujourd'hui du catalogage, activité longtemps première des bibliothécaires. En effet, si celui-ci s'est normalisé avec notamment la définition de l'ISBD (International Standard Bibliographic Description), l'évolution technologique permet d'alléger considérablement cette tâche en récupérant les notices bibliographiques réalisées par des fournisseurs spécialisés.

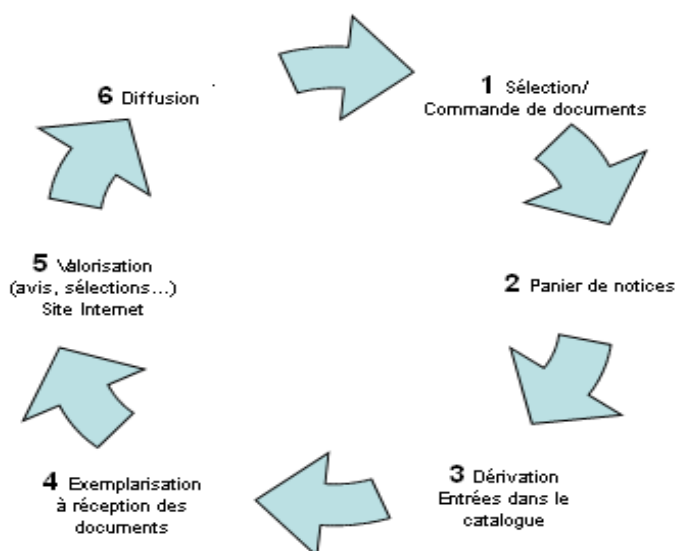
Des bibliothèques pionnières comme celle de Fresnes s'adonnent à cette pratique de dérivation de notices depuis la fin des années 1990 mais de trop nombreuses bibliothèques continuent encore de produire leurs propres notices.

Dérivation de notices pour la constitution du catalogue :

- en 2010 35,5 % des BM (taux moyen de notices entrées par dérivation 70,4 %)¹
- en 2008 52,1 % des BM (fonds constitués à partir de 65,3 % de notices dérivées)
- en 2007 44,7 % des BM (fonds constitués à partir de 60,4 % de notices dérivées)
- en 2007 97,7 % des BDP (fonds constitués à partir de 79,1 % de notices dérivées)

Source : Ministère de la Culture et de la Communication/Service du Livre et de la Lecture

Aujourd'hui les avantages que procurent le numérique et la normalisation dans la chaîne du traitement du document permettent aux bibliothécaires, en abandonnant le catalogage, de se centrer sur les acquisitions, la médiation et la valorisation auprès du public tant par des conseils de lecture que par une aide à la recherche documentaire et au repérage dans l'univers numérique. Cette mutation concourt à l'amélioration de cette chaîne : elle est riche d'impacts tant sur la cohérence d'un catalogue reposant sur une base bibliographique normalisée, interopérable et adaptée à l'environnement du web, que sur l'organisation du travail régie par davantage d'efficacité.



Mais déployer un processus de récupération de données requiert des connaissances préalables afin de l'organiser au mieux pour une application immédiatement performante et sous la forme la plus normalisée pour faciliter les évolutions attendues.

1 Pour les bibliothèques ayant répondu au questionnaire complet en 2010. Le mode de collecte et d'analyse des données pour le rapport statistique annuel des bibliothèques municipales a évolué en 2009. Se référer au site de l'Observatoire de la lecture publique pour davantage d'explication : http://www.observatoirelecturepublique.fr/observatoire_de_la_lecture_publique_web/FR/syntheses_annuelles.awp

1 Des normes et recommandations, piliers de la réussite

La connaissance des formats et des normes est indispensable pour être en mesure de négocier avec les fournisseurs de système de gestion de bibliothèques (SIGB). La présence d'une personne ressource experte au sein de la bibliothèque ou du réseau est spécialement requise.

1.1 Formats de structuration de données

La structuration des données grâce aux formats MARC (MACHINE Readable Cataloguing), et depuis quelques années au langage XML (acronyme pour *eXtensible Markup Language*) pour certains types de documents (manuscrits et documents d'archives), vise à faciliter les échanges.

UNIMARC : Format d'échange de l'information bibliographique

UNIMARC (acronyme pour UNiversal MARC) est le format développé par l'IFLA en 1977 pour permettre l'échange de l'information bibliographique informatisée et servir d'interface entre les formats MARC nationaux.

Il est maintenu au niveau international par le Comité permanent UNIMARC (Permanent UNIMARC Committee ou PUC) de l'IFLA. En France, c'est le format officiel d'échange de l'information bibliographique². Il est géré par le Comité français UNIMARC. Les formats UNIMARC existants s'appliquent aux notices bibliographiques et aux notices d'autorité. Le format UNIMARC des données locales n'est pas utilisé en France ; il est remplacé par des recommandations pour l'échange de données d'exemplaires (voir § 1.5).

UNIMARC est avant tout employé comme format de production dans la plupart des bibliothèques françaises.

XML : Extensible Markup Language³

XML est un langage informatique normalisé en 1998 par le W3C (World Wide Web Consortium⁴) qui permet de structurer toute information et de l'échanger sur le web. XML est à la fois un langage de description et d'échange de documents structurés. Il favorise d'une part le développement de formats permettant de rendre les données bibliographiques compatibles avec l'environnement du web. Et il facilite d'autre part le développement de protocoles d'échanges pour la fourniture ou l'échange de notices MARC comme les protocoles OAI-PMH ou SRU.

1.2 Normes et protocoles d'échange de données

Ces normes et protocoles définissent les modalités d'échange de l'information entre systèmes indépendants. Ils mettent en œuvre l'échange de données bibliographiques ou de métadonnées⁵.

La norme NF ISO 2709:2009 - Format pour l'échange d'information

Cette norme internationale définit une structure informatique générale pour l'échange de divers types de notices en MARC. Elle donne le cadre destiné spécifiquement aux échanges entre systèmes de traitement de l'information. Elle spécifie que chaque notice bibliographique préparée pour l'échange doit contenir :

2 Arrêté du 3 novembre 1993 relatif au format d'échange des données bibliographiques (J.O. n° 275 du 27 novembre 1993) ; <http://www.legifrance.gouv.fr/>

3 Pour des développements et des exemples sur XML voir la fiche 2 « Ré-informatisation et évolution des catalogues »

4 W3C, Organisme de normalisation des technologies du web

5 Voir chapitre 4

- un label de notice comportant 24 caractères ;
- un répertoire indiquant, pour chaque zone de données, son étiquette à 3 caractères, sa longueur et la position de son caractère de départ par rapport à la première zone de données ;
- des zones de données de longueur variable, séparées chacune par un séparateur de zones.

UNIMARC est une application particulière de la norme ISO 2709.

Le protocole Z39.50 (NF ISO 23950:1999 Définition du service de l'application et spécification du protocole et ANSI/NISO Z39.50)

Norme ISO depuis 1997, c'est un protocole utilisé pour la recherche et le transfert de notices (en ISO 2709). Il spécifie les modes d'interrogation de base de données bibliographiques, ainsi que la restitution des résultats. Il fonctionne selon une architecture client/serveur en mode synchrone : un poste client interroge en temps réel les bases bibliographiques et rapatrie depuis le serveur les résultats instantanément. Cette fonction est implantée dans la plupart des SIGB.

Le protocole FTP (File transfer protocol – Protocole de transfert de fichiers)

Il s'agit d'un protocole de communication destiné à l'échange informatique de fichiers entre ordinateurs. Il permet de se connecter à un serveur, d'en parcourir l'arborescence, d'en récupérer et télécharger les fichiers dans son SIGB.

Il faut disposer d' :

- un logiciel qui importe de l'ISO 2709 et gère les formats MARC ;
- une connexion Internet (protocole HTTP) avec un navigateur web (Mozilla Firefox ou Internet Explorer) ;
- un outil FTP (client FTP⁶ ou web FTP⁷) pour importer les notices depuis le serveur FTP du réservoir dans le catalogue.

Le protocole SRU (Search/Retrieve via URL)

Ce protocole propose les mêmes fonctionnalités que le protocole Z39.50 mais exprimées en XML. Il a été conçu comme une alternative au protocole Z39.50, en étant davantage adapté à l'univers du Web (protocole HTTP et XML). Développé à partir de 2002, il est maintenu par la Bibliothèque du Congrès.

Le protocole SRU est un outil performant, permettant d'interroger de façon synchrone le contenu des métadonnées d'une base. En revanche, il se heurte à nombreux inconvénients : l'instabilité (la dernière version en cours date de 2007), une documentation disponible faible, une communauté d'utilisateurs limitée et une certaine complexité. Même si des SIGB intègrent des connecteurs d'interrogation SRU, le nombre de serveurs interrogeables et stables ne dépasse pas la dizaine.

1.3 Encodage⁸ des caractères

Les caractères composant l'information doivent être encodés selon un système standardisé au niveau international qui permet de gérer tout type de caractère dans quelque écriture que ce soit. Il en existe plusieurs dont deux principaux utilisés par les fournisseurs de données en France.

ISO 5426 : c'est un jeu de caractères latins codés employés pour l'échange d'information bibliographique avec ou sans diacritiques.

⁶ Logiciel possédant une interface graphique ou en ligne de commande pour accéder à un serveur FTP

⁷ Il permet de se connecter sur un serveur FTP sans logiciel. Il s'installe en utilisant son navigateur

⁸ Manière de représenter les caractères (lettres, chiffres, symboles) dans un système informatique

UNICODE UTF-8 : c'est un jeu de caractères universels codés sur plusieurs octets, le codage étant sur 8 bits pour UTF8 ; norme à retenir si la bibliothèque gère des collections en caractères non latins.

1.4 Recommandations pour l'échange de données d'exemplaire en format UNIMARC

Ces recommandations ont été publiées en mai 1998, conjointement par l'Agence bibliographique de l'enseignement supérieur, la Bibliothèque nationale de France, la Direction du livre et de la lecture (Ministère de la culture) et la Sous-direction de la documentation et des bibliothèques (Ministère de l'éducation nationale, de la recherche et de la technologie – Direction de l'enseignement supérieur).

La partie 1 porte sur la Recommandation 995 pour la fourniture de données locales dans les échanges de notices bibliographiques en UNIMARC accompagnant le prêt ou le dépôt d'exemplaire. La dernière version (version 3) est disponible sur le site de l'ADBDP⁹.

Elle a pour objectif de faciliter les échanges de données liées au prêt ou au dépôt d'exemplaires entre établissements gérés par des systèmes informatiques différents, notamment dans les organismes dépositaires de documents des bibliothèques départementales de prêt (BDP) et dans les réseaux intercommunaux de bibliothèques.

La partie 2¹⁰ porte sur la description des données d'exemplaire pour l'échange d'information bibliographique en format UNIMARC hors BDP. Elles définissent une structure pour la description des données d'exemplaire associées aux notices bibliographiques échangées.

9 http://www.adbdp.asso.fr/IMG/pdf_r995.pdf

10 <http://www.abes.fr/Media/Fichiers/Sudoc-Fichiers/Produire-dans-le-Sudoc/Recommandations-pour-l-echange-de-donnees-d-exemplaire-en-format-UNIMARC>

2 Procéder à la récupération des données

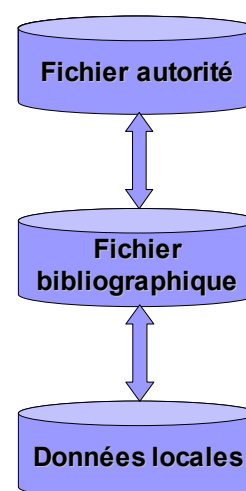
Pour être accessibles, interopérables dans des réseaux documentaires, réutilisables et adaptables, il importe que les données soient bien structurées conformément aux normes et formats de catalogage. **Il est impératif de ne pas développer de pratiques « maison ».** **La capacité de la bibliothèque à se projeter dans l'environnement du web en dépend.**

2.1 La structuration de la base de données

L'organisation de la base de données

Les catalogues actuels sont structurés en fichiers, composés de notices. On distingue trois types de fichiers :

- celui des **notices d'autorité** : notice établie pour un point d'accès (matière, personne, ...) sous une forme autorisée par les normes nationales. Selon les cas, elle peut contenir des informations d'identification, rappel des formes non retenues, associées, sources consultées ;
- celui des **notices bibliographiques** : ensemble des éléments comprenant la description bibliographique d'un document et des points d'accès (matière, personne, ...) à celle-ci ;
- celui des **données locales** : particularités de l'exemplaire (éléments d'identification tels que la cote, le code-barre ou numéro d'inventaire, la disponibilité...).



Les notices d'autorité (auteur, matière, titre uniforme...) sont liées uniquement aux notices bibliographiques.

Les données locales sont liées uniquement aux notices bibliographiques.

Il est indispensable de différencier ce qui relève, d'un côté, des données bibliographiques, et de l'autre, des données locales.

Les données locales ne doivent pas être insérées dans les zones de la notice bibliographique. Ainsi la cote, donnée locale, ne doit surtout pas se trouver dans une zone de la notice bibliographique. Si cette distinction n'est pas faite, il est impossible de faire évoluer séparément les deux types de données, par exemple d'écraser les éléments d'une notice de gestion provisoire par ceux d'une notice de référence, puisque par la même opération, on écraserait aussi les données locales déjà enregistrées.

Les notices bibliographiques et d'autorité sont récupérées automatiquement via des réservoirs existants. Seules les données locales sont saisies par la bibliothèque. Une attention particulière doit être portée à la récupération des notices d'autorité. Créer et maintenir des autorités demande à la fois du temps et des compétences. Il est donc préférable de les récupérer auprès d'un réservoir institutionnel. Toutefois, tous les SIGB ne permettent pas la gestion des liens entre les points d'accès et les notices d'autorité : un élément à prendre en compte lors du choix de celui-ci. Il est à noter que sans notices d'autorité, les possibilités de faire évoluer le catalogue de la bibliothèque seront réduites (Cf. Fiche 2).

Autant la récupération des données doit être exhaustive (sans suppression de zones), autant l'affichage doit être adapté aux utilisateurs (publics/professionnels). Ce dernier se paramètre dans le SIGB.

Les zones obligatoires

Les zones obligatoires pour la constitution d'une **notice bibliographique minimale** en UNIMARC¹¹, sont les suivantes¹² :

Zone	Libellé	Répétabilité	Application
100	Label de la notice	Non répétable	Zone permettant le traitement informatique de chaque notice.
001	Identifiant de la notice	Non répétable	Zone contenant des caractères attribués de façon unique. Cet identifiant est un numéro de contrôle qui permet d'identifier chaque notice et les éventuels doublons.
100 \$a	Données générales de traitement	Non répétable	Zone contenant la date de création de la notice (<i>obligatoire</i>), des informations générales sur le document et les jeux de caractères utilisés pour l'encodage (<i>données obligatoires</i> car nécessaires pour l'échange et le traitement des données).
200 \$a	Titre du document	Non répétable	Zone contenant le Titre propre (\$a), seule sous-zone obligatoire dans cette zone. Elle est exigée par le système pour charger les notices.
801	Origine de la notice	Répétable	Une zone par institution intervenant sur la notice.

Ces zones sont complétées par d'autres zones obligatoires dans certains types de notices :

Notices de ressources textuelles

101 Langue de la ressource

Notices de ressources cartographiques

120 Zone de données codées : Ressources cartographiques – Généralités

123 Zone de données codées : Ressources cartographiques – Échelle et coordonnées

206 Zone particulière à certains types de ressources : Ressources cartographiques – Données mathématiques

Notices de ressources électroniques

230 Zone particulière à certains types de ressources : Ressources électroniques

304 Note sur le titre et la mention de responsabilité

Notices de ressources musicales

125 Zone de données codées : Enregistrements sonores et musique notée

La zone adéquate pour la localisation et la cote est la zone 930 "Localisation et cote" (Cf. Les Recommandations pour l'échange de données d'exemplaire en format UNIMARC, partie 2)¹³. Les zones 9XX sont définies dans les présentes recommandations pour la structure des informations de localisation, de gestion des exemplaires.

11 http://www.bnf.fr/documents/section3_6_2011.pdf et <http://archive.ifta.org/VI/3/p1996-1/guid4.htm>

12 Contient en outre automatiquement un répertoire : voir http://multimedia.bnf.fr/unimarc_b_trad/BSection3-6-2011.pdf

13 <http://www.abes.fr/content/download/437/2470/version/1/file/Recommandations>

Les zones obligatoires pour la constitution d'une **notice d'autorité minimale** en UNIMARC¹⁴, sont les suivantes :

Zone	Libellé	Répétabilité	Application
000	Label de la notice	Non répétable	Zone permettant le traitement informatique de chaque notice.
001	Identifiant de la notice	Non répétable	Zone contenant des caractères attribués de façon unique. Cet identifiant est un numéro de contrôle qui permet d'identifier chaque notice et les éventuels doublons.
100 \$a	Données générales de traitement	Non répétable	Zone contenant la date de création de la notice (<i>obligatoire</i>), des informations générales sur le document comme la langue (<i>obligatoire</i>) et les jeux de caractères utilisés pour l'encodage (<i>données obligatoires</i> car nécessaires pour l'échange et le traitement des données).
152	Règles	Non répétable	Zones indiquant les règles suivies pour établir la vedette figurant dans la zone 2XX et les formes rejetées ou associées qui l'accompagnent.
2XX	Bloc de la vedette	Non répétable	Ce bloc contient la vedette pour laquelle on a créé la notice. Cette vedette sera une forme d'autorité dans une notice d'autorité.
801	Origine de la notice	Répétable	Une zone par institution intervenant sur la notice.

Lors de dérivation de notices, les zones indispensables

Lors de l'échange de notices bibliographiques ou d'autorité, la zone 035 (Identifiant de la notice dans un autre système) se doit d'être présente dans les notices dérivées. Cette zone joue en effet un rôle primordial dans la traçabilité en cas d'évolutions et de mises à jour des notices dérivées. Elle est alimentée à partir du contenu de la zone 001 des notices bibliographiques ou d'autorité dérivées.

De même, les liens entre les différents types de notices doivent être préserver lors d'une récupération. Ces liens se caractérisent au travers de sous-zones numériques présentes dans certaines zones des notices dérivées (sous-zones numériques telles que \$3 - identifiant de la notice d'autorité, \$0 - identifiant de la notice bibliographique, ...). Ces sous-zones numériques permettent le rebond vers des notices d'autorités, bibliographiques liées à la notice dérivée.

Depuis quelques années, la BnF ajoute à chaque notice un **identifiant pérenne de type ARK (zone 003)** dans le cadre de l'évolutivité vers le web¹⁵. **Il est recommandé de le conserver dès à présent et pour cela de le demander au fournisseur. Les filtres d'importation doivent être vérifiés et adaptés en conséquence.**

¹⁴ <http://www.bnf.fr/documents/UnimarcA.pdf> et <http://archive.ifla.org/VI/3/p1996-2/mlar.htm>

¹⁵ Voir fiche 2 Ré-informatisation et évolution des catalogues

Exemple de notice bibliographique avec notice d'autorité liée

Type : texte imprimé, monographie
Auteur(s) : [Laubier, Guillaume de](#) . Illustrateur
[Cazelles, Diane](#)
Titre(s) : Made in banlieues [Texte imprimé] : voyages inattendus aux portes de Paris / photographies, Guillaume de Laubier ; textes, Diane Cazelles
Publication : Paris : La Martinière, impr. 2010
Impression : 42-Saint-Étienne : Impr. Loire Offset
Description matérielle : 1 vol. (314 p.) : ill. en noir et en coul., couv. ill. en coul. ; 28 cm
Note(s) : Contient un entretien avec Manu Dibango. – Bibliogr., 1 p. Index
Sujet(s) : [Curiosités et merveilles](#) -- France -- [Île-de-France \(France\)](#) -- [Île-de-France \(France\)](#) -- [Guides touristiques et de visite](#)
Indice(s) Dewey : [914.436 04](#) (22e éd.)
ISBN 978-2-7324-4080-4 (br.) : 35 EUR
EAN 9782732440804
Notice n° : FRBNF42178940
<http://catalogue.bnf.fr/ark:/12148/cb421789401/PUBLI>

```

000 cx i22 45
001FRBNF119361791
00519991018
009http://catalogue.bnf.fr/ark:/12148/cb11936179g
039 $oOPL$a00157878X
100 $a19810226afrey50 ba0
106 $a211
152 $bRAMEAU$c2
250 $7ba0yba0y$8fre $9 2$aCuriosités et merveilles
305| $aVoir aussi la subdivision de forme Miscellanées à tous sujets noms communs et noms propres
450 $7ba0yba0y$8fre $9 $aCuriosités
450 $7ba0yba0y$8fre $9 $aÉnigmes
450 $7ba0yba0y$8fre $9 $aÉnigmes scientifiques
450 $7ba0yba0y$8fre $9 $aMerveilles
450 $7ba0yba0y$8fre $9 $aMystères et merveilles
450 $7ba0yba0y$8fre $9 $aSciences$xEénigmes
801 $aFR$bFR-751131015$c19991018
822 $aCuriosities and wonders$vLCSH (en ligne), 1999-10-18
http://catalogue.bnf.fr/ark:/12148/cb11936179g/UNIMARC

```

Identifiant de la notice bibliographique → <http://catalogue.bnf.fr/ark:/12148/cb421789401/PUBLI>

Identifiant ARK pérenne de la notice bibliographique et d'autorité → <http://catalogue.bnf.fr/ark:/12148/cb421789401>

Identifiant d'une notice d'autorité Rameau → <http://catalogue.bnf.fr/ark:/12148/cb11936179g>

Source de catalogue
Indicateur = 0 : catalogue original
\$a pays : France;
\$b agence : code de la BnF;
\$c date ; \$g règles utilisées : norme AFNOR ;
\$h identifiant d'origine ; \$2 code du format utilisé

Notice en affichage public du catalogue général de la BnF

Notice UNIMARC du catalogue général de la BnF

Notice d'autorité Rameau en UNIMARC

2.2 Quelles données, pourquoi et quand les récupérer ?

Avant d'adopter le principe de dérivation des notices au quotidien, il peut s'avérer utile – une fois réalisé le diagnostic reposant sur la rigueur normative – de procéder à un écrasement de la base initiale. On peut donc distinguer deux cas de récupération : l'un en bloc et rétrospectif, l'autre partiel au fil des acquisitions.

Types de notices	Quand	Pourquoi	Comment	Moyens
Notices bibliographiques	Mise à niveau d'une base existante	Repartir sur de bonnes bases : nettoyage	Par des produits rétrospectifs	- FTP - CD-ROM
Notices bibliographiques	Dérivation au fil des acquisitions	- Garantir la qualité de la base - Gagner du temps	Fourniture via des réservoirs	- FTP - Z 39-50 - Automate de « moissonnage »
Notices d'autorité	Mise à niveau d'une base existante	- Gérer des points d'accès normalisés - Évoluer vers le web de données - Changer d'indexation matière	Par des produits rétrospectifs	- FTP - CD-ROM
Notices d'autorité	Dérivation de notices bibliographiques	- Gérer des points d'accès normalisés - Évoluer vers le web	Fourniture via des réservoirs	- FTP

2.3 Comment préparer la récupération des données ?

Le diagnostic : se poser les bonnes questions

Avant tout, il convient de réaliser un diagnostic de l'état de la base. Celui-ci conduira à un nettoyage plus ou moins important. Le diagnostic consiste à explorer les pratiques de catalogage qui ont pu évoluer au fil du temps dans la bibliothèque. Il vise à vérifier la présence et la bonne utilisation des zones obligatoires.

Questions :

- Les zones obligatoires pour chaque type de support (imprimés, documents audiovisuels...) sont-elles présentes ?
- Les zones sont-elles utilisées et/ou modifiées selon les normes ?
- Les zones ont-elles été alimentées par saisie interne ou récupération ?
- Y a-t-il une gestion des notices d'autorité ou l'indexation ne fournit-elle que de simples points d'accès ? Les notices d'autorité sont-elles toutes liées aux notices bibliographiques ?
- A-t-on bien gardé les identifiants des notices de référence (zone 035 d'UNIMARC) ?
- Comment les notices d'autorité ont-elles été constituées (par saisie interne ou à partir de réservoirs) ?

Les opérations à mener

Si les notices ont été retravaillées, il convient d'identifier et de lister avec précision tous les champs qui ont pu faire l'objet de modifications ou d'interprétations et d'évaluer la volumétrie concernée ; ce qui permettra de décider :

- soit d'effectuer manuellement les corrections si le nombre de documents concernés est peu important ;
- soit de signaler les cas pour une prise en compte corrective lors du transfert des données vers la nouvelle base.

La reprise à partir de produits rétrospectifs pourra être demandée dans le cahier des charges de consultation des entreprises pour la fourniture du SIGB.

Les conditions d'une bonne récupération

Lors de la récupération, il est impératif de respecter les règles suivantes :

- Conserver l'identifiant d'origine (zone 001) de la notice source dans la zone 035 d'UNIMARC de la notice cible, facteur de traçabilité, indispensable pour les évolutions et les mises à jour ;
- Conserver les liens entre tous les types de notices (sous-zones numériques telles que \$3 - identifiant de la notice d'autorité, \$0 - identifiant de la notice bibliographique, ...) :
 - Entre les notices bibliographiques elles-mêmes,
 - Entre les notices d'autorité elles-mêmes (important pour les renvois de RAMEAU¹⁶ notamment),
 - Entre les notices bibliographiques et les notices d'autorité ;
- Lorsque la bibliothèque utilise des notices de gestion¹⁷, effectuer un dédoublement par l'identifiant de la notice de gestion (zone 035) pour permettre le remplacement par la notice de référence¹⁸, l'ISBN étant peu fiable (un même ISBN peut exister pour des ouvrages différents) ;
- Élaborer des filtres pour conserver les éléments d'enrichissement de la notice (résumés, vignettes numériques, extraits sonores et vidéos...) et les données locales.

Les deux modes de récupération

- **Mode manuel** (à l'unité ou par panier, par FTP ou Z39.50) : ce mode de récupération nécessite plusieurs allers retours entre le ou les sites extérieurs et le SIGB (interrogation des sites, constitution de paniers, récupération du panier et intégration dans le SIGB) :
 - La récupération par FTP n'est pas la plus pratique, car elle induit plusieurs opérations successives. Par contre, ce mode de récupération sera à privilégier pour les notices d'autorité, la récupération par le protocole Z39-50 n'étant pas permise pour ce type de notice chez certains fournisseurs.

Il est préférable d'envisager une récupération globale pour les autorités RAMEAU, afin de récupérer tous les liens (termes spécifiques, génériques associés à l'autorité

16 RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) est le langage d'indexation matière utilisé, en France, par la Bibliothèque nationale de France, les bibliothèques universitaires, ainsi que de nombreuses bibliothèques de lecture publique ou de recherche, d'organismes privés.

17 Notice minimale temporaire servant à la commande des acquisitions

18 Notice normalisée contenant les accès

RAMEAU), et une récupération au coup par coup pour les autres autorités (Personnes, Collectivités, Titres) en fonction des besoins.

- La récupération par Z39.50 présente l'avantage de rester dans son environnement de travail et de limiter les manipulations, d'où un travail plus rapide. Elle nécessite cependant que le SIGB dispose lui-même d'un client Z39.50 configuré pour se connecter au serveur Z39.50 du fournisseur de données et qu'il utilise le profil défini par celui-ci¹⁹. Même si cette technologie peut sembler dépassée, le protocole Z39.50 se caractérise par sa fiabilité et ses performances.
- **Mode automatique** (gestion depuis le SIGB par un automate de moissonnage et dérivation automatique dans le SIGB) :
 - Les outils automatiques de récupération présentent d'excellentes performances mais ne sont disponibles actuellement que dans quelques SIGB et nécessitent parfois de surmonter les réticences des services informatiques locaux pour l'ouverture de ports précis. Cette opération implique un paramétrage fin.
 - L'automate de moissonnage interroge le (ou les) site sélectionné d'un fournisseur de données et dérive les notices de gestion correspondantes aux critères de la requête. Ensuite, à un moment T paramétré, l'automate va de nouveau interroger le site du fournisseur à partir de ces notices de gestion, puis intégrer les mises à jour dans le catalogue. La notice de gestion est complétée par l'ajout de zones et de liens, elle devient ainsi une notice de référence.

¹⁹ Ces profils sont disponibles auprès des fournisseurs. Exemple sur le site de la BnF :
(http://www.bnf.fr/fr/professionnels/produits_et_services_bibliographiques/a.profil_z3950_bnf.html)

2.4 Exemples de procédures de récupération conformes au format

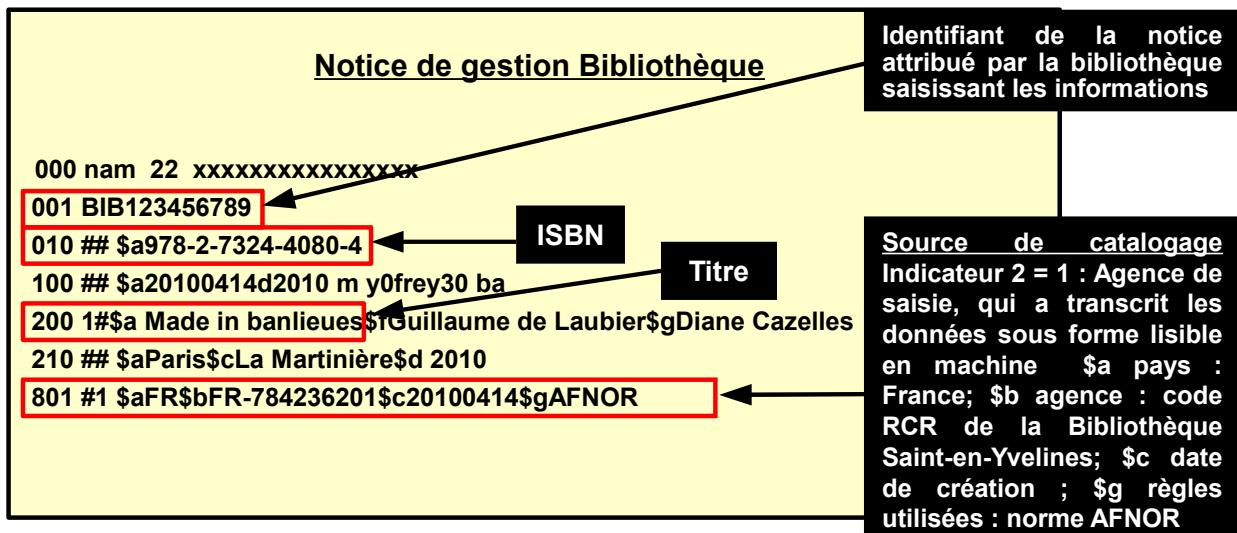
EXEMPLE 1 : Notice de gestion créée par la bibliothèque puis écrasée par la notice de référence provenant d'un fournisseur, ici la BnF

Dans ce cas-ci, le dédoublonnage ne peut se faire, à cette étape, que par la zone 010 (ISBN) ou bien la zone 200 (Titre), la zone 035 n'existant pas.

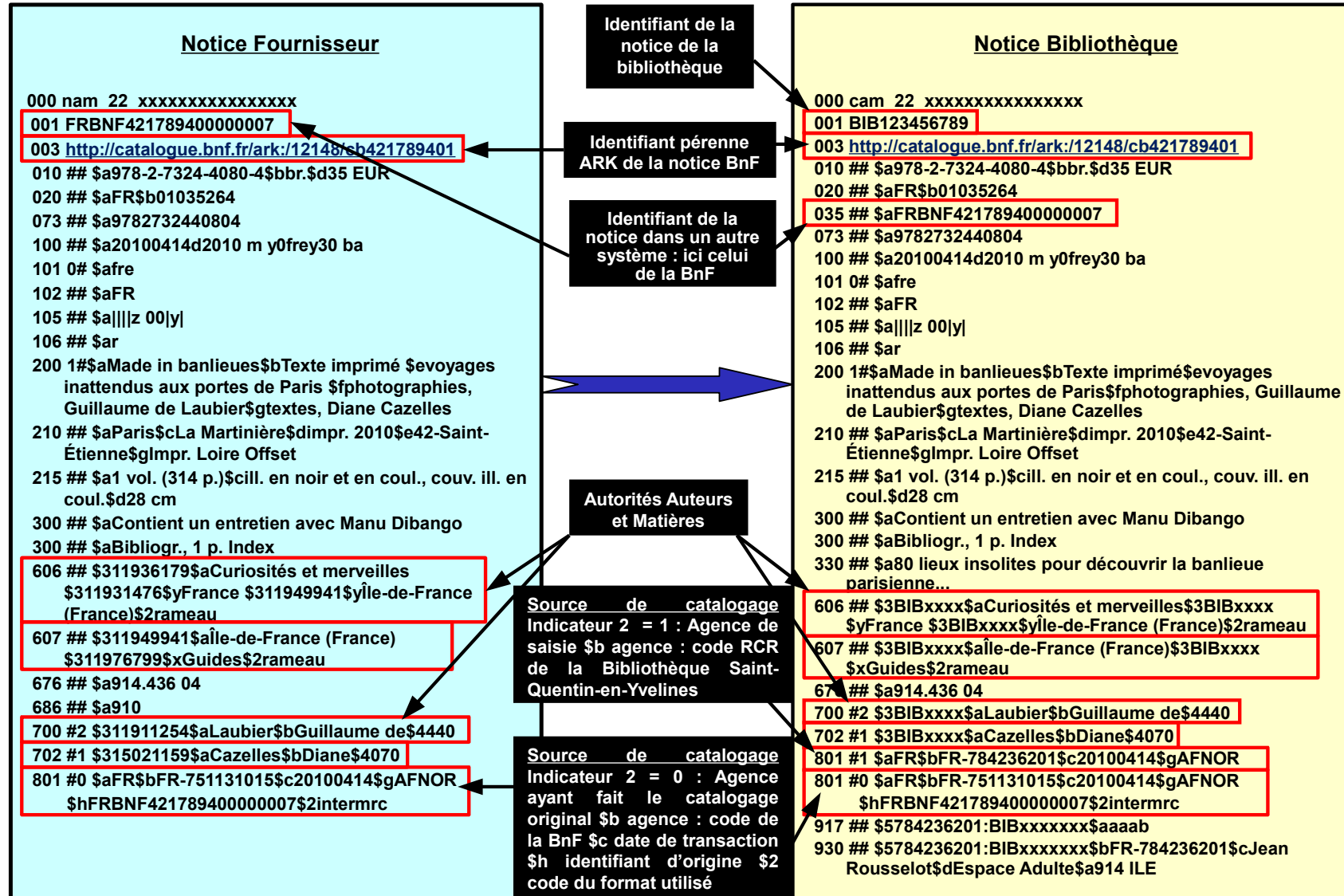
Le processus de récupération est le suivant :

1. Création de la notice de gestion avec saisie des données bibliographiques minimales par la bibliothèque.
2. Dérivation de la notice de référence depuis la base du fournisseur de données et écrasement de la notice de gestion via la zone 010 ISBN (ou bien la zone 200 Titre), seuls critères de dédoublonnage disponibles.
3. Implémentation des zones 035 (Identifiant de la notice dans un autre système) et 801 (Source de catalogage).
4. Ajout des zones propres à la bibliothèque qui font l'objet d'un filtre : zone 330 Résumé, zones 930 Localisation et cote, zones 917 Communicabilité
5. Ensuite les éventuelles mises à jour de la notice s'effectuent via la zone 035 et les sous-zones numériques.

1^{ère} étape : Création de la notice de gestion



2^{ème} étape : Récupération de la notice de référence depuis le catalogue général de la BnF



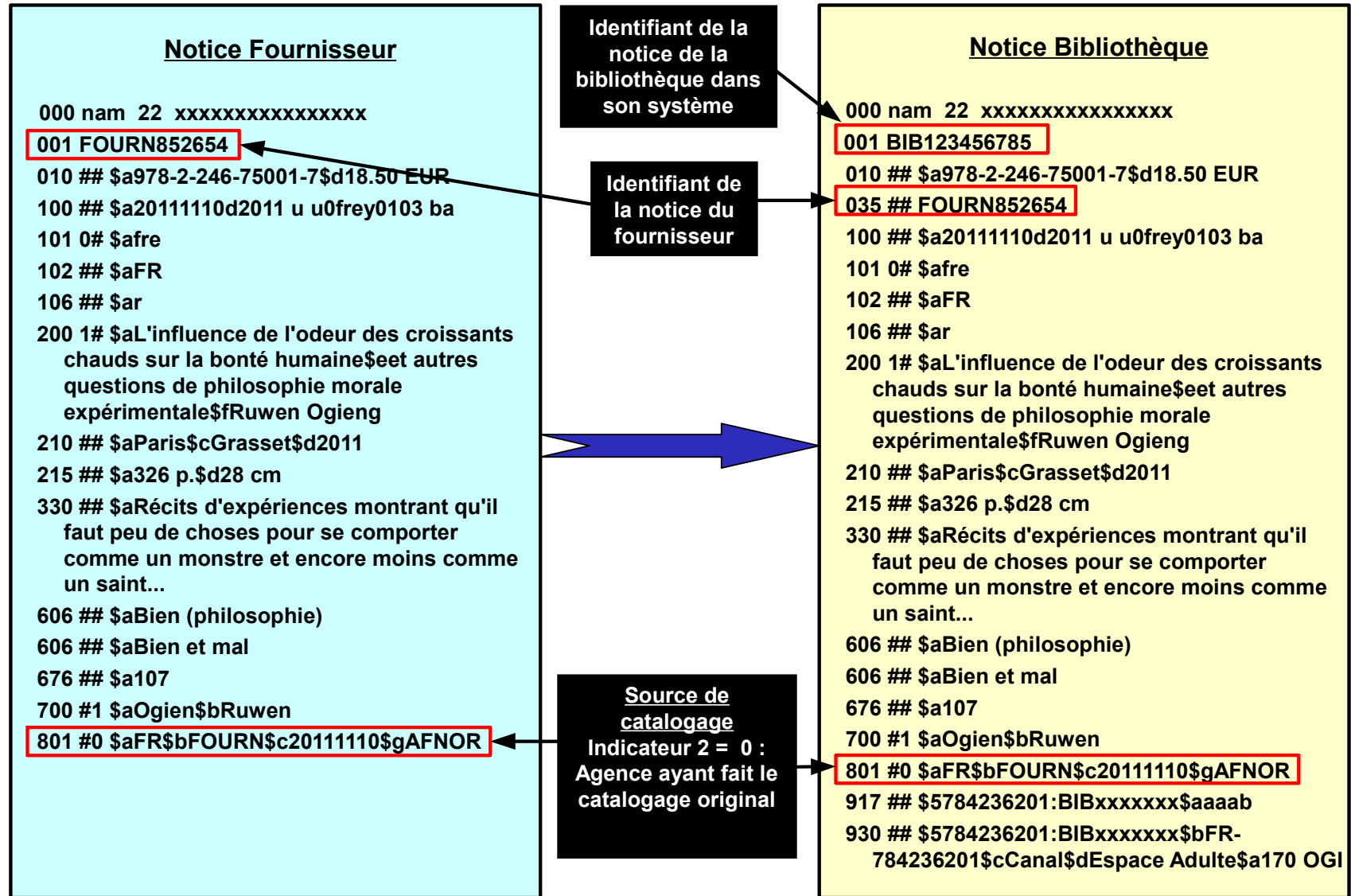
EXEMPLE 2 : Notice récupérée directement chez un fournisseur

La bibliothèque récupère directement ses notices chez un fournisseur, y compris le résumé. Certains fournisseurs complètent leurs notices avec des zones de la BnF (zones d'autorité-matière, auteurs...). Les notices de la bibliothèque sont ensuite enrichies de façon automatique et transparente par ces ajouts.

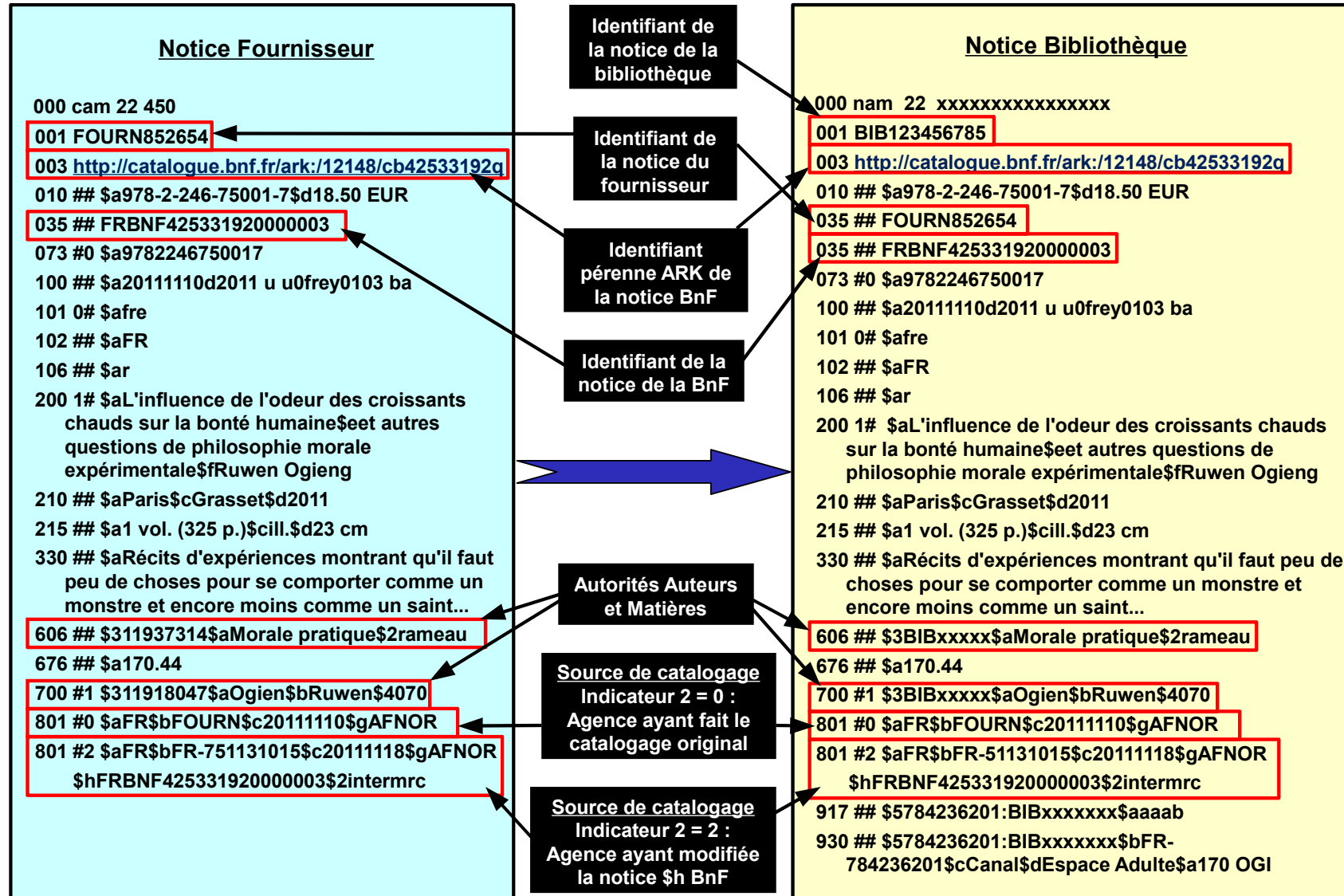
Le processus de récupération est le suivant :

1. Pas de création de la notice par la bibliothèque.
2. Dérivation depuis la base du fournisseur de données de la notice avec son résumé.
3. Implémentation des zones 035 (Identifiant de la notice dans un autre système) et 801 (Source de catalogage).
4. La zone 001 doit contenir le numéro d'identification de la notice attribué par la bibliothèque dans son système. Dans le cas où l'identifiant de la notice dans un autre système est utilisé comme identifiant de la notice en zone 001, il doit néanmoins être saisi en zone 035.
5. Ajout des zones propres à la bibliothèque qui font l'objet d'un filtre : zones 930 Localisation et cote, zones 917 Communicabilité.
6. Ensuite les éventuelles mises à jour de la notice (ici les données d'autorités) s'effectuent via la zone 035.

1^{ère} étape : Notice récupérée directement chez un fournisseur



2^{ème} étape : Enrichissement de la notice du fournisseur par l'ajout des autorités et récupérée automatiquement par la bibliothèque



3 Choisir des réservoirs de notices

Au réservoir institutionnel émanant de la BnF, composé de notices de tous types de supports, s'ajoutent ceux constitués par d'autres prestataires. Ceux-ci sont en général dédiés à un support unique, et sont payants, sauf exception : les bibliothèques s'y abonnent moyennant un coût annuel reposant sur des critères variés, sauf lorsque la livraison des notices est comprise dans le marché de fourniture de documents.

La stratégie de la bibliothèque, en termes de politique documentaire, de délai de mise à disposition des documents et de valorisation de ses collections via un site internet, guide le choix du/des fournisseur(s) et ce en tenant compte du budget qu'elle peut consacrer.

3.1 Les critères de choix

- La couverture du fournisseur (type de documents, langue,...) ;
- La disponibilité par rapport à la production éditoriale ou « fraîcheur » des notices ;
- L'enrichissement des notices (résumé, table des matières, feuilletage, vignette, extrait sonore ou vidéo) et leur qualité (normalisation, indexation) ;
- Le modèle économique : coût, autorisation de téléchargement de notices limité/illimité, nombre d'accès simultanés limités/illimités ;
- Le mode de récupération ;
- Import de notices autorité en lien avec le SIGB utilisé ;
- Mise à jour par l'automate de récupération, dit « moissonneur » ou « vendangeur » selon les SIGB.

Pour les livres, il est préférable de dériver les notices d'un fournisseur unique afin de proposer un catalogue le plus homogène possible et d'éviter l'abondance des procédures.

Il est à noter que certains prestataires indiquent que leurs notices sont ensuite écrasées par celle de la BnF – avec maintien toutefois de l'ajout de leurs spécificités – et ce de manière transparente pour la bibliothèque. Cette démarche est à privilégier.

Pour couvrir l'intégralité des besoins de notices catalographiques, la bibliothèque devra utiliser plusieurs réservoirs selon les types de supports qu'elle acquiert. Par exemple : BnF, Electre, Zébris, Notis ou Moccam... pour les livres, GAM pour les CD... et Colaco pour les DVD..., CVS pour l'un ou l'autre des supports audiovisuels... Ceci est parfaitement gérable à partir du moment où les notices des fournisseurs répondent aux normes et que les procédures de dérivation sont bien identifiées.

3.2 Les différents fournisseurs de données

La liste des différents fournisseurs de données bibliographiques et d'autorité est fournie en annexe à cette fiche et est disponible sur le site de l'ADBDP²⁰.

Elle peut être complétée par les réponses des fournisseurs au questionnaire du Comité français UNIMARC : http://cfu.enssib.fr/index.php?liste_publication=1 (partie 4 - Fournisseurs de notices en UNIMARC).

²⁰ <http://www.adbdp.asso.fr/spip.php?article822>

3.3 Le droit des données²¹

La question du droit des données se pose dès lors que des données sont mises à disposition et peuvent être récupérées. Dans ce cadre, la propriété des données répond ainsi à trois notions de droit.

Le droit d'auteur

Le droit d'auteur s'applique à une base de données dès lors que celui qui a pris l'initiative de la constituer démontre :

- un choix original des données et/ou
- une structure originale de la base.

Les résumés, les illustrations originales, les extraits de bandes annonces de film, etc. sont aussi soumis au droit d'auteur.

Le droit du producteur d'une base de données

Le producteur d'une base de données bénéficie d'une protection du contenu de la base qu'il a constitué. A ce titre, il définit les conditions d'extraction et de réutilisation de sa base. Ces droits peuvent être transmis ou cédés ou faire l'objet d'une licence.

Le droit de réutilisation des informations publiques

La réutilisation des données publiques est réglementée par l'ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques complétant la loi sur l'accès aux données de 1978. Elle stipule que :

- la réutilisation s'applique aux données déjà diffusées par l'organisme public ;
- les informations publiques ne doivent pas subir d'altération ou de dénaturation de leur sens ;
- la source et la date de dernière mise à jour doivent être mentionnées lors de la réutilisation des informations publiques.

La réutilisation peut donner lieu au paiement d'une redevance, ce qui occasionne la délivrance d'une licence fixant les conditions de la réutilisation des informations publiques ;

Des dérogations sont accordées aux établissements d'enseignement et de recherche ainsi qu'aux établissements culturels qui sont libres de définir les modalités de réutilisation et d'appliquer ou non une redevance.

Les bibliothèques ne sont pas toujours propriétaires de la notice. Dans le cas d'une récupération, la source de provenance de la notice doit toujours être conservée et mentionnée dans la notice (zone 801), de même que l'identifiant d'origine en zone 035 UNIMARC.

Il importe donc de prêter attention aux clauses juridiques du contrat de la bibliothèque avec le fournisseur de données. Il convient également d'être vigilant lors d'un changement de fournisseur à cette question du droit des données récupérées, quant à la perte éventuelle de services tels que les résumés, vignettes...

²¹ Ce droit relève des articles L112-1 et suivants, L341-1 et suivants du code la propriété intellectuelle

4 Perspectives d'évolution : vers l'interopérabilité et la visibilité sur le web

Le web est devenu un espace incontournable mais où, à l'heure actuelle, les données des catalogues des bibliothèques ne sont pas exploitées par les moteurs de recherche. Si les formats MARC proposent une information bibliographique très complète et très structurée, ils représentent des standards anciens, limités (hiérarchie, liens) et propres aux bibliothèques. Ils constituent donc un facteur d'isolement.

L'évolutivité revient à passer d'une normalisation « métier » (MARC, Z 39-50) aux standards du web pour être interopérable avec les autres fournisseurs/clients des métadonnées et être ainsi présent dans l'écosystème du web. Les bibliothèques doivent en effet pouvoir échanger avec d'autres communautés qui produisent de l'information bibliographique sur le web. Il faut donc aller vers le langage XML, format de l'information bibliographique sur le web.

4.1 Des solutions en cours

La conversion en XML des formats MARC est d'ores et déjà possible grâce au schéma MARCXML²², schéma XML pour le format MARC 21 développé et maintenu par la Bibliothèque du Congrès. Ce schéma propose une structure très simple, permettant une conversion sans perte d'information et réversible d'une notice MARC conforme à la norme ISO 2709 vers une notice MARC en XML. Il peut être utilisé dans le protocole OAI-PMH.

Le format MarcXchange (Norme NF ISO 25577:2009) a été élaboré dans l'optique d'étendre le principe de base du schéma MARCXML à tous les formats MARC. Il a été conçu comme une alternative à la norme ISO 2709 pour les bibliothèques ayant un catalogue en XML, et permettre l'échange de l'information bibliographique en XML. Sa structure est identique à celle définie par MARCXML. La limite de la norme dans sa version actuelle est de ne pas gérer la technique de l'imbrication des zones. Il ne permet donc pas de gérer complètement le format UNIMARC.

Le protocole OAI-PMH²³, protocole asynchrone et basé sur les standards du web (protocole HTTP et XML), est un moyen d'échanger sur Internet des métadonnées entre plusieurs institutions, afin de multiplier notamment les accès aux documents numériques. Il peut aussi contribuer à l'échange de données catalogographiques par l'introduction du format MarcXchange, mais ce protocole est moins adapté pour des mises à jour régulières de données. Aujourd'hui il n'existe toutefois pas encore de possibilité offerte pour cette modalité de récupération de données catalogographiques, sauf dans le cadre de portail collaboratif comme le CCFr avec les portails RACHEL et Lectura.

D'autres solutions de fourniture de données apparaissent avec le développement de Web services²⁴. Actuellement, les Web services proposés viennent enrichir les Opacs par l'intégration de liens pointant vers des contenus tel que les vignettes ou bien la 4e de couverture. Mais ces solutions ne sont pour l'instant pas implémentées dans les SIGB pour la récupération de données catalogographiques.

4.2 Un environnement en mutation

Cette période de transition se caractérise par un passage des formats MARC à un nouveau format en cours de définition, qui permettra une organisation de l'information bibliographique fondée selon le modèle FRBR²⁵, modèle conceptuel de données bibliographiques élaboré par

22 <http://www.loc.gov/standards/marcxml/>

23 Sigle de l'Open Archives Initiative – Protocol for Metadata Harvesting (Protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes)

24 Composant applicatif utilisant les standards du web (protocole HTTP et langage XML) permettant à un logiciel de communiquer avec d'autre logiciel

25 Functional Requirements for Bibliographic Records / Fonctionnalités requises des notices bibliographiques.

l'IFLA (éclatement de la « notice bibliographique » en entités reliées entre elles).

Cette évolutivité s'accompagnera d'une présentation plus claire des résultats à l'affichage, une meilleure visualisation et permettra de répondre ainsi aux besoins des usagers. Il faut que les catalogues des bibliothèques puissent être exploités par les moteurs de recherche afin qu'ils répondent davantage aux besoins des usagers et à leurs pratiques du web.

Ces différents aspects sont développés dans la fiche n°2 : Ré-informatisation et évolution des catalogues.

Recommandations

- **Veiller à la bonne structuration des données**
- **Distinguer les zones relevant des notices bibliographiques et les zones relevant des données locales ou d'exemplaire [Gérer les données locales indépendamment de la notice bibliographique]**
- **S'assurer de la provenance des autorités (auteurs, matières...) fournies par le réservoir (par exemple, mention de la BnF)**
- **Choisir le ou les fournisseurs en fonction des supports et/ou de la stratégie de la bibliothèque**
- **Choisir de préférence un fournisseur garantissant l'origine de la notice de référence**
- **Veiller à la configuration du SIGB pour une récupération normative et cohérente des notices bibliographiques et d'autorité**
- **Ne pas modifier les notices récupérées sauf exception**
- **Saisir uniquement les données locales ou d'exemplaire**
- **Ne jamais supprimer la zone des identifiants des notices de références [zone 035 ou 003],**
- **Ne jamais supprimer les liens entre notices (les sous-zones numériques telles que \$3 (identifiant de la notice d'autorité), \$0 (identifiant de la notice bibliographique), qui sont des données de liens entre notices ou zones.**
- **Conserver la source de provenance de la notice dans la zone 801**
- **Programmer le nettoyage de son catalogue si les données ne sont pas conformes à ce qui est expliqué dans cette fiche**

Pour en savoir plus

Sites officiels	
Format UNIMARC bibliographique	http://www.bnf.fr/fr/professionnels/anx_formats/a.unimarc_manuel_format_bibliographique.html
Structure du format UNIMARC Bibliographique	http://multimedia.bnf.fr/unimarc_b_trad/BSection3-6-2011.pdf
Format UNIMARC autorité	http://www.bnf.fr/documents/UnimarcA.pdf
Comité français UNIMARC	http://cfu.enssib.fr/index.php
RAMEAU	http://rameau.bnf.fr/index.htm
FRBR	http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html
Application format UNIMARC en France	
Format UNIMARC des produits de la BnF (Autorité et Bibliographique)	http://www.bnf.fr/fr/professionnels/outils_recuperation/a.produits_format_unimarc.html
Produits et services bibliographiques proposés par la BnF	http://www.bnf.fr/fr/professionnels/produits_et_services_bibliographiques.html
Recommandation 995 pour la fourniture de données locales dans les échanges de notices bibliographiques en UNIMARC accompagnant le prêt ou le dépôt d'exemplaires : Partie 1	http://www.adbdp.asso.fr/IMG/pdf_r995.pdf
Format des données d'exemplaire (Recommandation 995, Partie 2)	http://www.abes.fr/Media/Fichiers/Sudoc-Fichiers/Produire-dans-le-Sudoc/Recommandations-pour-l-echange-de-donnees-d-exemplaire-en-format-UNIMARC
Articles et journées d'étude sur la constitution des catalogues	
Exemple de mise en place de récupération des données dans la BM de Fresnes	Giappiconi, Thierry, « Les ressources bibliographiques de la BNF au service de la lecture publique. », <i>BBF</i> , 1998, n° 6, p. 26-33 [en ligne] < http://bbf.enssib.fr/ > Consulté le 24 mai 2012 http://www.slideshare.net/bibliobsession/privilgier-la-bibliographie-nationale-franaise-pourquoi-et-comment/
Article de Pierre-Yves Duchemin sur les catalogues	Duchemin, Pierre-Yves, « L'enrichissement des catalogues ? Et après ? », <i>BBF</i> , 2005, n° 4, p. 21-27 [en ligne] < http://bbf.enssib.fr/ > Consulté le 30 mai 2012
Présentation de Jacques Kergomard à la journée d'actualité « La constitution des catalogues », Béziers, 26 avril 2010	http://jkergomard.free.fr/jecat/co/Catalogue.html
Journée des utilisateurs des produits et services bibliographiques, BnF, 2010	http://www.bnf.fr/fr/professionnels/autres_journees_professionnelles/a.journee_utilisateurs_produits_2010.html

Protocoles d'échanges	
Z39.50	<p>http://www.loc.gov/z3950/agency/</p> <p>http://www.bnf.fr/fr/professionnels/protocoles_echange_donnees/a.proto_z3950.html</p> <p>RIVIER, Alexis. Aide-mémoire d'informatique documentaire. Paris : Editions du cercle de la librairie, 2007. (Collection Bibliothèques), p. 111 -114</p> <p>CZAPLINSKI, Jean-Marc, MORET, Yves. "Z39.50", Les mutations , <i>Bulletin d'information de l'ABF</i>, n°174, 1997, p.77-81, [en ligne], http://www.enssib.fr/bibliotheque-numerique/revues/afficher-45392, consulter le 22 mars 2012</p>
SRU (Search/Retrieval via URL)	<p>http://www.loc.gov/standards/sru/</p> <p>http://www.bnf.fr/fr/professionnels/proto_sru/s.proto_sru_intro.html</p>
Protocole OAI-PMH	http://www.bnf.fr/fr/professionnels/protocoles_echange_donnees/a.proto_oai.html#SHDC_Attribute_BlocArticle1BnF
Forums d'échanges entre les utilisateurs de logiciels de bibliothèque	
Guide pour l'administration de PMB	http://www.sigb.net/ml/user/pmb.user-200503/msg00047.html
Forum PMB	http://www.pmbservices.fr/nouveau_site/wiki.html
Communauté Archimed et Opsys portail Hermes	http://aduloa.asso.fr/
Site de la FULBI pour une liste de club utilisateurs	http://www.fulbi.fr/?q=content/membres
Utilitaires	
Exemple d'une récupération de notice efficace	http://www.dailymotion.com/video/xf48dc_cataloguer-un-livre-en-1-minute-et_lifestyle#from=embed
Librairies proposant des notices UNIMARC gratuitement en UTF8 et Marc21	http://www.librairiedialogues.fr/livre/1451388-sanctus-michel-quint-terrain-vague
Solutions pour transformer un serveur Z39.50 en serveur SRU	YAZ Proxy développé par Index Data
MarcEdit, utilitaire pour convertir des notices MARC ISO 2709 en MARCXML	<p>http://people.oregonstate.edu/~reese/marcedit/html/index.php</p> <p>http://www.slideshare.net/dok205/cataloging-with-marcedit</p>