

Diplôme de conservateur de bibliothèque

Mémoire d'étude / Janvier 2009

L'avenir des catalogues

Formats, données, outils, usages

Romain Wenz

Sous la direction de Christian Lupovici

Directeur du département Information bibliographique et numérique (DIBN)
Bibliothèque nationale de France

Remerciements

Mes remerciements vont en premier lieu à Christian Lupovici, Directeur du département Information bibliographique et numérique (DIBN) à la Bibliothèque nationale de France, qui m'a dirigé et orienté au cours de ce travail.

Merci à Emmanuelle Bermès, conservatrice à la Bibliothèque nationale de France, qui a proposé le sujet de ce mémoire et dont l'activité de veille et d'explications sur le Web est toujours éclairante.

Je remercie aussi tout particulièrement Clément Oury, Sébastien Peyrard et Matthieu Bonicel, conservateurs à la Bibliothèque nationale de France, pour leur aide et leurs conseils.

Merci à Sally Chambers, responsable de l'interopérabilité des métadonnées de TEL pour ses explications sur les enjeux techniques.

Merci à Christophe Catanese (ENSSIB) pour son soutien et ses relectures finales.

Merci aussi à ceux et celles qui m'ont aidé et formé par leurs enseignements au cours de ma scolarité, notamment Yves Desrichard, Isabelle Westeel et Benoît Éperon à l'ENSSIB, mais aussi Patrick Le Bœuf, Christine Nougaret, et particulièrement Gautier Poupeau à l'École nationale des Chartes.

Merci aux équipes de la BIUSJ-SCDM (Paris 6), et tout particulièrement à Isabelle Kratz, pour m'avoir donné l'occasion d'utiliser les outils de catalogage collectif.

Merci à tous ceux qui m'ont amené à m'intéresser au traitement informatique des supports de l'écrit, par leurs conversations ou leurs publications.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.
--

Résumé :

L'avenir des catalogues fait actuellement l'objet de nombreuses réflexions. En France, la BnF et l'ABES notamment mènent des travaux d'analyse et de remise en question dont les conclusions auront une importance pour les bibliothèques françaises. Ces enjeux concernent les bibliothèques du monde entier, avec l'action puissante de l'OCLC et la mise en place d'un catalogue mondial unique (Worldcat), mais aussi avec le développement de nouveaux modèles de structuration de l'informations bibliographique (FRBR), et une réflexion sur les règles de catalogage (ISBD consolidé, RDA), pour une évolution des formats (avec le langage XML), et éventuellement vers des évolutions plus fondamentales (Web sémantique).

Il ne s'agit pas de suivre mécaniquement l'évolution des technologies informatiques, mais de construire des outils pertinents pour répondre aux besoins des lecteurs, en ligne ou sur place. En effet, l'évolution générale du Web amène les bibliothèques numériques à rejoindre les autres types de ressources en ligne, et les anciens points d'accès prennent un autre sens dans des outils de recherche pondérée par des algorithmes. En outre, la navigation dans un environnement numérique amène la recherche d'information à se dissocier de la lecture de loisirs. Or, les outils de recherche des ouvrages doivent correspondre aux usages, et évoluer en conséquence, qu'il s'agisse de recherches à caractère studieux, d'informations pratiques, ou d'ouvrages de loisir

Un mémoire d'études ne peut avoir la prétention de faire un inventaire complet des travaux en cours. Il s'agira plutôt de décrire les enjeux et les outils place, existants ou possibles, pour présenter les formats et outils de travail sans adopter le point de vue d'un acteur particulier. On s'efforcera de considérer avec distance les outils pour lesquels l'engouement semble unanime, sans tenter de prédire l'avenir, mais de porter un regard aussi objectif que possible sur les héritages, évolutions et besoins techniques devant être pris en compte pour construire cet avenir.

On observera donc les formats de l'information bibliographique, les usages induits par le développement d'Internet, les nouveaux modèles de description au travers des métadonnées, et les changements que provoque le développement des bibliothèques numériques.

Descripteurs :

Catalogues de bibliothèques en ligne
Bibliothèques et Internet
Bibliothèques virtuelles
Numérisation
Bibliothéconomie**Innovations
Informatique documentaire
Indexation (documentation)
Bibliothèques ** Publics
Bibliothèques ** Services aux utilisateurs
Internautes
Web 2.0
Web sémantique

Abstract :

The « Future of Catalogs » is a subject many librarians talk about. The French national Library and ABES for instance are currently working on improvement programs, which will have a strong impact on French libraries. These matters are important for libraries worldwide, not only because the powerful OCLC is building a unique catalog (Worldcat), but also because new models for bibliographical description are being built (FRBR) together with a work on cataloguing rules (ISBD consolidated edition , RDA), so that the information formats can change (with XML language), and perhaps towards even more important changes (Semantic Web).

It is not just a matter of following new computer technologies, but of building the right tools in order to answer the reader's needs, whether online or not.

In fact, because of the way the Web is now changing, Digital Libraries start to be one online resource like any other. Besides, the terms of entries get a new meaning because of browsers and search engines. Moreover, the habit of surfing on a digital environment may widen the gap between information retrieval and leisure reading. Anyway, the search tools must comply with the user's needs, and change accordingly, for scholarly research, practical information or leisure books.

This dissertation cannot pretend to sum up all the undertaken works. Its aim is to describe what is at stake and the existing or possible tools. We will not consider the formats and work possibilities from any particular point of view, but try to remain neutral, specifically towards fashionable tools. Without trying to guess what the future of catalogs will be, we will describe the legacies, changes and technical needs which will interfere with this future.

We will therefore describe the tools of bibliographical description, the new Internet uses, the new description standards for metadata, and the changes which are linked to the development of digital libraries.

Keywords :

Online library catalogs
Libraries and the Internet
Digital libraries
Digitization
Library science**technological innovations
Information retrieval
Authority files (Information retrieval).
Public services (Libraries)
Library use studies
Indexing
Internet users
Web 2.0
Semantic Web

Table des matières

1. INTRODUCTION	11
1.1. QU'EST-CE QU'UN CATALOGUE ?.....	11
1.2. QUELS OBJECTIFS POUR UN CATALOGUE AUJOURD'HUI	12
2. LE CATALOGUE INFORMATISE	15
2.1 DES DONNEES	15
2.1.1 <i>Les éléments de données : leur établissement</i>	15
- Quelles informations ?.....	15
- Quelle description ?.....	16
2.1.2 <i>Les règles de catalogage</i>	16
- Des règles multiples	16
- Nouvel ISBD, nouvelles AACR	17
- Les normes nationales	18
- Les instances de normalisation.....	18
2.2 UNE STRUCTURE : LES FORMATS.....	19
2.2.1 <i>MARC générique (ISO2709) et MARC particuliers</i>	19
- Des champs de longueur variable	20
- Une grande famille	20
- Interopérabilité et évolutions	22
2.2.2 <i>MARC et structures XML</i>	22
- Une technologie prometteuse	22
- Echanger, indexer, décrire	23
- Un usage spécialisé	24
2.3 DES SYSTEMES DE GESTION DE BIBLIOTHEQUE	25
2.3.1 <i>Des outils techniques</i>	25
- L'origine de l'informatisation	25
- L'accès sur le Web	26
- L'intégration au Web.....	26
2.3.2 <i>Des moyens de gestion</i>	27
- Le circuit du document	27
- Des fonctions de gestion	28
2.3.3 <i>Une inertie à prendre en compte</i>	29
- Contrats et habitudes	29
- Des changements forcés.....	29
2.4 DES SYSTEMES D'INTERROGATION.....	30
2.4.1 <i>Indexer les champs</i>	30
- Une expertise ancienne	31
- Valider les termes.....	31
- L'indexation par des robots.....	32
2.4.2 <i>Les FRBR</i>	33
- Une nouvelle analyse théorique	33
- Repenser des données identiques	33
2.5 DES INTERFACES HOMME-MACHINE	34
2.5.1 <i>Apparence et ergonomie</i>	34
- Plaire à l'utilisateur	34
- Faciliter la recherche : quels choix ?	35
2.5.2 <i>Fournir les documents</i>	36

- Accéder aux documents.....	36
- Rechercher dans les documents	37
3. LES USAGES DES INTERNAUTES ET LEUR IMPACT SUR LES CATALOGUES	38
3.1 ERGONOMIE DU WEB ET ACCES AUX RESSOURCES	38
3.1.1 <i>Les moteurs pour la recherche d'information</i>	38
- De nouveaux outils.....	38
- De nouvelles habitudes.....	39
- Une logique de masse.....	39
3.1.2 <i>Consultation en ligne et téléchargement</i>	41
- Tout sur place	41
- Culture gratuite et immédiate	41
3.1.3 <i>Usagers : des lecteurs catalogueurs ?</i>	42
- Indexer selon les utilisations.....	42
- « Web 2.0 » : changements d'usages ?	43
- Changements de public ?	44
- Changements de contenu ?	45
3.2 DES CATALOGUES SUR LE WEB HORS DES BIBLIOTHEQUES	47
3.2.1 <i>Les catalogues privés</i>	47
- Catalogues thématiques	47
- Catalogues institutionnels de vente	49
3.2.2 <i>Les livres en ligne</i>	49
- Trouver des ouvrages papier.....	50
- Chercher dans le texte des livres.....	51
3.2.3 <i>Comment les bibliothèques s'y intègrent</i>	52
- Sur le chemin des internautes	52
- Dans des services existants ?	53
- Changements d'outils	54
4. L'AVENIR DES CATALOGUES DANS LE CONTEXTE DU WEB.....	57
4.1 LES METADONNEES A USAGE MULTIPLE	57
4.1.1 <i>Catalogue des métadonnées</i>	57
- Définition	57
- Métadonnées descriptives.....	58
- Métadonnées de structure	59
- Métadonnées administratives.....	59
4.1.2 <i>Nouvelles règles de création des métadonnées</i>	60
- Décrire dans le document	60
- Les possibilités du RDA	61
4.1.3 <i>Adapter des schémas de données déjà existants</i>	62
- Des schémas XML	63
- La diffusion du Dublin Core	65
- L'Open Archive Initiative	66
- Rompre l'isolement des bibliothèques.....	67
4.2 NOUVELLE ARCHITECTURE POUR LE CATALOGUE DE DEMAIN	68
4.2.1 <i>Une spécialisation nécessaire</i>	68
- Une standardisation mondiale.....	69
- Un besoin particulier de pérennité	69
- Un double travail ?	70

- S'adapter aux supports.....	71
4.2.2 <i>Les possibilités du RDF</i>	72
- Structurer l'information en triplets.....	72
- Un travail hors des bibliothèques	73
- Le « Web sémantique » : un grand catalogue ?	74
5. CATALOGUE ET BIBLIOTHEQUE NUMERIQUE	75
5.1 UN OUTIL POUR TOUS LES SUPPORTS.....	75
5.1.1 <i>Un fichier de métadonnées : identifier et localiser</i>	75
- Des changements communs.	75
- Un autre univers sémantique.....	76
- Des choix importants	77
5.1.2 <i>Les supports multiples des médiathèques virtuelles</i>	77
- Analogiques et numériques, multiples et uniques.....	77
- Généraux et spécialisés.....	78
- Quelle place pour l'audiovisuel ?	78
5.2 GERER LES DOCUMENTS.....	79
5.2.1 <i>Classer l'immatériel</i>	79
- Tout doit être dit.....	79
- Quels modes de recherche ?	80
- Quelle indexation ?	81
5.2.2 <i>Contrôler l'accès</i>	82
- L'ubiquité.	82
- Les collections à accès restreint	83
5.2.3 <i>Cycle de vie et de mort</i>	84
- Gérer des contenus	84
- S'associer à la production ?	84
5.3 ACCEDER AUX COLLECTIONS.....	85
5.3.1 <i>Depuis où ?</i>	85
- Depuis les catalogues ?	85
- Par des entrepôts moissonnables ?	86
- Par des pages Html ?	87
5.3.2 <i>Jusqu'où ?</i>	87
- Des services payants ?	87
- Un catalogue de lecteurs ?	88
- La bibliothèque partout ?	89
6. CONCLUSION	91
BIBLIOGRAPHIE.....	93
TABLE DES ANNEXES.....	103

1. Introduction

Ce mémoire d'étude correspondant à l'apprentissage technique d'un métier, il reflète le point de vue d'un apprenti conservateur, et non d'un professionnel de l'informatique. Par conséquent, nous nous attacherons principalement aux enjeux et perspectives du sujet selon les évolutions nécessaires ou souhaitables dans le monde des bibliothèques. Il s'agit ici des catalogues de bibliothèque au sens strict de l'instrument de recherche (à l'exclusion des catalogues d'exposition ou de vente par exemple).

1.1. QU'EST-CE QU'UN CATALOGUE ?

Le catalogue, est une liste des pièces ou éléments constituant une collection, qui doit d'une part décrire ces éléments pour permettre de les identifier, d'autre part les localiser pour permettre de les retrouver.

Par opposition à l'inventaire (outil interne recensant les pièces), le catalogue est destiné à l'utilisation par une clientèle. Il s'agit à la fois d'un outil technique et d'un instrument de visibilité de l'institution.

Selon l'ABF¹, le catalogue d'une bibliothèque doit permettre de

- « - retrouver un document à partir d'un seul de ses éléments connus [...]
- connaître ce que possède la bibliothèque [...]
- guider l'utilisateur dans le choix d'un document ».

Il s'agit donc d'un instrument de description et de localisation permettant au lecteur d'accéder aux ressources.

Le catalogue peut traiter des documents sous divers supports, et être accessible sous différentes formes. L'ADBS² établit une typologie pour cet outil rendu plus complexe avec la multiplication des supports, aussi bien des documents que des catalogues eux mêmes, en distinguant :

- 1- contenu intellectuel³
- 2- forme matérielle
- 3- mode de publication
- 4- forme de production⁴.

Ainsi, le catalogue reste un instrument de recherche composé par une liste de notices, quel qu'en soit le support : papier, microfiche, informatique.

Le passage des fiches papier au catalogue informatisé impliquait bien sûr des nouveautés techniques, mais sans changer vraiment la conception intellectuelle de l'outil.

¹ DELACOUR, Christiane E. s., et al. Cataloguer : mode d'emploi : initiation aux techniques du catalogage. Paris : ABF: diff. ABIS, 2006.

² BOULOGNE, Arlette; Vocabulaire de la documentation. Paris: ADBS Editions, 2004.

³ Exemple : musique notée ou enregistrée, son, image fixe, carte et plan, audiovisuel, multimédia, informatique : programme et caractères

⁴ Manuscrit/ Littérature grise/Publication

C'est ainsi que, par le passé, l'informatisation des catalogues a eu pour objectif de rendre l'outil plus commode et puissant, mais en le destinant au même usage. Certes, la structuration de l'information changeait, les fiches étant intégrées en bases de données, mais le catalogue restait un simple instrument de médiation entre la collection et les usagers, absolument indispensable au prêt voire à la consultation, avec la triple fonction de rechercher, identifier et localiser des ressources. En outre, il s'intègre au système de gestion de la bibliothèque, et est utilisé par les professionnels pour gérer les collections.

Pourtant, on a pu observer un changement récent, cette dernière décennie, avec le développement du Web. En effet, les catalogues informatiques étaient au départ élaborés de façon techniquement autonome. Certes, les catalogues de bibliothèque ont été mis sur les réseaux très tôt, d'accès gratuit et tout public, avant Internet. Mais la technologie Web amène l'informatisation et le réseau à se mêler. Ce qui implique d'une part que les données des différents catalogues apparaissent de façon comparable et pourraient sembler interchangeables, et d'autre part que les catalogues ne sont qu'une des sources d'information parmi les autres sur la Toile.

Surtout, il semble évident que l'évolution des modes d'accès, avec l'ergonomie du Web, change complètement les habitudes des lecteurs. Le catalogue disponible en ligne s'adresse à un public pour lequel les interfaces de catalogues et de portails de bibliothèques s'intègrent dans un paysage documentaire marqué par une ergonomie particulière.

Un des enjeux des catalogues en ligne est donc de faire en sorte que les ressources pertinentes soient repérées par un public dont les habitudes de recherche sont déjà fixées.

1.2. QUELS OBJECTIFS POUR UN CATALOGUE AUJOURD'HUI

Il ne fait pas de doute qu'un catalogue disponible en salle de lecture sert essentiellement à vérifier si la bibliothèque détient un ouvrage ou un type d'ouvrage que l'on recherche. On parle bien ici de l'outil permettant aux différents publics de repérer et se procurer les documents.

Accessible sur Internet, le catalogue standard conserve ces fonctions de décrire et identifier, localiser et retrouver les documents.

Pourtant, les objectifs des utilisateurs peuvent en ce cas être bien plus nombreux.

Avec un catalogue en ligne, on peut aussi chercher à trouver un ouvrage incertain ou retrouver ouvrage connu. Mais l'internaute peut en outre chercher un ouvrage à télécharger, voire chercher simplement une information. En ce cas, le passage par l'ouvrage lui-même est une étape jugée superflue.

On se heurte ici à un des problèmes principaux des catalogues en ligne : fournir une référence pertinente à des utilisateurs qui n'avaient pas forcément prévu au départ de faire appel aux outils d'une bibliothèque.

A ces objectifs multiples pourraient correspondre des outils multiples, en fonction des usages variés. Pourtant, cette adaptation ne saurait être résolue par une prolifération des

instruments, mais par un mode de description des ressources qui s'intègre aux modes d'accès des internautes. Cette évolution du service concerne tout particulièrement les internautes qui ne sont pas usagers de bibliothèques, mais peuvent y trouver des ressources, y compris virtuelles.

Il s'agit à la fois de faire en sorte que le catalogue soit repérable depuis les sites fréquentés par les internautes, et que son ergonomie corresponde aux pratiques de recherche de ces derniers, éventuellement en y intégrant des types de description des ressources et des fonctionnalités correspondant à l'évolution du Web.

En outre, avec le développement des bibliothèques numériques, les ressources décrites peuvent aussi bien être des ouvrages matériels que des documents numérisés ou électroniques, éventuellement disponibles en ligne.

Ces différents enjeux amèneront à décrire tout d'abord les contraintes techniques du catalogue informatisé, puis les usages des internautes, pour observer comment peuvent évoluer les catalogues dans le contexte du Web, en particulier pour inclure l'accès aux ressources numériques.

2. Le catalogue informatisé

2.1 DES DONNEES

L'outil informatique a été utilisé dès les années 1960 pour stocker les données et les rendre accessibles, avec les contraintes de production de l'époque, mais le stockage de l'information a très tôt été effectué en bases de données.

2.1.1 Les éléments de données : leur établissement

Les catalogues fonctionnaient sous une forme matérielle avant l'informatisation, avec des fichiers de notices papier. Sous une forme informatique, le catalogage continue à être l'action de créer une information secondaire à partir d'une information primaire.

- Quelles informations ?

Le mode de description des documents a été conservé, avec une notice par ouvrage, et quatre éléments d'information distincts dans chaque notice :

- 1- La notice bibliographique, décrivant le support et le contenu du document, en tant qu'objet matériel et que création intellectuelle.
- 2- Les points d'accès, c'est-à-dire l'auteur, le titre et les mots clés retenus⁵ pour retrouver le document, avec à la fois des éléments d'autorité et de description.
- 3- Les données d'exemplaire, concernant les exemplaires conservés et donc propres à l'établissement,
- 4- Une localisation pour se le procurer (cote).

Pourtant, la structuration des données change. En effet, ces quatre types d'éléments permettent une structuration en base de données, avec une possibilité de lecture de notice synthétique, et, de façon différente, des liens pour effectuer des requêtes simples à partir des auteurs, titres, et mots clés, avec un accès aux données locales.⁶ Pour des ouvrages identiques, seules les données locales sont vraiment différentes d'un établissement à l'autre. La question de la normalisation a donc rapidement été posée. Plus encore, le problème de l'interopérabilité est devenu essentiel, pour l'échange et la dérivation de notices, mais aussi pour les recherches fédérées.

Ce travail de normalisation a débuté en 1961 avec la Conférence internationale sur les principes de catalogage, au cours de laquelle ont été rédigés les « principes de Paris », clarifiant les objectifs et les normes des catalogues.⁷

⁵ Sujet, matières...

⁶ Cote, exemplaires, voire état de disponibilité.

⁷ CHAPLIN, A. et al. International Conference on Cataloguing Principles, Paris, 9th-18th October, 1961 : report. London: IFLA International Office for UBC, 1963., p. 91-96.

- Quelle description ?

Les « principes de Paris » d'octobre 1961 étaient établis pour des catalogues manuels, et pour des ouvrages imprimés essentiellement. Ils décrivent donc la constitution du catalogue avec l'ordre linéaire dans lequel il faut faire apparaître les notices. Les normes sur le contenu des notices en découlent. Il s'agit des éléments de l'ISBD,⁸ sous ses diverses formes.

Les ISBD émanent de l'IFLA⁹, et ont pour but d'établir un socle commun pour les normes nationales. Ils définissent les éléments de description bibliographique.¹⁰

Ce travail de normalisation est lié au programme de l'IFLA de CBU,¹¹ mis en place vers 1960 pour décrire uniformément les documents produits. Ce projet tend à se réaliser avec les catalogues collectifs, mais se heurte à la diversité des documents disponibles en ligne, pour lesquels la normalisation des informations secondaires est délicate. Publiés à partir de 1971 à l'issue de la Conférence internationale des experts en catalogage, réunie en 1969 à Copenhague par l'IFLA, les premiers ISBD correspondent à des principes inadaptés à l'environnement actuel. L'évolution pour une mise à jour des règles de catalogage est donc essentielle.

2.1.2 Les règles de catalogage

Il convient d'observer comment les règles multiples évoluent, au travers des efforts d'uniformisation mondiaux et nationaux.

- Des règles multiples

L'articulation des différents ISBD couvre l'ensemble des documents imaginables, avec l'ISBD(G) générale, et divers types.¹² Les éléments d'identification sont répartis en 8 zones¹³. Il s'agit seulement de description bibliographique (à l'exclusion des points d'accès), avec une structure intentionnellement rigide.

La multiplication des supports pour les mêmes oeuvres, rend moins pertinente la distinction entre des types de documents spécialisés.¹⁴ Surtout, la répartition des éléments des ISBD a été conçue pour des fiches, et est centrée sur l'auteur.¹⁵ Plusieurs éléments ne correspondent plus aux catalogues, ou plus précisément aux besoins des lecteurs. Ainsi, contrairement aux normes sur l'édition électronique, l'accès à l'éditeur n'est pas intégré dans les ISBD.

⁸ International Standard Bibliographic Descriptions, <http://www.ifla.org/VII/s13/pubs/cat-isbd.htm>

⁹ Fédération internationale des associations de bibliothécaires et de bibliothèques, principale organisation professionnelle, avec 1700 membres dans 150 pays, <http://www.ifla.org>

¹⁰ Avec l'ordre des éléments, les règles de transcription, les sources d'information, ainsi que la ponctuation, syntaxique et sémantique.

¹¹ Contrôle bibliographique universel, programme de recensement de toutes les publications, mis en place par l'IFLA aux environs de 1960. <http://www.ifla.org/VII/s13/pubs/cat-isbd.htm>

¹² Les ISBD(M), ISBD(PM), ISBD(CM), ISBD(NBM) proposent une distinction par nature du contenu, les ISBD(NBM), ISBD(ER) une distinction par nature du support, l'ISBD(CR) par mode de publication, l'ISBD(A) par date des documents.

¹³ Voir Annexe 1, précisions sur les ISBD.

¹⁴ Par exemple pour une publication en série sur support numérique contenant une carte ancienne...

¹⁵ Par exemple, pour les mentions de responsabilité, les trois premiers auteurs seulement sont cités, ce qui avait du sens pour des fiches papier mais est inadapté à l'informatique.

L'IFLA a entrepris en 2003 un projet de fusion de tous les ISBD en un seul ISBD.¹⁶ Cet « ISBD consolidé » publié en 2007 par l'IFLA propose de traiter l'ensemble des documents avec un seul modèle ISBD.¹⁷

- Nouvel ISBD, nouvelles AACR

L'ISBD consolidé fusionne tous les ISBD en un seul, pour permettre une internationalisation des règles. Les nouveaux principes décrivent toujours le pavé bibliographique ISBD comme « l'ensemble des données bibliographiques qui identifient une ressource », mais présentent plusieurs changements.

Ils couvrent tous types de documents, quels que soient leur contenu et leur support, ils permettent des accès par notices d'autorités ou par sujets, et ils prévoient que le catalogue intègre des fonctions permettant de rechercher le même type de données selon la nouvelle analyse théorique du modèle FRBR.¹⁸

Il s'agit par exemple de trouver toutes les ressources correspondant à une même œuvre, à une même expression ou à une même manifestation. Ou de rechercher les œuvres d'une même personne ou collectivité, ou sur un sujet, voire selon d'autres critères comme la date, le pays ou la langue.

Parallèlement à la révision de l'ISBD, les AACR¹⁹ sont devenues AACR 2000 ou « AACR2 ». Ces normes de catalogage ne sont pas une initiative de l'IFLA, mais de *l'American library association*, ainsi que d'autres organisations comme la *Canadian library association*.²⁰ Même sous leur forme révisée, les AACR suivent la même logique que l'ISBD, seuls certains éléments de structure et de ponctuation étant différents dans certains codes. Plus encore, le RDA (Resource Description and Access)²¹ se présente comme le successeur des AACR2, et propose une solution autre que celle de l'ISBD, en adaptant les FRBR à une logique de XML. Ce travail, parallèle et en pratique concurrent de celui de l'ISBD consolidé, propose l'élaboration d'un nouveau code de catalogage. Le RDA est élaboré par le JSC,²² comme les AACR, et ne dépend pas de l'IFLA. Il fait explicitement référence au modèle FRBR, et propose une façon nouvelle de structurer intellectuellement l'accès aux ressources.

Un des enjeux pour l'avenir des catalogues sera le choix d'une évolution sur le modèle de l'ISBD consolidé, ou du RDA, sous une forme transposée dans les normes nationales.

¹⁶ TILLET, Barbara B. É s. IFLA cataloguing principles : steps towards an international cataloguing code, 2 : report from the 2nd IFLA Meeting of Experts on an International Cataloguing Code, Buenos Aires, Argentina, 2004, München: K. G. Saur, 2005.

¹⁷ <http://www.ifla.org/VII/s13/pubs/cat-isbd.htm>

¹⁸ Fonctional Requirements for Bibliographical Records, Cf Infra.

¹⁹ Anglo American Cataloging Rules, <http://www.aacr2.org/>

²⁰ Elles sont utilisées notamment aux États-Unis, au Royaume-Uni et au Canada, mais aussi sous des formes voisines en Allemagne (RAK) et en Italie (RICA).

²¹ Resource Description and Access, voir <http://www.collectionscanada.ca/jsc/index.html>, et *Infra*, 4.1.2.

²² Le Joint Steering Committee a pour but l'établissement du RDA. Il regroupe l'ALA (American Library Association), l'Australian Committee on Cataloguing, la British Library, le Canadian Committee on Cataloguing, le CILIP (Chartered Institute of Library and Information Professionals), et la LOC (Library of Congress). <http://www.collectionscanada.gc.ca/jsc/index.html>

- Les normes nationales

Alors que le monde Anglo-Saxon utilise le terme de « standard » pour désigner les normes officielles et les standards de fait, la France est attachée aux normes validées par les autorités compétentes. Avant les ISBD, les règles de l'ABF²³ prévalaient en France, sans portée internationale. Elles permettaient une structuration des données lisibles par des machines pour l'échange et l'interprétation, mais se révélaient insuffisantes.

Le programme CBU avait mis en place une Agence Bibliographique Nationale par pays, pour recenser les publications nationales. Il s'agissait de la BnF pour la France. Elle se vit donc attribuer la traduction des ISBD, pour faciliter l'établissement et l'interprétation des notices.²⁴

L'ensemble de textes à portée internationale publié à partir de 1971 fut adapté dans des normes intégrant les particularités nationales, avec l'AFNOR en France.²⁵ Ainsi, plusieurs commissions AFNOR élaborent des fascicules de documentation sur les règles du catalogage.²⁶ Les multiples normes internationales ISBD se sont traduites par des fascicules nationaux aussi nombreux, qui ne sont pas des normes au sens réglementaire.

En effet, outre le nombre de textes, il faut mentionner leur diversité, avec trois niveaux :

- Les normes françaises homologuées NF entérinées.
- Les normes expérimentales FD.²⁷
- Les brochures non révisées mais à caractère normatif.

Un des objectifs pour la normalisation est de faire en sorte que le catalogage soit effectué selon des normes à jour, alors que des textes spécialisés antérieurs pourraient être utilisés. De fait, dans le cas des catalogues, diverses commissions produisent des textes utiles au niveau national, ce qui pose un problème de cohérence.

- Les instances de normalisation

On peut mentionner deux ensembles principaux pour la normalisation des données des catalogues : une Commission ISO et une Commission AFNOR. Au niveau international de la normalisation ISO, les questions d'interopérabilité des descriptions sont abordées, par exemple sur le sujet de l'archivage du Web, sur le *records management* et le *Dublin Core*, en TC 46²⁸. Pourtant, les catalogues ne sont pas normalisés au niveau international.

Le catalogage est donc normalisé seulement au niveau national : en CG 46²⁹ « documentation » de l'AFNOR pour la France, s'intéressant notamment aux catalogues de bibliothèques, avec la CN357GE6,³⁰ commission des normes de catalogage.³¹

²³ Association des bibliothécaires Français, <http://www.abf.asso.fr>

²⁴ <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancree=normes/no-isbd.htm>

²⁵ <http://www.afnor.fr/portail.asp>

²⁶ C'est ainsi que les multiples ISBD relatifs à des documents spécifiques ont été adaptés par des normes AFNOR correspondantes : Cf Annexe 1.

²⁷ Par exemple la FD Z44-050 : Norme française de catalogage transposant l'ISBD.

²⁸ Technical Committee 46 (TC46), comité de l'ISO responsable de toutes les questions relatives à l'information et à la documentation. Ainsi, la norme ISO 4689 sur le *records management*, utilisée notamment en Australie, et dans le sud-est asiatique, intéresse tout à fait la gestion de documents numériques natifs.

²⁹ Lié au CM11, sur le Records Management.

³⁰ Modélisation, production et accès aux documents, groupe d'experts 6.

Ces commissions AFNOR permettent un contact entre les bibliothèques et les archives. En effet, la production de normes supérieures aux standards est une particularité française.³² L'AFNOR est transversale (bien que les musées y soient peu impliqués).³³ Elle évite la confrontation entre les métiers et les institutions, en permettant l'élaboration de normes neutres. Pourtant, la grande majorité des normes AFNOR concernant les bibliothèques sont en pratique établies par des personnels de la BnF. La BnF publie d'ailleurs sur son site Internet³⁴ les variantes et interprétations possibles dans les cas de lacunes ou d'ambiguïtés des normes.³⁵ Ce travail correspond à des demandes extérieures.³⁶

Par ailleurs, la Bibliothèque nationale de France exerce en pratique un rôle normatif sur les points d'accès.³⁷ On peut distinguer les points d'accès simples et les points d'accès normalisés, correspondant à des notices d'autorité avec :

- L'entité figurant en vedette.³⁸

- La vedette autorisée, avec les formes rejetées, et les liens, sources et notes.

Par exemple dans le cas des notices d'autorité Rameau, on trouve à la fois la vedette et ses renvois.

La normalisation concerne aussi la structuration de l'information, et ses formats.

2.2 UNE STRUCTURE : LES FORMATS

L'inertie des formats déjà employés rend difficile leur remise en cause. Pourtant, le contexte de standardisation technologique lié au Web (avec l'action du W3C)³⁹ amène à s'interroger sur les formats.

2.2.1 MARC générique (ISO2709) et MARC particuliers

Le premier format MARC a été défini par une norme internationale d'échange, dès la fin des années 60 : l'ISO 2709, premier format à champs de longueur variable.

³¹ Bien qu'il s'agisse légalement de fascicules de recommandations AFNOR et pas de normes.

³² Dans les autres pays, ce sont en général des associations qui établissent des standards spécialisés.

³³ Par exemple pour l'EAD avec la CN4 (correspondant au SC4 de l'ISO), et avec la TC46 sur l'interopérabilité.

³⁴ <http://www.bnf.fr>, rubrique Professionnels et Normes, formats, modélisation, <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancree=normes/no-accueil.htm>

³⁵ De la même façon que la LOC publie la « jurisprudence » des AACR, notamment pour les périodiques avec CONSER, un programme de catalogage coopératif, <http://www.loc.gov/acq/conser/>

³⁶ Notamment des Bibliothèques Universitaires, par exemple dans le cas du catalogage de recueils qui n'est pas prévu par les normes et a fait l'objet de la première fiche publiée par la BnF.

³⁷ Cette action de normalisation concerne aussi bien les points d'accès auteurs que les autorités auteurs, les notices bibliographiques, les notices d'exemplaires, les notices d'autorités et dans une certaine mesure la classification (CDU et Dewey).

³⁸ Personne physique, collection, sujet en Rameau par exemple, nom géographique, titre uniforme.

³⁹ <http://www.w3c.org>

- Des champs de longueur variable

Les formats MARC⁴⁰ proposent une structure des données bibliographiques à deux niveaux (les zones/sous-zones), en donnant une indication de traitement de la zone (les indicateurs), et des positions avec des listes de valeurs codées pour certaines données.

Ces formats reposent sur la norme ISO 2709 « Format pour l'échange d'information », éditée en 1973, qui est adaptée à l'information bibliographique.⁴¹ Elle permet l'échange des données, et non le catalogage au sens strict. Les notices comportent un label,⁴² un répertoire,⁴³ et des données bibliographiques, avec une structure générique des zones.⁴⁴ Le répertoire situé en début structure l'enregistrement destiné à l'échange.⁴⁵

Les formats MARC devaient au départ permettre l'échange des informations bibliographiques, dans le cadre du programme de CBU devenu UBCIM en 1980.⁴⁶ Ce format d'échange sert pourtant de format de saisie pour la constitution des catalogues de bibliothèque, par les bibliothécaires, et les fournisseurs de SIGB, ce qui tend à figer la situation, imposée aux métiers d'édition de logiciels avec l'explosion documentaire des années 1960 mais actuellement maintenue par l'absence d'autres développements.

D'autres normes sont liées au format MARC.⁴⁷ Les formats MARC sont adaptés à l'ISBD,⁴⁸ avec une construction de l'information bibliographique où les accès sont normalisés,⁴⁹ mais ne permettant pas l'intégration de données de gestion. De fait, les multiples déclinaisons de ce format vont à l'encontre de sa fonction d'uniformisation.

- Une grande famille

Les premiers formats MARC ont été élaborés par la LOC,⁵⁰ avec le MARC I en 1965 pour l'échange de fiches papier, puis le Marc II en 1968, adapté à l'informatique, et le LCMarc (Library of Congress Marc). Ces formats conçus pour les échanges de notices ont rapidement été utilisés comme des formats de catalogage adaptés, dont la floraison a conduit dans les années 1970 à l'utilisation d'un MARC différent dans chaque pays ou presque.⁵¹

L'éclatement en des formats nombreux rendait difficile l'échange de données informatisées entre des SGBD⁵² ayant leur propre structuration, ce qui était pourtant le principe du MARC. C'est pourquoi l'IFLA développa en 1977 la première édition de l'UNIMARC (Universal MARC) pour servir de pivot entre les nombreux formats MARC, dans le cadre du programme UBCIM (évolution du CBU). L'UNIMARC est un

⁴⁰ « MACHine Readable Cataloguing »

⁴¹ Elle permet de gérer des zones et sous-zones de longueur variable, facultatives et répétibles.

⁴² Zone fixe de 24 caractères

⁴³ Sur le nombre de caractères et les parties des zones, avec N entrées de 12 chiffres (une par champ).

⁴⁴ Etiquette sur 3 caractères, indicateurs de traitement, codes de sous-zones et caractère de fin de zone.

⁴⁵ qui est notamment prévu par le protocole Z39-50.

⁴⁶ Universal Bibliographical control Marc Program

⁴⁷ En ce qui concerne les jeux de caractères, codes de langues et de pays, code de fonction (pour les auteurs), normes de translittération et règles de catalogage (choix et construction des accès).

⁴⁸ Voir Annexe 1, l'adaptation des ISBD en UNIMARC.

⁴⁹ Selon les standards IFLA, liés au CBU et à l'ISSD.

⁵⁰ Bibliothèque du Congrès, <http://www.loc.gov>

⁵¹ Le LC-MARC fut adapté avec le CANMARC pour l'usage simultané de l'anglais et du français, et l'INTERMARC, puis des développements nationaux comme l'IBERMARC Espagne/Catalogne ou le RUSMARC en Russie, avec un aspect identitaire dépassant les besoins bibliothéconomiques.

⁵² Système de gestion de bases de données.

format bibliographique définissant 10 blocs fonctionnels⁵³ pour permettre les échanges d'information bibliographique entre les multiples formats nationaux. Il est régulièrement mis à jour⁵⁴, et la maintenance en est assurée par le l'IFLA et le PUC.⁵⁵ Ce format d'échange est encore largement utilisé,⁵⁶ mais comme format de saisie (pour limiter les conversions), notamment par des bibliothèques françaises, avec le SUDOC⁵⁷ notamment. Il est maintenu par le CFU.⁵⁸ La BnF fournit ses notices en UNIMARC,⁵⁹ ce qui permet une uniformisation et une dérivation par les autres bibliothèques, pour les intégrer à leurs propres catalogues.

Pourtant, la situation française est atypique puisque le format national d'échange (UNIMARC) n'est pas le format de la BnF (INTERMARC), qui est pourtant l'agence nationale de catalogage. La BnF propose tous ses produits aussi bien en INTERMARC qu'en UNIMARC,⁶⁰ mais le fait que la France n'utilise pas le format de sa Bibliothèque nationale est une gêne pour le développement, certes peu visible par le grand public. De fait l'INTERMARC, créé en 1975 et remanié en 2002, devait à l'origine devenir un format européen, mais est en pratique essentiellement le format de travail de la BnF (qui en assure la maintenance) et de quelques autres bibliothèques françaises.⁶¹

Au niveau international, l'OCLC⁶² (dont les choix deviennent souvent des standards de fait) ainsi que la plupart des bibliothèques anglo-saxonnes utilisent un autre format, le MARC21. Il descend directement du plus ancien format MARC (1965), mais est issu d'un travail de fusion des multiples formats existants. entre 1997 et 1999, l'uniformisation des formats des Etats-Unis, du Canada du Royaume Uni et de l'Australie⁶³ aboutit à la publication du MARC21, format d'échange reconnu par l'IFLA. La Bibliothèque du Congrès en assure la maintenance,⁶⁴ avec le MARBI.⁶⁵

Les deux principaux formats utilisés aujourd'hui sont donc le MARC 21,⁶⁶ et l'UNIMARC.⁶⁷ Etant un produit Anglo-saxon, le MARC21 est très lié aux AACR2.⁶⁸

⁵³ Voir Annexe 1, l'adaptation des ISBD en UNIMARC.

⁵⁴ 5^e édition française : Fédération internationale des associations de bibliothécaires et des bibliothèques; Comité français UNIMARC. *Manuel Unimarc: format bibliographique*. München: KG Saur, 2007, et mise à jour sur le site de la BnF, <http://www.bnf.fr/pages/infopro/normes/no-unimarcb.htm>

⁵⁵ Permanent Unimarc Comitee, depuis 1991. <http://www.ifla.org/VI/8/puc.htm>

⁵⁶ Le programme ICABS de l'Ifla, IFLA-CDNL Alliance for Bibliographic Standards, <http://www.ifla.org/VI/7/icabs.htm>.

⁵⁷ Système Universitaire de DOCUMENTATION : le catalogue des bibliothèques de l'enseignement supérieur, <http://www.sudoc.abes.fr>

⁵⁸ Comité Français UNIMARC, participant au Permanent Unimarc Comitee. Le CFU a son propre site Web, <http://cfu.enssib.fr>, et répond aux questions des utilisateurs.

⁵⁹ Aussi bien notices bibliographique que d'autorité, <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancre=normes/no-acuni.htm>

⁶⁰ Et en ISO 5426 (écriture en caractères latins et diacritiques) ou en UNICODE UTF8 (toutes écritures) pour le codage des caractères.

⁶¹ Notamment quelques bibliothèques de prêt de Paris, et bibliothèques musicales.

⁶² Online Computer Library Center, le plus grand réseau de catalogage du monde, <http://www.oclc.org>

⁶³ USMARC, AUSMARC, CANMARC et UKMARC. Le MARC 21 correspond techniquement à la fusion de l'USMARC et du CANMARC, et s'est étendu plus largement, notamment au Royaume-Uni avec l'abandon de l'UKMARC en 2004.

⁶⁴ LOC, à Washington, <http://www.loc.gov/marc/>.

⁶⁵ Comité informatique Marc, employant environ 20 personnes à temps plein.

⁶⁶ Bibliothèques anglo-saxonnes, notamment OCLC.

⁶⁷ Que l'IFLA recommande et maintient, en lien avec l'ISBD, utilisé en France, Italie et Allemagne.

⁶⁸ et au RDA en cours d'élaboration.

C'est notamment un format associé aux AACR2 pour décrire des archives et des objets tridimensionnels. La plupart des SIGB actuels sont conçus pour le Marc 21.

Bien que la structuration des zones varie entre ces formats, la structure reste comparable, avec deux niveaux hiérarchiques,⁶⁹ ce qui est trop peu pour effectuer des descriptions fines, nécessaires face à la multiplication des supports et à l'évolution des techniques de recherche. Ils semblent en particulier désuets face aux langages à balises.

- Interopérabilité et évolutions

Agés d'une trentaine d'années, les formats Marc sont désuets. Complexes et lourds à gérer, ils manquent de souplesse, avec deux niveaux hiérarchiques seulement. Conçus pour les bibliothèques, ils posent des problèmes de visibilité et d'ouverture sur le Web, mais aussi d'interopérabilité, entre eux⁷⁰ et surtout avec les autres types de description de ressources, adaptés notamment aux moteurs de recherche.

Le protocole Z 39-50 (correspondant à la norme ISO Z 39-50) permet ainsi d'interroger les bases de données structurées en Marc, quel que soit le type de format Marc. Il est utilisé pour dériver des notices déjà produites dans un autre catalogue.⁷¹ Il sert aussi à fédérer des recherches dans plusieurs catalogues simultanément. Intégré à une architecture Client/Serveur sur des bases distantes, il permet de rendre des données interrogeables.⁷² Mais le protocole Z39-50 pose plusieurs problèmes. Sa lourdeur le rend lent et coûteux. Surtout, il est conçu exclusivement pour le MARC et est peu adapté à l'environnement Web. C'est pourquoi il est peu à peu remplacé par le SRU, un protocole XML qui en reprend les fonctionnalités. En effet, le passage à des structures XML est une des questions essentielles pour l'avenir des catalogues à court terme.

2.2.2 MARC et structures XML

Si ces structures sont utilisées pour l'échange de données, voire pour des indexations ou certains produits spécifiques, elles ne sont pourtant pas encore utilisées par les systèmes locaux des bibliothèques.

- Une technologie prometteuse

XML est un langage à balises, sur le principe de SGML⁷³ et de son dérivé Html, introduisant dans un texte des balises signifiantes sur le contenu, pour le structurer. Elaboré dans le contexte du Web à la fin des années 1990, il présente l'avantage de dissocier le contenu et la mise en forme (contrairement au Html).

Etant un standard élaboré par le W3C,⁷⁴ XML permet aux applications d'être indépendantes et connues. L'absence de discrimination entre le contenu structuré et le contenu non structuré, rend son utilisation souple. On peut en effet l'utiliser et partager

⁶⁹ Zone (« champ ») en 3 chiffres, et Sous-Zone (« Sous-champ ») avec code de sous-zone.

⁷⁰ Le format Marc devait servir de format pivot, mais il faut utiliser des protocoles d'échange...

⁷¹ Par exemple dans le SUDOC avec WinIBW permettant ce type de dérivation.

⁷² Par ailleurs, il peut être couplé avec un agrégateur Web, même s'il n'était au départ pas conçu pour un protocole http.

⁷³ Standard Generalized Markup Language, langage à balises normalisé en 1986 avec l'ISO 8879.

⁷⁴ voir notamment <http://www.w3c.org>.

des données entre différents services même si les systèmes sont différents, et partager des métadonnées à condition qu'elles soient interoperables. Il est même possible de l'utiliser pour décrire des structures sophistiquées (RDF),⁷⁵ ce qui impliquera probablement des changements sur le Web dans les années à venir.

XML est déjà largement utilisé pour gérer des contenus de documents numériques.⁷⁶ Il s'agit de gérer les informations de métadonnées, et d'intégrer le contenu dans d'autres applications, aussi bien pour des services Web (comme le Webdav),⁷⁷ que pour des services de gestion de *workflow* (actions, plannings, responsables...) ou de Gestion Electronique de Documents (durée d'utilité administrative, délais de communicabilité, d'élimination...),⁷⁸ pour la gestion de documents d'archives. Les schémas élaborés dans ce contexte sont très utiles au monde des bibliothèques. Son intérêt est tout à fait reconnu dans l'administration française, puisque la DGME⁷⁹ en recommande l'utilisation pour l'échange et l'archivage des documents électroniques.

Les aspects intéressant à court terme les catalogues de bibliothèques ont donné lieu à un travail important dans la profession, aussi bien sur les principes généraux que sur leurs applications pratiques, avec par exemple « XML pour les bibliothécaires ».⁸⁰

- Echanger, indexer, décrire

Les balises XML qualifient le contenu. Les seules contraintes sont de fermer toutes les balises ouvertes et de respecter l'agencement hiérarchique. Dans le contexte des technologies du Web, l'utilisation de XML suppose l'emploi de feuilles de styles (XSLT par exemple) et d'un navigateur (la machine, Parser) permettant la visualisation pour les humains. Surtout, il faut employer un schéma ou une DTD,⁸¹ c'est-à-dire une liste des balises possibles et nécessaires pour la description. L'établissement de DTD standardisées est un problème essentiel. Pour l'adaptation des notices de catalogues, il s'agit à la fois de répondre aux changements techniques, et aux changements d'utilisation.

Ainsi, les MARC sont des formats d'échange, linéaires, conçus pour des bandes magnétiques en série dans un répertoire, pas vraiment géré en machine au sens actuel. En fait, dans bases de données en MARC, les champs de bases de données conservent les attributs MARC. Par conséquent, le passage au MARC XML (ou à d'autres outils XML) peut être naturel, mais les balises restent celles du MARC.

Le MARC XML est extensible, mais est une simple transcription du MARC qui conserve deux éléments seulement, ce qui permet qu'il soit lu hors des bibliothèques mais n'améliore pas les niveaux hiérarchiques, qui étaient pourtant une des faiblesses principales du MARC. De ce fait, les premières fonctions du MARC, c'est-à-dire la

⁷⁵ Ressource Description Framework, *Infra*, 4.2.2.

⁷⁶ Dont il permet la réutilisation, avec des mécanismes de type Object Linking and Embedding (OLE).

⁷⁷ Webdav (pour World Wide Web Distributed Authoring and Versioning), projet développé par l'IETF et le W3C, est un protocole applicatif défini au-dessus de http et utilisant une syntaxe XML, protocoles et métadonnées permettant l'édition coopérative et la gestion de contenu de documents sur le Web.

⁷⁸ Voir notamment PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique documentaire. Paris: Dunod, 2004, 342 p., p.328.

⁷⁹ Direction Générale pour la Modernisation de l'Etat s'intéresse notamment aux standards administratifs

⁸⁰ *Getting started with XML, a manual and workshop*, Par Eric L. Morgan, <http://www.infomotions.com/> Traduction par Nicolas Morin, <http://www.nicolasmorin.com/xml/>.

⁸¹ Document Type Definition.

description, l'échange et la dérivation, sont conservées avec les formes actuelles de passage en XML, qui permettent de simplifier les échanges dans le contexte du Web. Ainsi la récente norme MarcXchange, (ISO 25577) permet avant tout l'échange de fichiers en XML. Par exemple, le SUDOC fournit ses notices à Google pour Google Books, et de nombreux SIGB proposent des fonctions d'Import/Export en XML en plus du MARC, pour faciliter les échanges même sans fonctions nouvelles.

L'abandon des formats MARC supposerait de créer des données directement en langage XML. En changeant de formats pour s'adapter au Web, on garde la fonction première des catalogues, avec l'identification, la localisation et le recensement des documents, et avec une qualité permettant la pérennité de l'outil.

Ce passage à des descriptions en XML est déjà effectué pour une partie des bibliothèques numériques. En effet, les documents numériques tirent en avant les évolutions techniques : ils ont provoqué le passage au XML des logiciels de gestion d'archives pour la GED, et encouragent les bibliothèques à faire de même, bien qu'il s'agisse pour le moment surtout d'adapter XML aux MARC.

La mise en place du langage XML en bibliothèque a fait l'objet de réflexions au début des années 2000, à la fois en raison de l'intégration de documents numériques et de l'adaptation au Web. Les DTD élaborées dans ce contexte correspondent en fait à la transposition des divers MARC.

- Un usage spécialisé

Du fait de l'importance de l'OCLC,⁸² les standards établis pour les Etats-Unis prévalent dans le monde entier. L'OCLC ayant fait le choix de passer au Marc XML, ce format a acquis une valeur de standard pour le passage des MARC à d'autres schémas de métadonnées exprimables en XML. MarcXML, élaboré par la Bibliothèque du Congrès à partir de 2000, permet de convertir en schéma XML des enregistrements Marc21.

Diverses autres DTD adaptées aux MARC existent.

La plus générale, ISO 12083 est une DTD bibliographique élaborée dès 1994, pour le catalogue des monographies, des périodiques et des articles de périodiques.

Les autres reflètent la diversité des formats de départ.

Les principales utilisées sont :

- La DTD BiblioML, élaborée en 1999 par la Mission de Recherche et de Technologie du Ministère de la Culture pour utiliser l'UNIMARC en XML.⁸³

- La DTD ONIX (ONline Information eXchange), créée en 2000 par l'Association des éditeurs américains, utilisée notamment par le groupe EDItEUR, pour fournir aux éditeurs une information enrichie dans une perspective de marketing. Le Cercle de la librairie l'utilise pour la base Electre.⁸⁴ L'ONIX permet l'intégration d'informations enrichies (image de couverture, liens vers d'autres sites).

- D'autres DTD de bibliothèques, comme MODS/MADS de la Bibliothèque du Congrès, basée sur MARC21, pour gérer de collections physiques et numériques,⁸⁵ et XOBIS (XML Organic Bibliographic Information Schema). Très novatrice, XOBIS est une

⁸² Online Cataloging Library Center, <http://www.oclc.org>

⁸³ Outil de conversion disponible sur <http://www.biblioml.org/fr/conversion.html>.

⁸⁴ Yves DESRICHARD, « Vers la convergence des formats bibliographiques ? ONIX, application XML du monde de l'édition », dans Bulletin des Bibliothèques de France, t. 49/5 (2004), p. 55-63. <http://bbf.enssib.fr>.

⁸⁵ Cf. Infra, partie 4.

expérimentation isolée de la Lane Medical Library (Stanford University, en Californie) mise en place en 2002, avec une structuration de l'information biobibliographique distincte des ISBD. Cette idée de s'affranchir des anciens formats est encore délicate à mettre en place, avec des DTD autres que MARCXML.

Ces différents schémas rendaient nécessaire l'utilisation d'une norme permettant l'échange. La récente norme MarcXchange (future ISO 25577 en cours de validation, actuelle DIS ISO 25577), qui est l'aboutissement de travaux menés depuis 2003, permet de standardiser les échanges en XML, mais ne vise pas à permettre plus que l'échange.

On pourrait avoir l'impression qu'il s'agit de poupées russes, avec un format d'échange entre formats d'échange. Mais cela correspond à la rationalisation d'une situation involontairement éclatée, qui est éclaircie par les nouveaux outils d'échange, avec souvent la difficulté locale de l'adaptation des systèmes de gestion.

2.3 DES SYSTEMES DE GESTION DE BIBLIOTHEQUE

En bibliothèque, le catalogue est un instrument de visibilité des collections. Il a la double vocation de présentation des collections et de gestion d'un fonds, depuis les *Tables* de Callimaque⁸⁶ d'Alexandrie jusqu'à l'OCLC-Worldcat.⁸⁷ Les changements techniques induisent de nouvelles modalités de gestion et d'accès.

2.3.1 Des outils techniques

Le passage à l'informatique avait déjà posé un problème de redéfinition des instruments de recherche. Ces changements continuent avec l'intégration au Web.

- L'origine de l'informatisation

L'informatisation des bibliothèques a commencé dès les années 1970, avec les premiers systèmes utilisant des ressources informatiques centralisées, quand la municipalité ou l'université dont relevait la bibliothèque disposait d'un service informatique. Il s'agissait d'un *mainframe*, machine centrale permettant toutes les applications avec un traitement différé (par bandes perforées puis disquettes), par exemple à la Part-Dieu dès 1972. Sans revenir dans le détail sur cette première informatisation, il faut remarquer que les acquisitions, le catalogage et les prêts constituaient des modules différents, avec une séparation nette des fonctions de gestion et de signalement, qui correspondait d'ailleurs au fonctionnement antérieur.

Puis, le développement des mini-ordinateurs moins onéreux dans les années 1980 permit la mise en place de *progiciels métier* élaborés par des sociétés de services pour les bibliothèques publiques. Ces premiers Systèmes Intégrés de Gestion de Bibliothèque intégraient toutes les fonctions, avec pour principal avantage de proposer une base de données unique pour tous les modules et toutes les fonctions. Outre le fonctionnement

⁸⁶ Citées par Christian Jacob, « Lire pour écrire : navigations alexandrines », In JACOB, Christian. *Lieux de savoir. [Volume 1], Espaces et communautés*. Paris: A. Michel, 2007, p. 61 sqq.

⁸⁷ Avec 100 millions de notices, 1,2 milliard de localisations, 9000 dans 90 pays différents, <http://www.worldcat.org> est un méta-catalogue élaboré par l'OCLC et regroupant de nombreux catalogues, dont désormais Bn Opale Plus et le SUDOC.

en temps réel dans un système d'exploitation classique, ces SIGB rendirent nécessaire une première normalisation des notices.

En particulier, les formats MARC furent intégrés aux SIGB, qui disposent habituellement dans leur base bibliographique de trois types de fichier :

- Un fichier de notices bibliographiques (en format Marc ou format compatible)
- Un fichier de notices d'autorité (en format Marc ou format compatible)
- Un fichier de notices exemplaires (en format du logiciel).

Ainsi, si l'informatisation des catalogues s'est imposée aux fournisseurs de SIGB, il faut remarquer que désormais leur fonctionnement de SGBD relationnelles conçues pour le MARC rend les évolutions difficiles.

- L'accès sur le Web

Déjà avant le développement d'Internet, l'échange de données entre SIGB avait été mis en place dans les années 1970 et 1980, avec le réseau Transpac X25. Pourtant, au cours des années 1990, Internet provoqua plusieurs changements, comme l'intégration d'images et de documents multimédia, et surtout l'utilisation de serveurs Web avec des bases de données centralisées, en ligne.

Par conséquent, les modules de consultation de notices des SIGB purent être accessibles sous la forme d'OPACs, Online Public Access Catalogs. Les SIGB sont devenus à cette occasion un module en *back-office*. C'est-à-dire que, dans la mesure où il s'agissait d'un outil dévolu à la gestion interne de la bibliothèque, il se dissocia intellectuellement du catalogue lui-même, manifestation du SIGB visible du public.

A l'exception de la consultation sur place en libre accès, le catalogue en ligne devint la forme normale d'accès aux collections. C'est pourquoi les divers éléments techniques de description ont dû être appréhendés sous une forme électronique, comme les éléments d'une base de données.⁸⁸ Ce relatif flou dans les objets et termes désignant les instruments de recherche est encore accentué par l'intégration au Web, qui rend peu efficace de concevoir la partie « Catalogue » du SIGB comme un outil autonome.

- L'intégration au Web

La bibliothèque peut proposer plusieurs types de services sur le Web : interface Web du catalogue, mais aussi d'autres ressources consultables en ligne comme des CD-Rom ou des Bases de Données en ligne d'éditeurs. Des moteurs de recherche fédérés peuvent donc permettre de trouver simultanément des résultats dans les divers types de ressources. Ces moteurs s'intègrent dans le SID, système d'information documentaire.⁸⁹ Pour les systèmes modulaires, intégrant les éléments de constitution des catalogues, il est donc absolument nécessaire de prendre en compte l'interopérabilité entre le SID et le SIGB d'une part, et entre l'ENT, le SI et les portails d'autre part. Du moins, si l'on souhaite que le catalogue puisse être intégré dans des outils plus larges.

⁸⁸ PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique documentaire. Paris: Dunod, 2004, 342 p., p.282, pour les « registres, fiches, catalogues, formulaires, procédures, rapports ... »

⁸⁹ Par exemple, la Jubilothèque de Paris 6-Jussieu donne accès à toutes les ressources disponibles, et éventuellement aux ressources sous droits après identification. <http://jubil.upmc.fr>

En outre, l'intégration dans des catalogues collectifs permet une plus grande visibilité, en diminuant considérablement les coûts de production, les notices élaborées étant collectives. En France, l'exemple principal en est le SUDOC,⁹⁰ catalogue collectif du réseau géré par l'ABES⁹¹ et regroupant principalement des Bibliothèques Universitaires. Pour les membres, il s'agit à la fois d'être repérables depuis le catalogue commun, et d'utiliser des notices déjà réalisées, en se localisant simplement si elles sont déjà dans le catalogue collectif, et en les dérivant si elles ont été élaborées par une institution partenaire. Pour des raisons évidentes d'économie de travail humain et d'uniformisation des descriptions, la localisation simple est considérée comme une forme souhaitable de catalogage. Il s'agit d'une exemplarisation dans le catalogue collectif, donnant lieu automatiquement à la création d'une notice dans le SIGB, avec l'éventuel écrasement des notices locales pour l'uniformisation.⁹² A l'échelle mondiale, un catalogue collectif regroupant toutes sortes de bibliothèques est développé par l'OCLC : Worldcat.⁹³

Les OPACs traditionnels présents sur le Web étant des bases de données interrogeables grâce à un formulaire, ils ne peuvent être indexés par les robots des moteurs de recherche. Ces catalogues font partie du Web invisible, c'est-à-dire des données disponibles mais difficilement accessibles car non repérées par les moteurs de recherche.

La plupart des organismes bibliographiques –même la puissante OCLC- acceptent l'idée que les moteurs de recherche classiques doivent désormais être considérés comme un moyen normal d'accéder aux ressources, les catalogues spécifiques n'en étant qu'un complément.⁹⁴ L'enjeu de la description des ressources est donc le passage de la description dans un catalogue aux métadonnées.⁹⁵

2.3.2 Des moyens de gestion

En une trentaine d'années, les catalogues informatisés sont passés d'un outil divisant les fonctions de gestion et de signalement, à des SIGB gérant tout, de l'achat au prêt aux lecteurs, puis à des systèmes modulaires. Les moyens de gestion y sont associés.

- Le circuit du document

Dans le cas des ouvrages de lecture publique, le document est acquis, avec une commande et sa réception, et fait à ce moment son entrée dans le catalogue (avec une description, dans le module d'acquisition du SIGB, donnant éventuellement un statut temporaire pendant la commande). C'est l'occasion de son indexation (avec un accès sujet, et une cote de classification). L'exemplarisation, la cotation, l'équipement, la mise en circulation, et le désherbage correspondent à la mise à jour du catalogue.

⁹⁰ <http://www.sudoc.abes.fr>

⁹¹ Agence Bibliographique de l'Enseignement Supérieur, EPA créé en 1994

⁹² Par exemple catalogue dans SUDOC avec WinIBW puis la création automatique de notices locales.

⁹³ <http://www.worldcat.org>, auquel participent pour le moment peu de bibliothèques Françaises (dont la BnF). Le *Hathi trust* amène à s'interroger sur les intégrations, voir <http://www.bibliobsession.net/2008/10/23/vers-un-catalogue-mondial-des-bibliotheques-et-moi-et-moi-et-moi/>

⁹⁴ M. YEE, Martha; SHATFRD LAYNE, Sara, *Improving public access catalogs*, ALA editions, 1998, 220 p.

⁹⁵ Voir Lupovici, Catherine, « Du catalogue informatisé à la navigation dans l'information numérique » *In Revue de la Bibliothèque nationale de France*, no 9, 2002, p. 80-82.

La plupart des bibliothèques utilisent des formats régis par des normes, pour les catalogues informatisés. Pour la cohérence et la qualité du catalogue, il faut permettre l'accès aux données par les lecteurs, et les échanges, à la fois pour l'interopérabilité et pour la récupération de notices. En outre, la conservation de données suppose la pérennité des formats et des données.

Avec le développement des catalogues collectifs, et l'habitude de chercher des notices déjà produites (comme celles de la BnF avec le dépôt légal),⁹⁶ les petites bibliothèques non responsables des notices bibliographiques et d'autorité interviennent seulement sur troisième type de notices : les notices exemplaires, avec les données locales. Il s'agit des fonctions de localisation, avec une cote, un code barre, et d'informations différentes comme l'état de disponibilité, l'état de collection, et des données d'exemplaires éventuelles. Avec le développement d'outils centralisés de production de notices, le travail intellectuel de catalogage au niveau local tend à se limiter aux informations liées à la gestion du document.

- Des fonctions de gestion

Ces fonctions sont habituellement prises en charge, par des modules du SIGB. Il s'agit principalement du « catalogage » et de l'« OPAC/CIEL ». Selon les produits, il existe aussi des modules pour le bulletinage, la circulation ou les commandes par exemple. Cette dissociation permet que les diverses tâches d'indexation, de recherche, de déchargement, d'affichage et d'impression par exemple, donnent lieu à des traitements distincts. La gestion des données catalographiques est donc clairement répartie.

Ces fonctions de gestion tendent à s'étendre, puisque les bibliothèques numériques amènent à gérer des documents dits « virtuels », en fait virtuels mais numériques, en lien avec les outils XML. Il s'agit de réutiliser les données bibliographiques pour différentes fonctions, et de gérer le document.⁹⁷ Il devient possible d'intégrer des fonctions de gestion des documents, avec les tâches proches de celles du *records management*⁹⁸, transposables en bibliothèque avec les droits d'auteur, l'obsolescence et le désherbage. Dans ce domaine ce sont les bibliothèques numériques qui conduisent à innover.⁹⁹

Ces fonctions peuvent être étendues à la gestion des documents papiers, dans des domaines où l'obsolescence des documents est rapide, ce qui permet de programmer le désherbage dès l'acquisition,¹⁰⁰ avec par exemple une adaptation de la grille Conspectus pour programmer par exemple l'élimination d'une partie des manuels techniques dès leur achat. Il peut s'agir aussi d'outils d'évaluation des collections, ou de renvois entre les ouvrages.

⁹⁶ Voir « Des métadonnées pour bien utiliser les ressources électroniques », Journée d'information AFNOR/CG46, 7 juin 2005

<http://www.bnf.fr/pages/infopro/journeespro/no-Afnor2005.htm>

⁹⁷ Par exemple pour SPAR, la BnF utilise des notices OAI pour créer les fichiers METS.

⁹⁸ Statut de communicabilité, durée d'utilité administrative, et élimination. Voir RIETSCH, Jean-Marc, et al. *Dématérialisation et archivage Électronique : mise en oeuvre de l'ILM (Information Lifecycle Management)*. Paris : Dunod: 01 informatique, 2006.

⁹⁹ Comme le système SOLON de la BnF qui gère les droits d'accès aux documents numérisés, avec des notices d'autorité et un encodage des durées légales, par exemple pour les auteurs morts pour la France.

¹⁰⁰ Voir la fiche pratique « désherber en bibliothèque » par Stéphanie David, <http://www.enssib.fr/bibliotheque-numerique/document-1735>, mentionnant l'intégration du désherbage au plan de développement des collections

Ces fonctions pourraient amener les catalogues à intégrer des fonctions de gestion plus développées, ou en tout cas plus liées au contrôle des documents.

2.3.3 Une inertie à prendre en compte

Les évolutions sont en partie conditionnées par les SIGB existant sur le marché.

- Contrats et habitudes

Même avec une vision théorique claire des évolutions souhaitables, les héritages techniques limitent les choix possibles. Les milliards de notices conservées en ISO 2709 depuis les années 1960, impliquent de continuer à exploiter les données elles-mêmes. C'est ainsi que la plupart des SIGB actuels utilisent le Marc 21, puisqu'ils sont en général élaborés aux Etats-Unis. Ainsi, certains SIGB ne supportent que le format Marc 21.

Outre l'inertie liée aux contrats conclus, et la lourdeur qu'implique un changement de fournisseur de SIGB, le problème principal est celui de l'inertie des produits du marché. Malgré le développement de SIGB libres¹⁰¹, et l'action d'organismes comme l'OCLC pour le développement d'outils autonomes, les bibliothèques françaises restent en général dans le cadre de progiciels commerciaux.

- Des changements forcés

La plupart des firmes des fournisseurs de SIGB ont été rachetées, au point que seules trois ou quatre sociétés se détachent vraiment avec des chances de rester indépendantes. Ces difficultés, pour lesquelles les circonstances économiques actuelles laissent peu d'espoir, ont des conséquences directes sur le maintien des SIGB en MARC et sur l'absence de développements. Certains systèmes comme Aleph¹⁰² sont en XML, mais il ne s'agit pas pour le moment d'une évolution générale. Certains changements sont pourtant nécessaires pour la gestion de documents numériques. C'est ainsi que depuis 2002, plusieurs SIGB¹⁰³ permettent l'intégration de collections numériques, avec des systèmes différents. Cette évolution vers une bibliothèque hybride¹⁰⁴ provoque l'évolution des instruments de recherche et de gestion.¹⁰⁵ On se heurte en particulier à l'ambivalence entre une production bibliographique de plus en plus adaptée aux documents numériques, et une masse de données valorisables en Web, avec une structure sémantique (à la fois pour les auteurs et les notices bibliographiques). En d'autres termes, les possibilités techniques offertes par le Web laissent entrevoir des

¹⁰¹ Par exemple KOHA, adopté en France par la bibliothèque de l'école des Mines, mais aussi Evergreen et PMB. Voir Déployer des solutions Open source dans les bibliothèques, centre de documentation et services d'archives, 13 novembre 2007, http://www.bnf.fr/pages/infopro/journeespro/jp_opensource.htm, SCOPSI, Claire ; SOUAL, Laurent ; FERRAILLE, Jean-François. *Mener un projet Open Source en bibliothèque, documentation et archives*. Paris: Éd. du Cercle de la librairie, 2007.

¹⁰² Par la société Israélienne Ex-Libris

¹⁰³ Notamment Aleph (Ex Libris), Koha, Serials Solutions (SirsiDynix, avec l'intégration d'outils de gestion existants), Encore (Innovative Interfaces), Primo (Ex Libris).

¹⁰⁴ Voir BROPHY, Peter, « La bibliothèque hybride », *BBF*, 2002, n° 4, p. 14-20. <http://bbf.enssib.fr>.

¹⁰⁵ Le Joint Information Systems (JISC) britannique a ainsi publié un rapport sur les différents SIGB employés, <http://www.jisc.ac.uk/media/documents/programmes/resourcediscovery/lmsstudy.pdf>, en particulier « 4.2.2 Electronic Resource Management (ERM) systems », p. 56,

modes de mise en valeur des données auxquelles la production bibliographique habituelle n'est pas prête. De cette façon, un groupe de réflexion réuni en 2008 à Fairbairn (Université du Connecticut) pour réfléchir à la nouvelle génération d'OPACs, concluait par un simple encouragement à renouveler les interfaces graphiques.¹⁰⁶ Du moins pour les petites bibliothèques, cet encouragement à se limiter à des outils locaux correspond à la fois à une perspective économique, à une confiance en l'action d'OCLC pour les travaux bibliographiques de grande envergure, et à l'idée que la diffusion de l'information sur le Web se passe de l'action des bibliothèques.

La question des héritages et de la demande est importante à la BnF, puisqu'il s'agit à la fois d'encourager à l'amélioration des systèmes des bibliothèques françaises, et de répondre à leurs attentes comme consommatrices de notices bibliographiques et d'autorité. La réflexion qui y est actuellement menée sur l'évolution des catalogues correspond donc à la conciliation d'un système ancien (mis en place dans les années 1990), de demandes évoluant peu, et de nécessaires évolutions liées aux bouleversements techniques.

Au niveau mondial, les restructurations des sociétés fournissant des SIGB, et la diminution des ressources des bibliothèques publiques, dans un contexte de difficultés économiques, pourraient conduire à un nouveau modèle d'élaboration des progiciels de gestion de bibliothèque. De fait, l'OCLC, qui occupe déjà une position hégémonique dans le monde des bibliothèques, travaille sur le développement de SIGB échappant à la logique des fournisseurs privés. Il ne s'agit que d'hypothèses, mais cette évolution aurait des conséquences sur les modèles économiques et sur l'évolution des catalogues.

2.4 DES SYSTEMES D'INTERROGATION

Le catalogue informatisé est conçu comme un outil de réponse à des questions, formulées par les bibliothécaires puis par les usagers. Les points d'accès permettant de trouver l'information sont donc essentiels. L'élaboration de schémas de données se traduit par une indexation des champs de la notice, pour permettre la recherche.

Pour l'utilisateur, l'affichage à l'écran est une réponse présentée au départ par analogie avec les fiches, puis par champs principaux. L'accès aux données parvenant peu à peu à s'affranchir des anciennes contraintes, il devient possible de mettre en place de nouveaux modèles, pour répondre à la lourdeur des catalogues et à l'immensité du Web.

2.4.1 Indexer les champs

On parlait au départ d'indexation pour désigner la production d'index papier.¹⁰⁷ Avec le traitement informatique en bases de données, la création de tables d'index prit un sens nouveau. Pour autant, le principe d'élaboration restait de devancer les futures requêtes en établissant les points clairs reliant les différents documents.

L'actuelle indexation automatique du Web par les moteurs de recherche, constituant des bases de données dans lesquelles s'effectue la recherche, donne encore un autre sens aux index. Les bibliothèques doivent s'y intéresser, à la fois pour s'y intégrer et afin d'adapter des procédures similaires pour le traitement des documents numériques.

¹⁰⁶ MARCIN, Susan; MORRIS, Peter. "OPAC: the next generation: placing an encore front end onto a SirsiDynix ILS". In *Computers in libraries*. Westport, CT: Meckler, 1989-, Mai 2008.

¹⁰⁷ Avec des fichiers regroupant noms de personnes, lieux ou sujets, et renvoyant aux ouvrages.

- Une expertise ancienne

L'indexation de données numériques est proche du travail sur fiches papier, avec des points d'accès normalisés.¹⁰⁸ Il s'agit des éléments permettant de trouver la notice bibliographique, qui se répartissent en trois ensembles :

- Les éléments de description bibliographique, dont l'ISBN (ou ISSN) et le titre propre.
- Les éléments que le catalogueur construit.
- Les éléments normés (Par exemple l'auteur,¹⁰⁹ ou le lieu d'édition¹¹⁰).

Ces éléments peuvent donner lieu à des renvois vers les notices les contenant aussi, par exemple des autres éditions du titre ou des autres ouvrages du même auteur, et même des ouvrages sur un même sujet. Bien avant l'informatisation, il existait des « fiches de regroupement » dans les fichiers de bibliothèques. Il s'agissait de fiches intercalaires permettant le renvoi entre les notices.¹¹¹ Dans les systèmes informatisés, il est possible d'effectuer automatiquement le renvoi que permettaient ces fiches. Il s'agit aussi bien de passer d'une forme rejetée vers une forme retenue, que de trouver les notices liées à cette forme. La notice d'autorité uniforme et validée est essentielle.

- Valider les termes

Les autorités validées sont la seule façon de résoudre la question des homonymes, d'établir des titres uniformes, et de décrire les contenus dans un langage unifié. Peu d'institutions valident des notices d'autorités auteurs : il s'agit essentiellement de la Library Of Congress, et de la BnF en France. En France, les notices d'autorité sont créées et modifiées uniquement par les catalogueurs de la BnF et ceux du SUDOC, mais sont utilisées partout pour éviter les ambiguïtés et permettre les renvois. Pour que les utilisateurs du catalogue puissent effectuer des recherches par sujet sans se limiter aux titres et auteurs, il faut une indexation par sujet décrivant le contenu des documents. Il est nécessaire d'utiliser des termes précis, qui peuvent être définis dans un index limité, sous forme de liste hiérarchisée (avec des relations entre les termes génériques, spécifiques est associés), ou dans un thésaurus (décrivant des relations sémantiques spécifiques pour un domaine concerné). Ce qui permet de compléter la description bibliographique par des points d'accès utilisant un vocabulaire contrôlé, avec des listes comme la Blanc-Montmayeur,¹¹² des listes d'autorités comme Rameau¹¹³ ou Motbis.¹¹⁴ Les principes d'indexation manuelle des sujets sont parfois remis en cause, face aux modes d'indexation automatique des moteurs de recherche, de la recherche en plein texte. L'indexation par vedettes matière garantit pourtant la précision de la description des contenus. Surtout, l'évolution des habitudes de recherche vers une formulation de mots-clés contenus dans le document rend plus que jamais nécessaire de rendre le document accessible par les termes de son sujet, et non simplement par titre ou auteur.¹¹⁵

¹⁰⁸ Notamment l'IME-ICC (IFLA Meetings of Experts on an International Cataloguing Code), 2007 et 2008, voir http://www.ifla.org/VII/s13/icc/imeicc-statement_of_principles-2008_french.pdf

¹⁰⁹ Z44-059.

¹¹⁰ Z44-060.

¹¹¹ Cette technique a été utilisée par exemple pour l'établissement du Catalogue général des imprimés de la BnF. *Catalogue général des livres imprimés...* Paris: Bibliothèque nationale, 1972-1978.

¹¹² Utilisée notamment en bibliothèques de lecture publique.

¹¹³ Répertoire d'autorité matière encyclopédique, alphabétique et unifié

¹¹⁴ Pour les sciences de l'éducation, <http://www.motbis.fr>.

¹¹⁵ Voir les deux journées sur l'indexation et le lien des catalogues et SIGB avec l'indexation et les vedettes matières : Réunion d'information- débat des utilisateurs des produits et services bibliographiques, 29 mai 2008, http://www.bnf.fr/pages/infopro/journeespro/jp_produits2008.htm.

- L'indexation par des robots

L'idée de devancer la recherche des utilisateurs en décrivant les ressources avec des termes prend un sens nouveau avec les moteurs de recherche. On peut désormais distinguer deux types d'indexation : intellectuelle et physique, qui prennent la forme de l'indexation manuelle, et de l'indexation automatique par des robots, de plus en plus importante, au point de tenir lieu de catalogue. Ces robots, de façon schématique, se comportent comme un internaute en ouvrant tous les liens possibles sur les sites et en y « moissonnant » les termes signifiants, indexés automatiquement. Les moteurs de recherche ont donc deux activités différentes :

- D'une part, un robot d'archivage qui parcourt le Web en suivant les liens et en copiant des données.

- D'autre part, un robot d'indexation permettant l'indexation en plein texte.

C'est le cas par exemple avec Google pour traiter la masse du Web, avec une indexation et l'utilisation d'un algorithme de traitement des résultats très puissant et connu dans ses grandes lignes, le Page Rank.¹¹⁶ C'est la seule solution efficace, au point que cette méthode est transposée pour les outils comparables. Par exemple, le service du Dépôt Légal du Web à la BnF utilise principalement un robot d'archivage, mais pour créer un instrument de connaissance des collections (comme la Bibliographie nationale française pour les ouvrages), le Web collecté est indexé en plein texte.

Cette indexation automatique n'exclut pas l'emploi de termes normalisés, avec un cluster sémantique permettant de dissocier les différents sens des termes polysémiques (comme le fait Wikipédia par exemple),¹¹⁷ et de relier des « grappes de concepts » contribuant à un sujet d'intérêt principal.¹¹⁸ Cette idée de préciser le sens des termes indexés automatiquement est liée à la technologie des traductions automatisées, du moins dans les développements et améliorations envisagés par les principaux moteurs.¹¹⁹ Il est donc probable que l'indexation des contenus des objets numériques progresse dans les années à venir. Les techniques d'indexation des moteurs de recherche, constituant des bases de données séparées avec les termes indexés automatiquement, a une double conséquence. D'une part les utilisateurs prennent l'habitude d'effectuer des recherches simples par mots-clés, et d'autre part l'indexation automatique dans des contenus non normalisés rend nécessaire d'adapter les catalogues dans le Web. Il s'agit de proposer une granularité des descriptions correspondant à la précision de l'information et aux types d'accès nécessaires, selon les domaines et institutions. Il s'agit aussi de garantir la pertinence des points d'accès, et de permettre à la fois une recherche par champs, et une pondération des recherches plus générales en augmentant l'importance des éléments contenus dans les champs validés. Ces évolutions supposent une nouvelle façon d'envisager les données des catalogues.

¹¹⁶ <http://infolab.stanford.edu/~backrub/google.html>, présentation par Sergey Brin et Lawrence Page.

¹¹⁷ <http://www.wikipedia.org/>

¹¹⁸ PRAX, Jean-Yves; LARCHER, Simon. *La gestion électronique documentaire*. Paris: Dunod, 2004, 342 p., p.289-290.

¹¹⁹ VISE, David A.; MALSEED, Mark. *The Google story*. New York: Delacorte Press, 2005.

2.4.2 Les FRBR

Les FRBR¹²⁰ sont un modèle de description bibliographique qui constitue une nouvelle approche du catalogue, conçue entre 1992 et 1997 par un groupe d'étude de l'IFLA. Il s'agit de repenser les données selon une analyse théorique des œuvres, dans un but pratique avec notamment le travail de l'OCLC pour Worldcat.¹²¹

- Une nouvelle analyse théorique

Depuis les années 1990, divers changements remettent en cause l'élaboration des données bibliographiques. En raison de la diversification des formats de publication, il est habituel de trouver le même contenu sous plusieurs formes. Du fait des processus de localisation et de dérivation des notices, il fallait rationaliser les procédures et clarifier les concepts bibliographiques pour en permettre l'interopérabilité. Les éléments de données eux-mêmes ne sont donc pas remis en cause, mais leur agencement renouvelé par le modèle FRBR,¹²² qui reformule les concepts de l'information bibliographique. L'objectif est de faire apparaître la structure du catalogue. Il s'agit de modèles applicables dans une interrogation sur le Web. En ce sens les FRBR des bibliothèques sont comparables aux CRM des musées, élaborés eux aussi il y a une dizaine d'années face à l'évolution de la description des œuvres. Ce modèle a encore peu d'applications, mais a le mérite de recentrer la description autour l'œuvre.

- Repenser des données identiques

Les FRBR formulent des données anciennes, selon un schéma novateur. Il s'agit de considérer l'œuvre elle-même comme une unité intellectuelle, distincte des objets conservés. C'est ainsi que se superposent l'œuvre, les expressions de l'œuvre (adaptations, diverses versions), les manifestations des expressions (traductions, éditions différentes), et les items des manifestations (objets, documents que l'on conserve et consulte).¹²³ Ce découpage permet de présenter des résultats de recherche sans juxtaposer des réponses qui sont en fait liées, et surtout d'établir des relations entre les entités décrites. Dans le modèle FRBR, les entités « Œuvre - Expression - Manifestation – Item » constituent ainsi le groupe 1, organisé avec des relations « structurelles ». La Personne et la Collectivité (auteur) constituent le groupe 2, et ont une relation de « responsabilité » avec le groupe 1. Enfin, les Concept, Objet, Événement et Lieu constituent le groupe 3, lié à l'œuvre par une relation « sujet ».

L'application de cette analyse théorique des œuvres suppose des changements importants.¹²⁴ Les nouveaux principes internationaux de catalogage sont pensés pour des catalogues en ligne. D'où le besoin de couvrir tous les types de documents et d'accès, et d'être compatibles avec ce modèle conceptuel relationnel, reposant sur la notion

¹²⁰ *Functional Requirements for Bibliographical Records*, Fonctionnalités requises des notices bibliographiques.

¹²¹ BnF - Lettre d'information - Actualités du catalogue : produits et services bibliographiques, Mai 2008 | n° 11, <http://www.bnf.fr/pages/lettres/produits/produits11.htm>

¹²² Texte anglais : <http://www.ifla.org/VII/s13/frbr/frbr.pdf>, Traduction française Patrick Le Bœuf et Françoise Leresche.: <http://www.bnf.fr/pages/infopro/normes/pdf/FRBR.pdf> *IFLA. Spécifications fonctionnelles des notices bibliographiques : rapport final.*

¹²³ Voir annexe 2, exemple de modélisation.

¹²⁴ Voir annexe 2, exemples d'application.

d'œuvre. Outre une étude de la BnF¹²⁵ à ce sujet, on peut mentionner le travail de l'OCLC¹²⁶ et de l'IFLA¹²⁷ sur les diverses applications FRBR. L'adaptation des nouveaux principes de catalogage à ce modèle relationnel correspond donc à une attente réelle.

2.5 DES INTERFACES HOMME-MACHINE

Le catalogue pose la question de l'ergonomie, qui doit être agréable et efficace.

2.5.1 Apparence et ergonomie

L'impression que le catalogue donne à l'utilisateur implique l'apparence graphique, mais aussi la simplicité de fonctionnement, et la pertinence des résultats

- Plaire à l'utilisateur

La question du *design* des catalogues amène à présenter un paradoxe important. Il pourrait sembler évident qu'une interface esthétique, recherchée et chaleureuse est un atout pour la mise en valeur d'un instrument de recherche. Pourtant, on observe la réussite de sites à l'esthétique dépouillée, à commencer par celui de Google, qui refuse d'afficher des publicités sur sa page d'accueil,¹²⁸ et évite même toute illustration.¹²⁹ Il s'agissait certes au départ surtout de fournir une connexion rapide en réduisant les temps de latence, et cette simplicité est devenue une marque de fabrique. Mais on peut aussi voir une esthétique dépouillée permettant de se concentrer sur la recherche. De même, Youtube¹³⁰ permet le partage de vidéos avec une interface peu originale et d'apparence déstructurée. L'aspect simple et désorganisé n'en empêche pas le succès, au contraire. Cependant, l'idée d'interfaces comparables à celle de Google pour les catalogues ne semble pas fonctionner, puisqu'en ce cas, les lecteurs préfèrent utiliser directement Google et Wikipédia, ne voyant pas en quoi une recherche similaire leur apporte.

L'ergonomie des OPACs est certes liée aux *packages* possibles avec la fourniture des logiciels, mais il est d'autant plus nécessaire de réfléchir aux évolutions souhaitables. C'est ainsi que le projet Europeana a permis une réflexion sur les interfaces graphiques avec, dans le prototype proposé par la BnF en mars 2007,¹³¹ une reprise de la Dewey permettant de naviguer dans une arborescence thématique,¹³² puis dans le modèle tout récemment lancé,¹³³ une interface plus dépouillée s'inspirant des sites de partage de documents multimédia. L'interface doit permettre d'accéder aux documents de la

¹²⁵ Fédération internationale des associations de bibliothécaires et des bibliothèques. Groupe de travail sur les spécificités fonctionnelles des notices bibliographiques. Paris: Bibliothèque nationale de France, 2001. <http://www.bnf.fr/pages/infopro/normes/pdf/FRBR.pdf>.

¹²⁶ Voir <http://www.oclc.org/fr/fr/memberscouncil/meetings/2003/october/interestgroups/cataloging.pdf> pour les réflexions du groupe « Catalogue et métadonnées », sur la « FRBRisation » de Worldcat.

¹²⁷ On trouve sur le site de l'IFLA une bibliographie spécifique de 53 pages, http://www.ifla.org/VII/s13/wgfrbr/FRBR_bibliography.pdf

¹²⁸ VISE, David A.; MALSEED, Mark. *The Google story*. New York: Delacorte Press, 2005.

¹²⁹ A l'exception des occasionnelles fantaisies graphiques sur le nom de la marque.

¹³⁰ Le cinquième site le plus visité au monde, GERVAIS, Jean-François. *Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups..* Paris: Dunod, impr. 2007, 216 p, p. 110.

¹³¹ Maquette d'Europeana élaborée par la BnF : <http://maquette.bnf.fr/labs/scenario/Europeana.demo.html>

¹³² Comme avec les « Constellations » de la BM de Lille, <http://portail.bibliotheque.bm-lille.fr/>.

¹³³ Europeana dans sa version financée par l'Union Européenne, <http://www.europeana.eu/portal/>

bibliothèque, et d'utiliser ses services de façon intuitive. En ce sens le système de navigation à facettes, avec une première requête par mots-clés puis des propositions selon un cluster sémantique, est un type d'interface adaptable en bibliothèque.¹³⁴ Avec la définition de l'interface, se pose la question des différents services rendus au lecteur.

- Des fonctions pour les lecteurs

Le premier élément de présentation du catalogue est une interface graphique lisible, avec des libellés en clair à la place des zones, pour que le sens soit transparent pour l'utilisateur. Cette fonction implique la présence de notices séparables, avec des feuilles de style pour présenter différemment des données de départ identiques. Par exemple avec le catalogue de la BnF, Bn Opale Plus, il est possible d'utiliser l'affichage en INTERMARC ou en UNIMARC (pour les professionnels), ou « public » (transparent). Il est aussi possible d'exporter des données, pour les professionnels mais aussi pour un public de chercheurs (vers un logiciel de bibliographie ou pour une simple copie de référence). Cette fonction existe par exemple dans Bn Opale Plus et dans le SUDOC.

Diverses fonctions peuvent intéresser l'utilisateur en ligne. En premier lieu, la vérification de l'état de disponibilité de l'ouvrage, puisque le catalogue en ligne doit permettre à l'utilisateur de savoir s'il est nécessaire de venir à la bibliothèque. Cette fonction liée aux modules de circulation du SIGB a été mise en place dans la plupart des OPACs. La possibilité de réserver l'ouvrage en est la suite logique. Limitée aux lecteurs inscrits, cette fonction est essentielle pour les ouvrages exclus du prêt. La réservation en ligne fonctionne par exemple très bien pour les ouvrages de recherche à la BnF. On peut aussi afficher le suivi de l'ouvrage (par exemple, en commande ou en restauration), ce qui n'est pas à proprement parler une fonction pour le lecteur, mais a presque ce statut lorsqu'on utilise le catalogue. Au-delà de ces fonctions simples, la façon d'envisager les recherches est très liée aux choix stratégiques effectués par l'établissement.

- Faciliter la recherche : quels choix ?

En premier lieu, la bibliothèque peut conserver un catalogue restant autonome mais repérable (par exemple sur le site de la municipalité d'une bibliothèque municipale). Il est possible d'effectuer un renouvellement graphique, voire une réorganisation de l'ergonomie.¹³⁵ En ce cas, l'aspect de l'interface est changé, mais l'OPAC conserve le même statut. Il peut de cette façon fournir des services annexes, par exemple avec des liens et des éléments cartographiques intégrés par des *mashups*. Par exemple, une proposition de « catalogue du futur » par l'ABES¹³⁶ intègre ce type de fonctions. Il peut même s'agir de services à distance accessibles par le catalogue mais sortant des ressources documentaires, comme le service du Guichet du Savoir à la Bibliothèque Municipale de Lyon.¹³⁷ Le catalogue comme reflet de l'identité de la bibliothèque n'exclut pas divers choix possibles pour mettre en valeur les ressources.

¹³⁴ Worldcat utilise ces techniques récentes, comme les « facettes » et les « Worldcat identities ».

¹³⁵ Par exemple à Fresne, avec la structuration de la grille Conspectus, pour une recherche par sujets : <http://bm.fresnes94.fr/bibliotheque/accueil.asp>. L'innovation pour « rechercher un document » y est telle que le terme de « catalogue » ne figure nulle part...

¹³⁶ L'ABES a élaboré un modèle d'interface intégrant des webservice complétant les données indexées : [http://www.abes.fr/abes/documents/reseau/journees_reseau/05_ExperimentationWebSudoc\(2\).ppt](http://www.abes.fr/abes/documents/reseau/journees_reseau/05_ExperimentationWebSudoc(2).ppt)

¹³⁷ Même si en ce cas la conservation des réponses publiées les transforme partiellement en ressources classiques. <http://autonomy.bm-lyon.fr/retina/public/login.do?username=guest&password=guest>.

Ainsi, il est possible d'intégrer le catalogue dans un catalogue collectif. En ce cas, on suppose que l'on utilise la fonction de localisation (éventuellement automatisée, comme dans Worldcat) pour trouver une ressource près de chez soi. Si l'utilisateur cherche à l'inverse les ressources qui se trouvent dans une bibliothèque précise, il peut toujours être renvoyé à l'OPAC. Le catalogue peut aussi s'intégrer dans un moteur de recherches fédérées, par exemple sur un portail universitaire.¹³⁸ Ce type d'outil n'empêche pas l'OPAC de fonctionner indépendamment, mais permet une plus grande visibilité. Enfin, le catalogue peut s'intégrer dans des moteurs généraux de recherche sur le Web.

2.5.2 Fournir les documents

Les catalogues, pour permettre d'accéder aux documents, doivent s'adapter et reprendre à leur compte diverses évolutions récentes.

- Accéder aux documents

Dans leur fonction première, les OPACs permettaient de préparer une visite en bibliothèque en repérant les ouvrages que l'on allait consulter.¹³⁹ Cette fonction change avec la possibilité d'interroger plusieurs catalogues simultanément, qu'il s'agisse de vrais catalogues collectifs, comme le SUDOC, ou de catalogues fédérés permettant d'interroger différents catalogues avec la même interface, comme le CCFR,¹⁴⁰ TEL,¹⁴¹ ou le KVK.¹⁴²

Plus encore, ce type de fonctions permet d'intégrer les catalogues dans des sites autres que ceux des bibliothèques. Ainsi, que Google books¹⁴³ propose, à partir des notices d'ouvrages, de « trouver ce livre près de chez vous », ce qui met en œuvre un *mashup* avec une localisation à partir de l'adresse IP de l'utilisateur, et une interrogation automatique de Worldcat. Ce mode de recherche intuitif se situe sur le chemin habituel des internautes, et constitue donc une concurrence rude pour les interfaces traditionnelles. En effet, plusieurs problèmes restent à résoudre : Worldcat, le plus grand catalogue du monde, n'intègre qu'une maigre partie des bibliothèques françaises ; en l'absence de mise en place suffisamment fine des FRBR, il est difficile pour l'utilisateur de trouver d'autres manifestations de la même œuvre ; il est donc nécessaire que les catalogues rejoignant Worldcat présentent une intropérabilité totale, mais aussi un degré de précision et des types de descriptions cohérents. Surtout, ce type d'outil laisse croire à l'utilisateur que la ressource est inexistante s'il ne l'a pas trouvée. L'aspect hégémonique de ces outils risque donc de mettre à l'écart les bibliothèques moins fortunées, sans qu'il soit question des volontés des institutions. Mais, à l'inverse, cela

¹³⁸ Permettant d'interroger les ressources des Bibliothèques, de la bibliothèque numérique et des publications et bases en ligne.

¹³⁹ Les OPACs des bibliothèques peuvent avoir une fonction de client pour choisir des cibles, pour interroger d'autres catalogues, et créer un catalogue collectif virtuel dans le SCD ou la bibliothèque.

¹⁴⁰ Catalogue Collectif de France, qui est en fait un méta-catalogue et non un catalogue collectif, interrogeant notamment Bn Opale Plus, le Sudoc, et la Base Patrimoine (catalogue des fonds anciens de 60 bibliothèques municipales et spécialisées). <http://ccfr.bnf.fr/portailccfr/servlet/LoginServlet>.

¹⁴¹ The European Library, méta-catalogue des catalogues de bibliothèques nationales européennes, <http://search.theeuropeanlibrary.org/portal/en/index.html>.

¹⁴² Karlsruhe Virtueller Katalog, méta-catalogue maintenu par la bibliothèque de l'université de Karlsruhe, interrogeant diverses grandes bibliothèques mondiales et des libraires en ligne, http://www.ubka.uni-karlsruhe.de/kvk/kvk/kvk_fr.html.

¹⁴³ <http://books.google.fr/>, à partir des pages de résultats de recherches de livres.

peut être une occasion d'apparaître sur le chemin des internautes pour des petites bibliothèques dont peu de lecteurs consultent les OPACs.¹⁴⁴

- Rechercher dans les documents

Avec l'accès à Worlcat par l'intermédiaire de Google Books, se pose la question de la recherche dans les livres eux-mêmes. En effet, en indexant le contenu des ouvrages numérisés, y compris pour des ouvrages sous droits dont l'utilisateur ne voit que des extraits, Google propose d'inverser complètement la logique du catalogue. Il s'agit de chercher les informations dans le document lui-même, puis d'avoir affaire aux données de description pour valider la ressource. Cette logique de recherche par mots-clés contenus dans le texte impose une réflexion. Il s'agit non seulement d'une concurrence forte imposée par Google Books, mais aussi d'une nouvelle logique de recherche, intégrant les ouvrages numériques dans le monde des publications, au sens où la recherche des ouvrages publiés s'effectue par l'intermédiaire de leur forme numérique.

Ces changements liés à l'accès direct au document sont indissociables des évolutions de pratiques de recherche induites par l'utilisation d'Internet. Il s'agit aussi de changements d'usages, puisque le livre devient un objet du Web, dans lequel on cherche une information sans s'intéresser nécessairement à l'oeuvre en elle-même.

¹⁴⁴ Les jeux de caractères multi-écritures permettent de ne plus utiliser des données translittérées dans les notices, mais plusieurs types d'écritures. Par exemple à la BULAC, <http://www.catalogue.bulac.fr>. C'est une étape importante pour l'évolution mondiale d'un catalogue collectif américain.

3. Les usages des internautes et leur impact sur les catalogues

Les changements technologiques récents, avec la relative démocratisation de l'Internet, la puissance de calcul des machines et l'accès de plus en plus fréquent au haut débit, ont entraîné de nouveaux usages. Il s'agit de pratiques de consommation culturelle avec des produits immédiatement disponibles sous une forme numérique, et d'un accès à l'information selon une ergonomie devenant intuitive.

3.1 ERGONOMIE DU WEB ET ACCES AUX RESSOURCES

L'utilisation des outils de recherche en ligne modifie les techniques utilisées, le rapport aux ouvrages, et la place des lecteurs. Seulement la moitié des foyers français environ sont connectés à Internet, mais les habitudes de recherche sont modifiées de façon générale. Les informations données par Médiamétrie sur l'audience de l'Internet,¹⁴⁵ permettent de resituer dans un contexte plus large l'analyse des outils des bibliothèques.

3.1.1 Les moteurs pour la recherche d'information

En premier lieu, l'habitude d'utiliser des moteurs de recherche offre de nouvelles possibilités, avec une logique de traitement de l'information à échelle mondiale.

- De nouveaux outils

La diffusion des connexions en haut débit dès le début des années 2000 a constitué une avancée décisive pour l'utilisation d'Internet par les particuliers. Ne plus payer en fonction du temps de connexion encourageait à rester en ligne, et la possibilité d'échanger des données rapidement allait accroître la place du multimédia. Surtout, la gratuité des connexions et la facilité de transfert des données rendent moins nécessaire le téléchargement proprement dit. Sans retracer l'histoire des moteurs de recherche, on peut remarquer que leur développement à partir de 1995 avec Altavista a abouti dans les années 2000 à une situation où Google occupe une position dominante.¹⁴⁶

Ce développement inquiète souvent les professionnels des bibliothèques, avec l'idée que la simplicité et l'efficacité des recherches sur Google rendent inutile le recours aux catalogues de bibliothèques.¹⁴⁷

Pourtant, l'action de Google utilise autant que possible l'images des bibliothèques. On peut citer notamment la communication publicitaire autour des bibliothèques fournissant les ouvrages de Google Books.¹⁴⁸ La mise en valeur de cette image de bibliothèques et

¹⁴⁵ http://www.mediametrie.fr/resultats.php?resultat_id=585&rubrique=net

¹⁴⁶ Et sur les applications : ÉLIE, Jean-Philippe. *Tout sur les outils Google*. Paris : Dunod, 2008, 212 p.

¹⁴⁷ Voir BYRNE, Alex, « qui a besoin des bibliothèques à l'heure de Google », *BBF* 52, n°5, p. 92-93

¹⁴⁸ A la fin de l'année 2004, cette firme avait annoncé son projet de numérisation et de mise en ligne de 15 millions d'ouvrages d'ici 2015, avec dès le départ la volonté affichée de créer une bibliothèque

de fonds n'est pas anodine, bien que « Google print » soit devenu « Google book search »¹⁴⁹ en 2006 justement pour évoquer l'image de la recherche par moteurs et pas de la bibliothèque de papier.¹⁵⁰

Les accords avec les grands catalogues correspondent aussi à un service intéressant pour les bibliothèques. On peut citer en particulier « trouver ce livre en bibliothèque près de chez vous » dans Google books, renvoyant à Worldcat. Outre l'important service rendu, la référence aux bibliothèques est liée l'image que développe Google, avec une dimension de gratuité et de désintéressement. Le fait d'y associer le service public que sont les bibliothèques a un aspect stratégique, assimilant l'offre gratuite au service public. Cette action de Google peut constituer une aubaine pour les bibliothèques. Elle doit en tout cas être prise en compte, car elle modifie les pratiques de recherche.

- De nouvelles habitudes

Diverses pratiques de recherche sont possibles. La typologie de lecteurs établie pour l'étude qualitative de la réception du prototype Europeana correspondait ainsi à des publics très divers.¹⁵¹ Il est difficile pour les bibliothécaires de répondre efficacement à la diversité de ces publics. On peut simplement rappeler que les outils mis en place doivent à la fois s'adapter aux nouveaux usages, et laisser une possibilité de se contenter des ressources disponibles sur place pour les publics laissés de côté par la « fracture numérique », ou ne souhaitant pas utiliser l'informatique pour leurs lectures de loisir.

Après une première vague d'euphorie pour les outils collaboratifs, il était ressorti de cette étude¹⁵² que les différents publics, dans le cas d'Europeana, désiraient unanimement que la bibliothèque leur fournisse des contenus sélectionnés, les lieux d'expression pour les internautes existant déjà par ailleurs. La validation de l'information correspond à un rôle ancien des bibliothèques, avec une analyse des contenus mis à disposition, mais elle s'inscrit dans une logique différente avec la masse des informations disponibles sur le Web.

- Une logique de masse

La quantité d'informations disponibles sur le Web change complètement la façon de chercher une ressource pertinente, quel que soit le type de document.

numérique mondiale. Ce projet s'appuie sur des collections empruntées à des bibliothèques prestigieuses, principalement d'universités anglo-saxonnes : Les universités de Harvard (Widener Library), de Stanford (Green Library), du Michigan et d'Oxford (Bodleian Library), et la New York Public Library au départ, puis celles Californie (août 2006), du Wisconsin-Madison (octobre 2006), du Texas à Austin (janvier 2007), et la Staatsbibliothek de Bavière (mars 2007), l'université de Lausanne (mai 2007) et l'université de Mysore en Inde (mai 2007), puis le CIC.

¹⁴⁹ <http://books.google.com> (version française : <http://books.google.fr>); réflexions par exemple sur <http://booksearch.blogspot.com>

¹⁵⁰ SALAÜN, Jean-Michel, « Bibliothèques numériques et Google-Print », 2005 http://archivesic.ccsd.cnrs.fr/docs/00/06/26/68/PDF/sic_00001576.pdf

¹⁵¹ BOUVIER-AJAM, Laurent ; LESQUINS, Noémie ; RIPON, Romuald. Europeana : étude sur les usages et les attentes relatifs à l'interface de consultation de la future Bibliothèque numérique Européenne : rapport final. 21 mars 2007, 53 p.

http://bibnum.bnf.fr/usages/BnF_BNuE_Ourouk_Rapport_final.pdf.

¹⁵² Ibid.

Par exemple, malgré le peu de rigueur des modes de descriptions, le nombre d'images permet souvent de retrouver des ressources,¹⁵³ car les moyens sont suffisamment importants pour un résultat efficace. C'est une logique de recherche inverse de celle du catalogue, puisque l'utilisateur ne cherche pas dans des descriptions avant de trouver des ressources, mais dans des ressources avant de les valider par une description.

Malgré cette différence de démarche de recherche, la description des contenus reste essentielle, en vue de leur indexation automatique pour les recherches. Bien qu'ils soient en partie inconnus, les divers éléments utilisés par l'algorithme de pondération des résultats de Google doivent être pris en compte, aussi bien pour les moyens techniques que pour les critères intellectuels de pondération. Cette réflexion est cruciale pour l'établissement de moteurs de recherche internes, quand il s'agit d'outils proposés par Google mais surtout pour la construction de moteurs de recherche propres comme pour Gallica2, avec un algorithme spécifique de pondération. Des moteurs internes peuvent aussi être ajoutés de façon optionnelle, comme le moteur SFX à Lausanne, avec RERO (catalogue roman),¹⁵⁴ qui fournit une réponse enrichie s'il est mis en œuvre, en fournissant des « slots » de livres supplémentaires.

En établissant une barre avec un moteur de recherche interne (comme le font Wikipedia ou les sites de ventes commerciaux), un site établit un outil limité, qui permet à l'internaute de continuer ses recherches en ne sortant pas du site en question. Il suppose cependant que l'internaute soit sur ce site. La présence de moteurs de recherche internes n'exclut donc pas que les ressources doivent être indexées par les moteurs généraux.

L'intégration dans des moteurs existants peut être effectuée automatiquement dans le cadre des collectes par les robots des moteurs. Des accords peuvent intervenir pour faire apparaître les ressources de façon plus pertinente. Ainsi, l'accord entre la BnF et l'OCLC pour l'intégration des données de la BnF dans l'OCLC-Worldcat correspond aux critères d'indexation par Google. Il est même possible de passer des accords directement avec les moteurs pour faire en sorte que les références soient bien placées. Par exemple, des accords existent entre Rero¹⁵⁵ et Google, et entre l'ABES¹⁵⁶ et Google, pour que les institutions fournissent des fichiers et données dans le bon format, en échange de quoi le moteur s'engage à en faire une présentation, bien qu'il en soit maître.

En définitive, les deux seuls moyens de faire apparaître les données de façon vraiment efficace pour la collecte des moteurs seraient d'en permettre un affichage en HTML (ce sur quoi travaille la BnF) pour simplifier l'indexation par Google, ou de faire figurer les notices dans des entrepôts OAI (c'est-à-dire des réservoirs de notices), à condition que ces derniers soient moissonnés.

Il s'agit en tout cas de répondre à une évolution incontestable des pratiques, « chercher sur Internet » étant devenu la locution vernaculaire usuelle signifiant « formuler une recherche par mots-clés dans Google ». La langue Anglaise propose la même équivalence, dans l'autre sens, puisque « to Google » signifie « effectuer une recherche et trouver sur Internet ».

¹⁵³ En particulier grâce à Google Images. Voir Association de pédagogie et de didactique de l'École normale de Lille. *Spirale : Revue de recherches en éducation*. Lille, 1988-, N°40, 2007, Présentation par Annette BÉGUIN-VERBRUGGE, <http://spirale-edu-revue.fr/spip.php?rubrique115> .

¹⁵⁴ <http://opac.rero.ch/gateway>

¹⁵⁵ <http://opac.rero.ch/gateway>

¹⁵⁶ <http://www.abes.fr>.

3.1.2 Consultation en ligne et téléchargement

Sur le Web, le catalogue décrit des ressources qui sont pour certaines accessibles en ligne, mais dont la consultation sous forme papier peut continuer à présenter un intérêt. Il faut donc intégrer les ressources très diverses que les bibliothèques peuvent fournir, dans un ensemble de ressources immédiatement disponibles.

- Tout sur place

Le premier avantage offert par le contenu des pages Web est que le lecteur peut effectuer une recherche d'informations avec une seule interface pour la requête et les résultats. Or, les ouvrages physiques signalés en ligne ont un statut ambigu : ce sont à la fois des objets matériels que l'on peut chercher à se procurer, et des sources d'informations que le lecteur peut chercher à obtenir indépendamment du support. L'instrument de recherche en ligne implique donc deux fonctions : permettre d'identifier les ressources numérisées et physiques comme éléments des collections de la bibliothèque, et laisser la possibilité d'accéder à autant d'informations que possible. En ce sens, l'évolution des catalogues autonomes ne va sans doute pas vers une réduction mais au contraire vers un enrichissement des données.

Pour les recherches précises, aussi bien pour des chercheurs habitués à fréquenter une bibliothèque que dans le cadre de la lecture publique, il est donc important de conserver un outil qui corresponde à des demandes précises des internautes.

Comme dans les catalogues commerciaux comparables, il est important que le catalogue de bibliothèque présente les deux avantages d'être précis et efficace quand on l'utilise en tant que tel, et d'être facile à trouver sur les moteurs de recherche.

Pourtant, l'apparition dans les résultats des requêtes simples sur des moteurs pourrait sembler hors de propos quand il s'agit de bibliothèques physiques. Dans la mesure où ce service ne tient pas compte de la localisation de l'utilisateur et où il s'agit d'ouvrages physiques peu rares, cette présence peut constituer du bruit et polluer la recherche. La question des objectifs est donc cruciale : l'idée de se faire voir et d'attirer les lecteurs n'a de sens que si cela leur permet de mieux répondre à leurs besoins. En l'occurrence, il pourrait s'agir à la fois d'une réponse technique intégrant la localisation de l'utilisateur, ou plus simplement d'un renvoi permettant de trouver un service comparable près de chez soi. Cette possibilité de navigation entre les sites est essentielle pour rendre utiles les ressources dont les bibliothèques proposent une description en ligne. La question est comparable pour l'accès aux ouvrages numérisés, que nous décrirons plus tard. Il s'agit en tout cas de s'inscrire dans une culture de l'immédiateté.

- Culture gratuite et immédiate

Les bibliothèques ont commencé à mettre leurs catalogues en réseaux très tôt, avant même le développement d'Internet, de façon gratuite et avec un accès libre au signalement des documents, sans que l'accès aux documents eux-mêmes y soit associé. Les catalogues se sont donc intégrés sans difficulté à la culture de gratuité et d'immédiateté liée à Internet. Ainsi, lors du procès achevé le 29 octobre 2008,¹⁵⁷

¹⁵⁷ L'outil Google Book Search a donné lieu à un procès avec notamment *The Authors Guild* et l'*Association of American Publishers*, contre Google qui est finalement libre de continuer à scanner les livres, contre 125 millions de dollars versés aux éditeurs.

opposant Google aux sociétés des éditeurs et des auteurs Américains, le premier argument avancé pour justifier la mise en ligne gratuite d'ouvrages était le fait que ce service soit rendu par les bibliothèques de façon comparable et incontestée.¹⁵⁸

En revanche, l'évolution du Web implique une ergonomie nouvelle et des habitudes de recherche devenant systématiques pour les utilisateurs. Il est nécessaire que les catalogues et portails de bibliothèques s'y adaptent. Il s'agit à la fois de s'intégrer efficacement aux recherches par moteurs, de fournir des liens pour circuler entre les ressources, et de permettre une appropriation au travers des nouveaux outils du Web, en intégrant éventuellement les usagers aux processus.

3.1.3 Usagers : des lecteurs catalogueurs ?

L'évolution de l'utilisation du Web implique des ressources numériques mises en ligne massivement, un besoin de les décrire avec précision, et une participation des usagers de plus en plus grande à la production de contenus ou de descriptions. Les bibliothèques peuvent y voir l'occasion d'adapter les descriptions aux usages, en laissant les utilisateurs participer à leur élaboration.

- Indexer selon les utilisations

L'idée d'établir une indexation en fonction des usages, selon les demandes et non à partir de listes prédéterminées, est déjà ancienne. Difficile à mettre en place, elle a été essayée dans le domaine de l'audiovisuel, où l'indexation des contenus est la plus délicate. C'est ainsi que l'INA¹⁵⁹ a proposé dès 1975 environ une indexation par mots-matières selon les demandes des personnes, pour informatiser les bobines de télévision. C'étaient les questions de journalistes qui servaient à établir les termes, avec une enquête sur leur vocabulaire. L'idée d'indexer à partir de la demande et non à partir des bobines trouva pourtant ses limites, et le langage libre dut rapidement être contrôlé, la demande servant simplement à établir un corpus des termes d'indexation.

Dans le domaine de l'indexation des images en bibliothèques spécialisées, il s'agissait de mettre en place un vocabulaire d'indexation continuellement adapté. On peut citer par exemple l'« iconothèque » élaborée dans les années 1980 par Henri Hudrizier,¹⁶⁰ pour indexer les images selon leur légende et non selon leur contenu. De la même façon, Michel Melot pose les principes d'une indexation évoluant selon ce que les gens cherchent.¹⁶¹ De fait, l'indexation des images à partir des légendes est pour le moment la seule façon de permettre des recherches par mots-clés. C'est donc le mode de recherche habituel de Google Images, ou de sites spécialisés comme Flickr par exemple. En ce cas, le fait que les utilisateurs soient aussi producteurs du contenu et de sa description amène évidemment à s'interroger sur les changements d'usages.

¹⁵⁸ Voir ERTZSCHEID, Olivier « Le marché du livre comme algorithme », http://affordance.typepad.com/mon_weblog/2008/11/le-march%25C3%25A9-du-livre-comme-algorithme-suite-.html

¹⁵⁹ Institut National de l'Audiovisuel, <http://www.ina.fr/>.

¹⁶⁰ Par exemple HUDRISIER, Henri, L'iconothèque, documentation audiovisuelle et banques d'images, La Documentation française, collection « audiovisuel et Communication », 1982.

¹⁶¹ MELOT, Michel. *Une brève histoire de l'image*. Paris: l'œil neuf, 2007.

- « Web 2.0 » : changements d'usages ?

On parle couramment du « Web 2.0 », par analogie avec les versions révisées des logiciels, pour désigner l'évolution du Web vers des outils rendant l'utilisateur actif sur les contenus. Cette expression a été lancée par Tim O'Reilly en 2004 pour une conférence, par manière de boutade.¹⁶² Il n'a pas défini ce terme, pour lequel il a pourtant donné plusieurs pistes de réflexion en raison de sa popularité.¹⁶³

Ce terme qui était au départ une plaisanterie a été unanimement repris. C'est en fait une façon de désigner l'ensemble des outils collaboratifs, qu'il s'agisse à proprement parler de sites dont les contenus sont fournis par les internautes comme Youtube et Flickr, de sites de partages de favoris, de « réseaux sociaux » comme Myspace ou Facebook, ou simplement de techniques permettant d'agrèger des liens d'information à un portail avec la technologie des Fils RSS. Il s'agit donc de techniques et de pratiques très différentes. Le « Web 2.0 » correspond pourtant aux concepts de départ de l'Internet, visant à permettre des outils collaboratifs.

Ces évolutions inquiètent souvent les professionnels des bibliothèques, au point que le terme finit par décrire l'ensemble des évolutions techniques pouvant inquiéter. On peut citer par exemple l'article d'Emilie Barthet et Amélie Church, « l'avenir des bibliothèques à l'heure du Web 2.0 »,¹⁶⁴ où l'interrogation sur les technologies prend nettement la forme d'une inquiétude. Il peut aussi s'agir d'une simple volonté d'adaptation, la « bibliothèque 2.0 » répondant au « Web 2.0 », en cherchant à selon Emmanuelle Bermès à « renouveler les technologies, les pratiques, les services et finalement l'esprit de la bibliothèque ». ¹⁶⁵

En pratique, le terme de « Web 2.0 » est souvent utilisé pour désigner un ensemble d'applications qui sont en fait très variées du point de vue du bibliothécaire. Il s'agit :

- de ressources proposées par des usagers.¹⁶⁶
- d'indexation, de signalisation ou d'évaluation par des usagers,
- de visibilité des usagers avec des profils personnels (les « réseaux sociaux »)
- d'outils de veille automatique, comme les flux RSS et agrégateurs de liens.

Ces aspects convergents posent des questions d'utilité et de légitimité variables. Si la légitimité de l'intervention des bibliothèques publiques dans les outils de veille documentaire est incontestable en elle-même, la mise en application reste parfois discutable. D'une part l'apport des nouveaux services peut être maigre par rapport à la masse d'informations disponibles sur le Web ; d'autre part la participation à des services déjà existants constitue une forme d'encouragement à partir du moment où l'on développe des outils correspondant à un produit. En ce cas la participation de services publics peut être ambiguë, avec l'intégration dans des services gratuits pour l'utilisateur mais de fonctionnement opaque et commercial (comme Facebook).

¹⁶² Tim O'Reilly est le fondateur des éditions O'Reilly, dans le domaine informatique, cité par GERVAIS, Jean-François. Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups.. Paris: Dunod, 2007. 216 p., p. 4.

¹⁶³ Voir son article <http://www.oreily.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

¹⁶⁴ Emilie Barthet, Amélie Church, « l'avenir des bibliothèques à l'heure du Web 2.0 », BBF 52, n°5, p. 92-93,

¹⁶⁵ Emmanuelle Bermès, « Gestion d'équipe : accompagner le changement, In Arhimag, Guide pratique « La bibliothèque à l'heure du Web 2.0 », juillet 2008.

¹⁶⁶ Aussi bien sur des sites collectifs que personnels avec la création des blogs.

A l'inverse, les techniques d'édition partielle de contenus automatiquement mis à jour, comme l'abonnement à des Flux RSS qui ne concernent que l'utilisateur, correspondent à une information au même titre que les autres services rendus par les bibliothèques. Ces évolutions sont liées aux catalogues de bibliothèques par deux aspects : d'une part la mise en place de techniques par lesquelles le catalogue répond aux façons nouvelles de repérer des sources d'information, et d'autre part l'adaptation aux pratiques et itinéraires d'un public changeant.

- Changements de public ?

Il est difficile d'étudier avec rigueur les publics intéressés par les catalogues des bibliothèques, et par les bibliothèques numériques, qui peuvent toucher un autre public. Les utilisateurs des bibliothèques numériques ont donné lieu à diverses enquêtes, notamment les publics du prototype d'Europeana, pour la BnF.¹⁶⁷ Mais il s'agit ici de la question plus large des internautes et de leurs habitudes. Si la moitié environ des foyers français sont connectés à Internet,¹⁶⁸ en pratique de nombreux services commerciaux¹⁶⁹ et même des services publics¹⁷⁰ tendent à considérer l'ensemble des Français comme des internautes.

Il serait illusoire de vouloir étudier les pratiques de l'ensemble des internautes. On peut cependant constater plusieurs évolutions. En l'absence de chiffres fiables sur des enquêtes de public, on raisonnera principalement à partir de l'adaptation des outils. D'une part, il est de plus en plus simple de participer aux contenus publiés sur le Web. Qu'il s'agisse de créer un site Web avec un logiciel intuitif,¹⁷¹ un blog ou de participer simplement à un forum ou au blog d'une autre personne en postant un message, les outils de création de contenus permettent de faire de plus en plus avec de moins en moins de compétences spécialisées. L'évolution des outils de recherche en ligne les plus utilisés correspond tout à fait à cette dynamique.¹⁷² Ainsi, le navigateur Chrome¹⁷³ intègre indifféremment les recherches par mots-clés et les saisies d'URL, dans l'idée que l'utilisateur doit taper ce qu'il veut trouver et non connaître le fonctionnement technique.

Mais d'autre part, la diffusion de linux et de logiciels libres supposant une utilisation un peu moins intuitive de l'environnement bureautique laisse supposer un début d'évolution inverse. Pour une partie du grand public, il s'agit d'un engouement pour la compréhension du fonctionnement des systèmes, et d'une découverte de logiques

¹⁶⁷ BOUVIER-AJAM, Laurent ; LESQUINS, Noémie ; RIPON, Romuald. Europeana : étude sur les usages et les attentes relatifs à l'interface de consultation de la future Bibliothèque numérique Européenne : rapport final. 21 mars 2007, 53 p.

http://bibnum.bnf.fr/usages/BnF_BNuE_Ourouk_Rapport_final.pdf

¹⁶⁸ Observatoire des Usages de l'Internet,

http://www.mediametrie.fr/contenu.php?rubrique=net&rubrique_id=259&menu_id=257

¹⁶⁹ Par exemple la SNCF pour tous ses programmes de fidélisation

¹⁷⁰ Par exemple pour les inscriptions aux concours ou à la faculté.

¹⁷¹ Les WYSIWYG (What You See Is What You Get) permettent de produire du code Html à partir d'une interface graphique simple.

¹⁷² Ainsi, le site de partage de signets au nom humoristique de « del.icio.us » est devenu « delicious.com », nom plus insipide mais plus simple, plus accrocheur que la plaisanterie technique. On peut y voir l'adaptation à un public moins connaisseur, utilisant les mêmes outils du fait de leur simplification. Voir « Delicious, Chrome et quelques réflexions sur les URI », <http://www.figoblog.org/node/1928>.

¹⁷³ Google Chrome : <http://www.google.com/chrome/?hl=fr>

informatiques. Cela n'exclut pas la participation aux contenus par des interfaces simples, mais cela suppose un intérêt pour le fonctionnement des outils. C'est à cette démarche que répondent les modes de recherche avancée de Google, avec la possibilité de saisie manuelle de commandes plus complètes que les opérateurs booléens simples de l'interface graphique.¹⁷⁴ On peut en tirer l'idée que formuler une recherche saisie manuellement et utiliser des opérateurs sera peut-être de plus en plus intuitif, du moins pour les lecteurs habitués à effectuer des recherches en ligne. De cette façon, l'écart entre les néophytes et les amateurs éclairés se creuse probablement, y compris parmi les internautes disposés à participer aux contenus.

Il faut aussi prendre garde aux présupposés qui peuvent naître de notre propre participation à Internet. On peut remarquer la surreprésentation des professions liées à l'information dans la blogosphère, avec une activité forte des professionnels des bibliothèques.¹⁷⁵ Cette évolution ne doit pas faire oublier que le développement n'est pas forcément comparable dans les autres secteurs. La validité de ce type de sources pour la bibliothéconomie nous pousse donc à les associer aux ressources validées par les bibliothèques. Le débat sur la signature des blogs, la collaboration dans les Wikis et la validité de l'information correspond à l'idée que l'information est juste si elle semble de bon sens. Or cet aspect est vrai dans les secteurs où les renseignements sont vérifiables, comme l'informatique ou le droit, avec des renvois à des sites officiels techniques ou à des articles de presse, mais la situation est moins simple dans d'autres secteurs.

Plus encore, les pratiques professionnelles et personnelles de l'informatique tendant toujours à se lier, et l'habitude de l'utilisation des outils de réseaux en bibliothèque encourage à estimation haute de la maîtrise des outils par les autres utilisateurs. Une remarque comparable pourrait être formulée au sujet des « réseaux sociaux », où les élèves de grandes écoles et les étudiants aisés sont surreprésentés.

Pour autant, un élément essentiel devant être relevé est la fusion entre des outils de divertissement et de recherche. Ces deux aspects paraissent même parfois indissociables, à la fois en raison de l'utilisation possible des ressources, et du fait de la participation bénévole des utilisateurs pour la création d'instruments de recherche.

- Changements de contenu ?

La participation des utilisateurs est un élément essentiel de l'évolution du Web, à la fois pour la production des contenus et pour la description des ressources. Cette participation semble intéressante pour les bibliothèques, à la fois pour se rendre visible sur les sites concernés en offrant le service le mieux adapté, et pour utiliser la puissance de travail que représentent les internautes.¹⁷⁶

Cette adaptation du travail des internautes prend principalement la forme des « folksonomies » (mot-valise de « folk » et « taxonomy »), c'est-à-dire de l'indexation des contenus effectuée par les utilisateurs.¹⁷⁷ Il s'agit du mode habituel d'indexation libre par des mots-clés ou « tags », utilisé notamment pour les sites dont les utilisateurs

¹⁷⁴ ÉLIE, Jean-Philippe. *Tout sur les outils Google,...*, *Op. Cit.*, par exemple « define : », « site: »

¹⁷⁵ On a même pris l'habitude de mentionner constamment les blogs dans les bibliographies professionnelles, en raison de leur qualité d'outil de travail.

¹⁷⁶ Voir LEBRETON, Claire, *Bibliothèques, tags et folksonomies : L'indexation des bibliothèques à l'ère sociale*, Mémoire ENSSIB, 2007, 117 p.

¹⁷⁷ PEYRARD, Sébastien, *L'accès aux collections numériques de masse*, Mémoire ENSSIB, 2007, 85 p., p. 41 et suivantes.

fournissent les contenus. Il s'agit dans ce cas de la légende proposée sous la forme d'une liste de mots pour trouver le document.

Le site de partage le plus connu est Youtube, site de vidéos.¹⁷⁸ Mais celui qui s'intègre le plus dans les réflexions sur l'avenir des catalogues est Flickr, un site de partage d'images fixes, en particulier de photographies. Il s'agit en effet d'un type de document dont disposent de nombreuses bibliothèques publiques, pour lequel il est souvent pertinent d'effectuer des numérisations, et qu'il est fastidieux et difficile d'indexer de façon utile. Or, la mise en ligne sur un site préexistant pose alors peu de problèmes, puisque Flickr n'a pas de vrai concurrent.¹⁷⁹ Plusieurs bibliothèques publiques ont donc fait le choix de placer sur Flickr des numérisations de photographies issues de leurs fonds, avec des licences « Flickr commons » conçues pour les organismes publics. Il s'agit d'un genre nouveau de catalogage collectif, le contenu étant décrit par les utilisateurs. On peut citer l'essai effectué par la Library Of Congress mettant en ligne 3000 images, indexées par les utilisateurs. D'autres actions de ce type ont été effectuées : La Boston Public Library y place ainsi ses *Fore-Edge Paintings*, photographies des peintures sur tranches de livres anciens, qui sinon ne seraient pas mises en valeur, et la région basse-Normandie y met en ligne des photographies du débarquement de Normandie.¹⁸⁰ Cette utilisation de Flickr pose une question double. Il s'agit d'une part d'utiliser la masse de travail fournie par les usagers (qui sur le Web dépasse largement les possibilités de mobilisation de catalogueurs professionnels), et d'autre part de savoir en quoi les documents continuent à relever de la bibliothèque : est-ce un moyen de la faire apparaître, ou est-ce que seul le document est important ?

Si c'est le cas, comment effectuer le choix du fournisseur et des modalités de mise à disposition des contenus, dans un contexte commercial fort ? Plus encore, pourrait-il y avoir lieu de négocier avec les fournisseurs ?¹⁸¹ En effet, la mise à disposition des images sur Flickr permet une grande visibilité des documents, mais rend la bibliothèque dépendante, et diminue sa visibilité. Faut-il donc à l'inverse envisager d'intégrer un renvoi à des ressources collaboratives depuis des catalogues institutionnels ?¹⁸²

Surtout, la participation des usagers au contenu des bibliothèques numériques n'est pas forcément souhaitable, en raison de la fonction d'expertise que remplit la bibliothèque. Ainsi, les fonctionnalités proposées par la première maquette d'Europeana,¹⁸³ qui permettait notamment aux lecteurs d'évaluer et de commenter les ouvrages, n'avait pas donné satisfaction sur ce point, selon les enquêtes effectuées ensuite. Il y apparaît que les utilisateurs désirent une bibliothèque numérique qui reprenne les atouts de la bibliothèque physique, c'est-à-dire un lieu où les ressources sont validées.

¹⁷⁸ 6^e site le plus visité au monde en 2006 toutes catégories confondues, après Yahoo!, MSN, Google, Baidu, (Le moteur de recherche chinois, <http://www.baidu.com/>) et Myspace.

¹⁷⁹ A part des sites liés à des laboratoires de tirages (Kodak, Snapfish) ou des logiciels de catalogage (Photoshop Elements, Picasa), qui remplissent une fonction différente. GERVAIS, Jean-François. Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups.. Paris: Dunod, 2007, 216 p., p.86.

¹⁸⁰ Par un service de presse (réunissant des journalistes des Etats-Unis, du Canada et de France), et dont un groupe de passionnés fait une analyse complète, allant du repérage des insignes à celui des matériels et des personnes.

¹⁸¹ Comme le prévoit Google avec la rétribution d'auteurs sur Youtube en fonction des pages vues.

¹⁸² Comme dans le prototype de la nouvelle version d'Europeana où la démonstration des « chaussures de Van Gogh » conduisait à Youtube

¹⁸³ Maquette d'Europeana élaborée par la BnF : http://maquette.bnf.fr/labs/scenario/Europeana_demo.html

Or, il est possible que le travail des lecteurs soit validé par une institution. Par exemple, la numérisation et la mise en ligne des registres d'Etat Civil et du cadastre par les Archives Départementales de Mayenne¹⁸⁴ donne lieu à une participation de lecteurs à description, validée par les professionnels.

Malgré le développement de contenus créés par des utilisateurs, la dissociation des outils reste importante, avec diverses formes de catalogues en ligne validés par une institution ou une communauté.

3.2 DES CATALOGUES SUR LE WEB HORS DES BIBLIOTHEQUES

Les catalogues en ligne restent un mode d'accès courant aux produits. Les livres ne font pas exception, avec divers catalogues de vente, arrivant parmi les premiers résultats lorsque l'on cherche le titre d'un ouvrage récent dans un moteur. En revanche les catalogues de bibliothèques, généralement non indexables par des moteurs, sont probablement sous-exploités. Il faut donc adapter les pratiques anciennes, en conservant un outil qui garantisse l'accès aux ressources avec une description validée.

3.2.1 Les catalogues privés

Dans le contexte du Web, le terme de catalogue est parfois utilisé pour désigner des descriptions d'objets disparates ne constituant pas des collections. A l'inverse, on y désigne parfois autrement des instruments de recherche qui s'apparentent tout à fait à ceux des bibliothèques.

Nous prenons donc ici la liberté de désigner par le terme de « catalogue » dans son sens bibliothéconomique, pour désigner l'ensemble des outils de recherche s'apparentant aux catalogues de bibliothèques et aux services qui y sont associés.

- Catalogues thématiques

Le premier exemple d'instrument de recherche organisé thématiquement est la liste de sites, qu'elle soit élaborée de façon institutionnelle ou collaborative.

Les annuaires avaient constitué un des premiers moyens d'organiser la recherche d'informations aux débuts de l'Internet, avec Yahoo !¹⁸⁵ depuis 1995.¹⁸⁶

Ils ont trouvé un écho dans divers domaines spécialisés, et en particulier dans le monde des bibliothèques, se prêtant à la fois à la question de la validation des contenus et à l'éclectisme des domaines couverts. On peut citer les signets de la BnF,¹⁸⁷ un des premiers services de ce type mis en place. Ces principes d'élaboration ne sont pas complètement opposés à celui des moteurs, puisque le critère des liens établis notamment depuis des annuaires intervient pour définir la pertinence dans les algorithmes de classement des résultats des moteurs.

¹⁸⁴Avec la mise en place d'une indexation de l'Etat Civil par les lecteurs internautes, <http://www.lamayenne.fr/?SectionId=397>. Des projets comparables existent aux archives Municipales de Rennes et aux Archives Départementales des Yvelines.

¹⁸⁵ LELOUP, Catherine. *Moteurs d'indexation et de recherche : Environnements client-serveur, Internet et Intranet*. Paris: Eyrolles, 1997, <http://fr.dir.yahoo.com/>

¹⁸⁶ *Ibid.*

¹⁸⁷ <http://signets.bnf.fr/>.

Les catalogues thématiques de sites conservent plusieurs atouts majeurs à l'heure actuelle : permettre que des ressources soient validées par une institution ou une collectivité, et faire émerger des ressources difficiles à trouver, qu'il s'agisse à proprement parler du *Web profond* non indexé par les moteurs, ou de sites mal référencés et donc mal classés lors des recherches. En outre, ils permettent à l'utilisateur de se confronter à une liste complète et cohérente, rassurante par rapport à la dimension aléatoire des recherches par moteurs.

Ils connaissent en fait un développement croissant avec la participation des usagers au contenu, qui permet de couvrir des domaines très vastes, et s'attache à l'idée de mettre en valeur des sites recommandés. Parmi les catalogues de sites établis par des particuliers, Delicious¹⁸⁸ permet le partage de favoris avec une indexation par mots clés. Il s'agit de combiner une indexation en langage libre avec une synthèse effectuée sur les indexations en langage libre des utilisateurs précédents, constituant en définitive une liste de vedettes.¹⁸⁹ De cette façon, « Digg »,¹⁹⁰ qui propose des sites recommandés par les internautes, donne accès à son « laboratoire »¹⁹¹ avec des schémas qui montrent en temps réel les évolutions selon les votes. Lorsqu'on parle d'indexation par les utilisateurs, il s'agit donc de mécanismes complexes, déjà étudiés dans des travaux de ce type.¹⁹² On peut retenir l'idée que ce type d'outil est tout à fait d'actualité, pour accéder aux ressources et pour intégrer les utilisateurs à la description.

A l'inverse de ces outils, où la description correspond à une fusion de différentes contributions, il existe aussi des catalogues où un utilisateur seul produit une description. Il s'agit des catalogues de ventes privées, encadrés par des organismes (comme Ebay¹⁹³ ou PriceMinister¹⁹⁴), mais où les articles sont mis en vente à titre personnel, et où le vendeur établit sa notice de façon autonome, avec une description libre. Pour autant, ces notices sont organisées selon une structuration à la fois typologique et thématique tout à fait fonctionnelle, pour permettre aux acheteurs de trouver les articles. La très grande disparité de qualité de descriptions (par exemple pour les livres anciens sur Ebay) montre bien que cet outil ne fonctionne que pour une audience disparate et pour des produits disponibles dans un temps court.

En revanche, l'approche du mode de navigation des clients est intéressante pour les catalogues de bibliothèque, puisqu'on y prévoit une utilisation conjointe de descripteurs libres et d'une arborescence fixe de classement, à la fois typologique et thématique.

¹⁸⁸ <http://www.delicious.com> (deli.cio.us a changé de nom septembre 2008, Voir <http://www.figoblog.org/node/1928>.)

¹⁸⁹ Delicious intègre les « popular tags » les plus souvent proposés par les utilisateurs dans les « recommended tags » proposés au moment de la sauvegarde d'une URL dans ses signets. <http://del.icio.us/help/tags>, cité par PEYRARD, Sébastien, *L'accès aux collections numériques de masse*, Mémoire ENSSIB, 2007, 85 p.

¹⁹⁰ <http://www.digg.com>

¹⁹¹ <http://labs.digg.com>

¹⁹² LEBRETON, Claire, Bibliothèques, tags et folksonomies : L'indexation des bibliothèques à l'ère sociale, Mémoire ENSSIB, 2007, 117 p.

¹⁹³ <http://www.ebay.fr/>.

¹⁹⁴ <http://www.priceminister.com/>.

- Catalogues institutionnels de vente

Pour de nombreux magasins et enseignes commerciales, le catalogue en ligne est un moyen d'accéder à une description des produits, et un reflet de l'activité de l'entreprise. Ce double reflet des ressources et de l'institution est identique pour les bibliothèques.

Le premier élément d'évolution des catalogues commerciaux en ligne est souvent l'intégration du client. Ainsi, de nombreuses possibilités habituellement associées au « Web 2.0 » existent depuis la fin des années 1990 sur Amazon, leader de la vente d'ouvrages en ligne. Il s'agit d'intégrer le consommateur comme prescripteur, pour les conseils eux-mêmes et en raison de la force de vente d'arguments donnés par des personnes n'ayant pas d'intérêt direct dans la transaction. Depuis dix ans, sur Amazon, le conseil du libraire est donc remplacé par les commentaires des internautes.

Une offre complémentaire est permise avec des blogs d'auteurs parfois liés à leurs livres. Un algorithme de recommandation suit la navigation de l'utilisateur, selon l'indexation thématique des ouvrages et les achats communs effectués par d'autres clients, pour proposer d'autres titres. Plus encore, un programme de fidélisation propose depuis 2004 l'ouverture d'un compte client avec un « Plog ». C'est en fait un blog client¹⁹⁵ permettant de créer une interface intégrant des recommandations et des contenus de blogs d'auteurs auxquels le client est abonné. Il s'agit d'un moyen de fidélisation permettant au client de s'approprier les ressources. On pourrait tout à fait envisager des fonctions comparables pour les catalogues de bibliothèques.

Il reste cependant plusieurs différences notables entre les catalogues de vente et ceux des bibliothèques. En premier lieu, la vente par correspondance est associée à presque tous les catalogues commerciaux en ligne, ce qui les place dans un contexte différent de celui des bibliothèques, qui nécessitent toujours un déplacement pour la consultation des ouvrages matériels, même s'il peut s'agir d'une bibliothèque plus proche dans le cadre du PEB. Surtout, les catalogues de vente s'inscrivent dès le départ dans une logique de dissociation typologique : qu'il s'agisse de la Fnac,¹⁹⁶ de la Redoute¹⁹⁷ ou de Virgin,¹⁹⁸ il faut rechercher un support pour une navigation traditionnelle. A l'inverse, le problème actuel des OPACs est de permettre l'orientation vers des ressources multiples qui émanent toutes de la bibliothèque.

3.2.2 Les livres en ligne

Pour les bibliothèques, la gestion et la description de différents supports sont des problèmes anciens. Pourtant, l'approche change avec les modes de recherche en plein texte, le livre pouvant devenir un simple objet numérique. En effet, contrairement aux images et vidéo, son contenu peut répondre au mode de recherche textuelle. Ce qui concerne aussi bien la recherche d'ouvrages papier signalés en ligne, que la recherche d'informations en ligne éventuellement contenues dans des ouvrages.

¹⁹⁵ Visible si le consommateur active son compte.

¹⁹⁶ <http://www.fnac.com/>.

¹⁹⁷ <http://www.laredoute.fr/>.

¹⁹⁸ <http://www.virginmegastore.fr/>.

- Trouver des ouvrages papier

La recherche d'ouvrages papier sur le Web a provoqué plusieurs changements d'habitudes, aussi bien en ce qui concerne les pratiques de lecture que du point de vue des modes de recherche.

L'évolution des pratiques a été étudiée par Chris Anderson,¹⁹⁹ qui a développé le concept de « *long tail* »,²⁰⁰ notamment dans son ouvrage « La longue traîne, la nouvelle économie est là », en étendant aux livres des considérations qui s'appuyaient au départ sur la vente des films et des disques.²⁰¹ Bien que le travail s'appuie au départ sur des services comme Netflix²⁰² (Location de DVD) ou iTunes,²⁰³ cette théorie éclaire le développement de sites marchands comme Amazon et Fnac.com, selon un modèle commercial différent de leurs homologues ou branches physiques. En comparant les modes de distribution des produits culturels en magasin et en ligne, on constate que la vente en ligne a modifié les pratiques de consommation. Dans la distribution classique, les *hits* et *best-sellers* constituent la majorité des ventes, puisque 20% des produits représentent 80% du chiffre d'affaires, suivant le « principe de Pareto ». Ce mode de distribution met en avant certains produits au détriment d'autres. Or, sur le Web, même les ouvrages très peu vendus continuent à être recherchés et achetés.²⁰⁴

Ainsi, là où les deux premières surfaces de vente de livres en France, c'est-à-dire la Fnac et Leclerc, proposent de 20000 à 80000 références de livres selon leurs magasins,²⁰⁵ leurs homologues en ligne Fnac.fr et Amazon.fr proposent 600000 à 900000 références,²⁰⁶ ce qui reflète des évolutions importantes des modes de consommation, comme l'a montré Jean-François Gervais²⁰⁷.

Ces évolutions sont très importantes pour l'action des bibliothèques, aussi bien pour la mise en place d'un signalement collectif efficace, que pour la description des multiples ressources d'un même établissement. En effet, toutes les notices sont traditionnellement présentées sur un mode identique dans les OPACs. En revanche, comme dans les catalogues commerciaux, diverses bibliothèques entreprennent de mettre en valeur des ouvrages particuliers, comme à la Bibliothèque de la Cité des Sciences et de l'industrie, où différents livres sont présentés sur la première page du catalogue en fonction des expositions présentées dans la Cité.²⁰⁸ En ce cas, l'objectif est d'éclairer un aspect culturel et de prolonger ses découvertes par un approfondissement livresque, qui n'a pas de vocation commerciale mais une simple volonté de succès, où la mise en valeur des ressources de la bibliothèque est importante.

Avec la mise en place d'objets numériques intégrés aux ressources de la bibliothèque et disponibles depuis le catalogue, la question de la « longue traîne » est essentielle. Elle implique en effet que l'inflation sur le nombre d'ouvrages à numériser, avec la mise en

¹⁹⁹ Rédacteur en chef de Wired, magazine de référence de la cyberculture créé en 1993

²⁰⁰ Dans un article paru en octobre 2004, <http://www.wired.com/wired/archive/12.10/tail.html>.

²⁰¹ « The long tail : why the future of business is selling less of more » (p.166 bis de notes)

²⁰² Netflix, site de location de DVD, <http://www.netflix.com/>.

²⁰³ iTunes, <http://www.apple.com/fr/itunes/download/>.

²⁰⁴ En d'autres termes cette courbe des 80/20 devient rapidement asymptotique, c'est-à-dire très faible mais pas inexistante.

²⁰⁵ http://www.e-leclerc.com/c2k/portail/enseigne/enseignes_fiche_culturel.asp

²⁰⁶ <http://www.journaldunet.com/0509/050908amazon.shtml>

²⁰⁷ GERVAIS, Jean-François. Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups.. Paris: Dunod, 2007, 216 p., p.184-186.

²⁰⁸ <http://www.cite-sciences.fr>, puis « bibliothèque ».

place de programmes de numérisation de masse,²⁰⁹ correspond au besoin de nourrir une courbe asymptotique, dont l'étendue va loin au-delà du nombre d'ouvrages consultables au cours d'une vie humaine, mais dont l'utilisation potentielle par des internautes constitue une justification valable. En effet, le point principal de la théorie de la *long tail* est l'opportunité de marchés de niches, qui se traduit par l'efficacité possible de publications très spécialisées pour un public spécialisé, et par le bien-fondé de rendre disponibles des ressources diverses et spécialisées, leur communication correspondant à une demande. De cette façon, la diffusion d'ouvrages numériques inscrit les bibliothèques dans une démarche particulière, où l'ouvrage est assimilé aux contenus particuliers au Web.

- Chercher dans le texte des livres

La possibilité d'accéder au contenu d'ouvrages numérisés a provoqué une « révolution culturelle » en bibliothèque. Disposer d'outils pour interroger le contenu des documents sembla normal en centres de documentation, mais bouleversait en revanche une culture professionnelle attachée à des descriptions expertes et incontournables. Pour accéder aux documents, le fait de disposer d'un fichier de texte permet de gérer des unités minimales, avec des entrées d'index. Le problème principal pour l'accès aux ressources disponibles en mode texte est en fait la définition des unités décrites. Par exemple, dans une encyclopédie en plusieurs volumes, c'est l'article qui est une unité, mais l'ouvrage en lui-même peut aussi en constituer une. Par exemple, le Larousse du XIX^e siècle, numérisé,²¹⁰ peut être décrit à différents niveaux. Pour ces ressources, la possibilité d'effectuer des recherches en *full text* est donc essentielle, à moins de se limiter à une organisation thématique ou aux modes d'accès déjà permis par le papier. Rechercher dans le texte des livres ne s'oppose donc pas au besoin d'élaborer des descriptions bibliographiques. Bien au contraire, cette possibilité crée un besoin de balisage pour permettre à l'utilisateur d'identifier rapidement les ressources présentées comme résultat d'une recherche, et pour élaborer des réponses prenant en compte les éventuelles différences de pertinence entre les documents.

La façon de concevoir intellectuellement la bibliothèque change de façon importante, avec un nouveau mode d'accès aux collections, y compris en vue de trouver des documents papier. On peut en particulier citer les fonctionnalités offertes par Google Scholar,²¹¹ avec l'interrogation des ressources de Google Books. En effet, il est possible d'effectuer une recherche dans des contenus numérisés, même dans les ouvrages sous droits, qui ne peuvent être consultables librement, pour des raisons légales. En ce cas, le moteur les utilise pour l'élaboration des résultats puis ne laisse lire que les extraits de passages contenant les termes. Ce système puissant et utilisé illustre le problème commun des banques d'ouvrages, où les données virtuelles sont stockées matériellement, mais accessibles uniquement par le langage, ce qui change la constitution intellectuelle de la bibliothèque. Surtout, l'intégration des ouvrages imprimés (même sous droits) dans un mode d'accès en plein texte encourage à chercher d'abord sur Google, puis à localiser éventuellement des exemplaires. C'est là qu'interviennent les instruments de recherche des bibliothèques, pour signaler la présence des ouvrages à partir des moteurs.

²⁰⁹ Par exemple les 30000 (contrat avec Jouve) et avec Safig à la BnF.

²¹⁰ Larousse en ligne : sur <http://www.larousse.fr/>

²¹¹ <http://scholar.google.fr>.

3.2.3 Comment les bibliothèques s’y intègrent

Dans ce contexte, les ressources des bibliothèques peuvent être mise en valeur autrement que par leurs OPACs. Il s’agit de se situer sur les chemins de navigation des internautes, de traiter efficacement la diversité des ressources, et d’utiliser les nouveaux outils.

- Sur le chemin des internautes

Etre visible sur le parcours des internautes peut supposer des démarches diverses : apparaître dans les moteurs, dans les applications Web, ou dans les circuits commerciaux de recherche d’ouvrages. Même le plus grand service de catalogue au monde, OCLC, estime que les moteurs constituent désormais un moyen normal pour trouver des livres.

De cette façon, des accords existent entre des catalogues ou agrégateurs (comme OCLC²¹²), et des moteurs dont Google et Yahoo, pour être bien indexés par ces moteurs et permettre d’être visibles pour des flux d’internautes importants. Cette façon d’orienter les utilisateurs vers les sites locaux suppose des négociations, comme pour les accords conclus par l’OCLC avec Yahoo, MSN et Google (en dernier), mais aussi pour des institutions plus modestes, comme les accords de l’INIST avec Google.²¹³ Surtout, la mise en valeur d’outils fédérant des catalogues ne met bien sûr en valeur que ceux qui y sont présents. En l’occurrence, l’OCLC-Worldcat a une vocation mondiale, mais les bibliothèques françaises y sont pour le moment peu présentes.²¹⁴ Il faut remarquer l’ambivalence de ce genre d’outil : le fait de s’y intégrer permet aux ressources d’être mieux repérées, mais les outils collectifs présentent le danger d’encourager à se limiter aux résultats proposés. La logique des actuels moteurs consiste à toujours proposer un résultat, même non pertinent, en amenant à reformuler la question plutôt que de changer de moyen de recherche. Or, les bibliothèques peuvent avoir un fonctionnement inverse, en renvoyant volontiers l’utilisateur vers des institutions voisines spécialisées. Les catalogues collectifs amènent donc à envisager une démarche du lecteur différente, avec des résultats divers fournis avec la même interface et non au terme d’une navigation. Ces outils sont considérablement améliorés par l’intégration de *mashups* permettant des techniques de géolocalisation, à partir des adresses IP des utilisateurs. Il s’agit de repérer automatiquement quels établissements se situent à proximité de l’utilisateur. C’est ainsi que Google Book Search -Via Worldcat- propose simplement de « trouver ce livre près de chez vous », d’une façon plus intuitive que les localisations manuelles.

L’évolution des usages implique des changements d’habitudes, qui peuvent rendre pertinente l’intégration des bibliothèques dans des applications Web. De fait, les internautes sont marqués par l’ergonomie du Web, de la façon la plus générale puisque la masse de ceux qui n’utilisent que les outils les plus simples sont probablement ceux qui se limitent aux interfaces intuitives. Pourtant, les catalogues ne sont pas sur le chemin naturel des internautes, celui des moteurs de recherche et des sites de partage. Il est habituellement admis, selon Lorcan Dempsey²¹⁵ et des auteurs comme Hervé Le Crosnier²¹⁶ en France, que le catalogue de bibliothèque doit être là où sont les internautes, c’est-à-dire dans Google et dans les applications Internet habituelles.

²¹² <http://www.oclc.org>.

²¹³ Pour l’interrogation des données de Pascal et Francis par Google.

²¹⁴ Bien que BN Opale Plus (BnF) et le SUDOC (ABES) soient en passe d’y être intégrés.

²¹⁵ <http://orweblog.oclc.org/>.

Il s'agit d'attirer l'attention en faisant apparaître le catalogue lui-même, des éléments du catalogue, ou directement des ressources.

Cette intégration pose plusieurs problèmes de légitimité. D'une part, pour ne pas polluer les recherches, il est important que seuls les ouvrages rares ou réellement disponibles pour l'utilisateur apparaissent. D'autre part, il est délicat de définir dans quelle mesure les bibliothèques publiques doivent s'intégrer et améliorer des outils éventuellement contestables. L'intégration d'outils de bibliothèque dans des réseaux de sociabilité comme Facebook ou Myspace correspond à une prestation gratuite pour l'utilisateur, mais liée à un aspect commercial fort,²¹⁷ dégagant des profits importants et entretenant une relation particulière aux informations à caractère personnel. Plus encore, la présence de certaines bibliothèques dans des jeux en ligne, notamment la Library Of Congress sur *Second Life* avec la possibilité de consulter des ouvrages numérisés et d'accéder au catalogue,²¹⁸ correspond à une démarche de communication hors de la sphère habituellement dévolue aux bibliothèques.

L'intégration dans les circuits commerciaux proprement dits pose une question similaire, même si la situation est souvent plus claire du fait de la nature contractuelle des accords conclus. En tous cas, les catalogues de vente de produits culturels donnent toujours accès aux ressources en en distinguant les types de support et les modes d'acquisition, ce qui se prête tout à fait à l'intégration de recherches dans les collections des bibliothèques, quels que soient les types de documents et de supports.

- Dans des services existants ?

L'imitation voire l'intégration des bibliothèques dans des catalogues commerciaux pose la question des objectifs à atteindre. Il s'agit d'une possibilité intéressante, d'une part parce que certains catalogues commerciaux sont bien référencés et apparaissent en tête lorsqu'on cherche un produit culturel, ce qui permet d'attirer l'attention des utilisateurs, et d'autre part parce que les catalogues commerciaux en ligne sont, pour certains, connus et situés sur le chemin habituel des internautes. Notamment, Amazon, le premier site de vente de livres en ligne,²¹⁹ est désormais ouvert au développement de *mashups*, avec Amazon Web Services (AWS), la plate-forme d'Amazon pour les développeurs.²²⁰ Les bibliothèques du Royaume Uni s'y intègrent, à la suite d'un accord. Le principe est d'attirer l'attention d'internautes souhaitant se procurer un livre, en les géolocalisant à l'aide d'un *mashup* sur Amazon, plutôt que d'avoir à utiliser le site particulier d'une bibliothèque.²²¹ L'intégration de Worldcat dans Google Books a un objectif identique : permettre de repérer les ouvrages en bibliothèque au cours d'une navigation fluide, et alors qu'on n'avait pas forcément envisagé cette possibilité.

²¹⁶ Voir par exemple les conférences d'Hervé Le Crosnier sur le « Web 2.0 », <http://www.gin-ebisi.umontreal.ca/confmidi/2006/web20/index.html>.

²¹⁷ GERVAIS, Jean-François. Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups.. Paris: Dunod, 2007, 216 p. p. 110 : sur l'achat de Myspace par Rupert Murdoch en 2005 pour remplacer les études d'audiences, dans l'idée que « là où il y a de l'audience, il y a de l'argent » en analysant les profils de Myspace.

²¹⁸ Avec le Reference Desk d'Info Island, la "terre des bibliothèques" de Second Life.

²¹⁹ <http://www.amazon.fr/>

²²⁰ <http://imagine-it.org/amazon/vissimweb.htm>.

²²¹ GERVAIS, Jean-François. Web 2.0 : les internautes au pouvoir... *Op. Cit.*, p. 188.

L'intégration des catalogues des bibliothèques dans l'environnement habituel des internautes peut s'effectuer de façon moins ambitieuse. En particulier, la présence d'un lien vers le catalogue des B.U. dans les portails d'universités, voire dans les Environnements Numériques de Travail d'étudiants, est de plus en plus répandue. C'est un usage habituel des bibliothèques américaines, où la présentation de la bibliothèque la fait apparaître comme un des lieux de travail principaux, sur le portail de l'université.²²² On peut contester la volonté d'appliquer cette démarche hors du contexte universitaire, notamment avec l'intégration éventuelle d'un catalogue dans les ENT des écoles, qui va peut-être au-delà des besoins des utilisateurs.²²³ En revanche, l'intégration dans l'environnement de travail universitaire peut permettre de découvrir des ressources, en particulier avec la mise en place de métamoteurs fédérant les ressources,²²⁴ puisque ces outils dépassent le simple lien vers le catalogue.

Le catalogue lui-même peut intégrer divers services à distance, qui pour certains existent depuis longtemps et apportent de vrais avantages pour l'utilisateur, en permettant d'économiser du temps et d'être certain que le service qu'il attend lui sera rendu. Il peut s'agir de services correspondant à des fonctions du SIGB, comme la réservation d'ouvrages, utilisée depuis longtemps pour des bibliothèques de recherche comme le Rez-de-jardin de la BnF, mais utile aussi pour des bibliothèques de lecture publique, en permettant aux usagers de préparer des déplacements utiles, ainsi que des fonctions comme le suivi de l'ouvrage dès la commande pour éviter un second traitement via les suggestions d'acquisition, ou que le lecteur aille se procurer autrement l'ouvrage attendu. Ces fonctions de réservation de places et de prolongation des prêts, font partie du soin d'une image favorable en apportant des services dont l'utilisation doit être intuitive, ce qui nécessite d'élaborer des outils adaptés, y compris pour permettre l'appropriation des ressources par les utilisateurs.

- Changements d'outils

Dans l'idée que les informations primaires et secondaires interviennent dans les normes de service et l'évaluation, la qualité du catalogue fait partie du service. Mais sa fonction dépasse celle de l'outil local de recherche topographique et de contenu. C'est non seulement un outil de proximité permettant de réserver des places et des ouvrages, ou de prolonger un prêt, mais aussi un moyen de copier des documents libres de droits. Cette possibilité de télécharger ce qui peut l'être pose des questions de gestion des ressources, et de public, puisque le catalogue n'est plus sur le Web mais dans le Web, s'adressant non seulement aux chercheurs ou aux publics proches de la bibliothèque, mais potentiellement à tous les internautes. Les outils mis en place peuvent participer aux techniques de veille sur le Web et de validation des informations. En particulier, les outils de gestion des références bibliographiques²²⁵ encouragent à mettre en place des données collectables avec des balises compatibles et des données extractibles.²²⁶ L'intégration d'outils collaboratifs permet à la fois l'amélioration par les utilisateurs participant à la description des contenus, et une forme d'appropriation. Enfin, l'interrogation des données peut être regroupée, avec un lien entre les institutions et

²²² Par exemple la Johns Hopkins University Library, <http://webapps.jhu.edu/jhuniverse/libraries/>.

²²³Voire en créant des blogs et services spécialisés pour des actions de médiation en ligne. <http://www.bibliobsession.net/tag/comprendre-les-enjeux-de-demain/>.

²²⁴ Comme la Jubilothèque de Paris 6-Jussieu, <http://jubil.upmc.fr>.

²²⁵ Comme Zotero et Refworks.

²²⁶ Comme le propose notamment le SUDOC, <http://www.sudoc.abes.fr>.

éventuellement des bases de données qualifiées par les bibliothèques et les utilisateurs, comme le prévoyait l'idée de méta-moteur fédérant le CNERTA, la BPI, et la CSI.²²⁷ Au travers de ces outils, se pose la question des protocoles et des réseaux. Si le Web simplifie et uniformise, encourageant à l'interopérabilité des descriptions, il faut étudier la façon dont les modes de description sont changés, et dont s'intègrent les objets numériques des bibliothèques, musées et archives.

²²⁷ Projet de 2006, par Algora, voir http://www.centre-inffo.fr/IMG/pdf_edl_bureautique_1_.pdf

4. L'avenir des catalogues dans le contexte du Web

Avec l'importance du langage XML et de l'implication des usagers, l'évolution des standards de description des données laisse entrevoir des possibilités bien plus grandes pour l'avenir des catalogues. Le contexte du Web tend à imposer des formats standards, essentiellement HTML²²⁸ et XML²²⁹. Ce ne sont donc pas des formats spécifiques, mais des formats génériques correspondant à des usages professionnels différents, et liés aux outils de traitement de l'information. La description des ressources doit en effet répondre à de multiples fonctions, ce qui peut laisser entrevoir des évolutions profondes du Web.

4.1 LES METADONNEES A USAGE MULTIPLE

Le terme de métadonnées est issu des langages à balises, où des « données sur les données » sont intégrées au document. Elles sont utilisées dans d'autres secteurs que le Web, comme la téléphonie, et pour des sources de toutes natures, comme le texte, l'image, le son ou la vidéo.

4.1.1 Catalogue des métadonnées

Elles n'ont pas vocation à être perçues telles quelles par l'utilisateur final, mais permettent la description, la structuration et la gestion du document, selon des règles de création essentielles pour son utilisation.

- Définition

Le terme de métadonnées désigne, depuis le milieu des années 1990, des données renseignant un ensemble de données numériques auquel elles appartiennent. Par extension, on utilise aussi ce terme pour désigner les données renseignant d'autres contenus, même non numériques.²³⁰ Et inversement, même pour des ressources électroniques, il est possible de séparer physiquement les données de leurs métadonnées.

²²⁸ Format fermé (avec une liste de balises est fixe et non modifiable) utilisé pour presque tout le Web.

²²⁹ Format ouvert utilisé selon des DTD et schéma multiples, permettant même la création de balises, et nécessitant donc l'interprétation par une feuille de style XSLT ou un système des feuilles de style CSS pour que les navigateurs Web (IE, Firefox,...) le transforment en Html.

²³⁰ Ainsi, le modèle OAIS, qui fait référence pour la préservation du numérique, et peut convenir à des contenus physiques. Modèle de référence OAIS, p. 11 : « l'accent est mis sur l'information numérique, d'une part comme forme principale de l'information pérennisée, d'autre part comme information complémentaire de ce qui est archivé aussi bien sous forme numérique que sous forme physique. Par conséquent, le modèle s'adapte aux informations qui sont par nature non numériques ».

Disponible sur http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf

La NISO²³¹ les définit simplement comme une « information structurée » qui décrit pour retrouver, tandis qu'il s'agit selon l'ISO de « données qui définissent et décrivent d'autres données ». ²³² Il s'agit en tout cas d'outils permettant la gestion de l'information numérique, en documentant son cycle de vie et les droits qui s'y attachent, et en permettant l'interopérabilité entre les diverses sortes de ressources présentes sur le Web. Elles peuvent, selon les cas, distinguer plusieurs niveaux de granularité d'information, depuis le niveau le plus fin du fichier numérique, jusqu'au niveau le plus global de la collection. Ce qui suppose de déterminer les différents niveaux de description utiles, ²³³ et de structurer les descriptions avec efficacité.

Les métadonnées sont de diverses natures. La distinction entre les métadonnées descriptives, techniques et administratives correspond donc à des différences essentielles. ²³⁴

- Métadonnées descriptives

Les métadonnées descriptives permettent d'identifier et de retrouver des ressources. C'est le plus souvent d'elles que l'on parle lorsque l'on évoque les « métadonnées » sans plus de précision, car ce sont les plus apparentées au travail de description bibliographique traditionnel. Au lieu de créer une notice décrivant une information primaire, on crée des données secondaires décrivant des données primaires. Elles correspondent donc à la forme informatique du catalogage proprement dit, puisqu'il s'agit des données de description et d'identification, depuis les plus simples (auteur, titre) jusqu'à la forme plus complexe que ces informations peuvent prendre, en particulier le renvoi à des données d'autorités, au travers de références, éventuellement via des URL pérennes. En effet, l'intégration de renvois est possible, y compris dans des champs principaux du Dublin Core²³⁵ avec des identifiants.²³⁶ Il s'agit toujours d'un travail de description bibliographique. Les métadonnées descriptives sont donc souvent extractibles depuis les catalogues « anciens », et correspondent au même contenu.²³⁷

Les données de valorisation et de communication peuvent y être assimilées, en tout cas celles qui permettent le lien vers des éléments associés, pouvant ensuite être rattachés au document décrit, comme une image de couverture, un résumé ou des commentaires. En revanche, les métadonnées permettant d'établir sous une forme numérique l'intégrité, les composantes et la cohérence du document décrit ne sont pas descriptives, mais structurelles.

²³¹ National Information Standards Organisation (NISO) Understanding metadata, 2004, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

²³² ISO / IEC 11179-3 « Metadata : data that defines and describes other data »

²³³ Il s'agit de déterminer le niveau documentaire, c'est à dire celui que l'utilisateur manipulera en priorité, par défaut (de la collection au fichier).

²³⁴ Voir LUPOVICI, Catherine, In BURESI, Charlette; CÉDELLE-JOUBERT, Laure. *Conduire un projet de numérisation*. Villeurbanne : ENSSIB ; Paris : Éd. Tec et Doc: Lavoisier, 2002.

²³⁵ Notamment dc:relation et dc :identifier, Cf. Infra

²³⁶ Voir <http://bibnum.bnf.fr/identifiants>, avec notamment les identifiants ARK (Archival Resource Key), créés par la California Digital Library pour donner un identifiant unique à chaque notice bibliographique, et utilisés notamment par la BnF dans BN Opale Plus et Gallica.

²³⁷ Voir Marie-Elise Freon, Journée d'information AFNOR CG 46, 7 juin 2005, Les métadonnées de préservation, Journée d'information AFNOR CG 46, <http://www.bnf.fr/pages/infopro/journeespro/pdf/AFNOR2005/Metadonneespreservation.pdf>

- Métadonnées de structure

Aussi appelées « techniques », les métadonnées structurelles reflètent la granularité des ressources. En effet, les fichiers informatiques séparés doivent permettre de reconstituer techniquement et intellectuellement les liens entre les niveaux de description, pour permettre une navigation efficace en énonçant explicitement les relations entre les composants de la ressource. Elles permettent de décrire des objets numériques constitués de fichiers éclatés. On doit en effet conserver l'unité des ouvrages et des collections en rattachant les fichiers, au travers des liens entre ces composants. C'est ainsi qu'on reconstitue la structure logique et physique des documents numérisés, et que l'on rend explicite celle des documents numériques natifs.

Il s'agit aussi bien des publications numérisées (avec les niveaux de collection, document, volume, page et détail) que des documents graphiques par exemple (avec des collections, des lots, des images et des vues de détail).²³⁸ Le principe est en tout cas d'ordonner utilement un entrepôt de données, en conservant une structure qui permette de respecter la granularité des ressources.

Il existe deux types de cartes de structure : physique, restituant la navigation linéaire dans les fichiers du document,²³⁹ et logique, permettant une navigation non linéaire à partir de renvois et de liens.²⁴⁰ Les problèmes de navigation et de granularité se posent parfois différemment, selon les formats employés.²⁴¹ Ils sont cependant liés, avec la définition des différents niveaux. On trouve le niveau le plus global, qui peut agréger plusieurs documents,²⁴² éventuellement de façon multiple,²⁴³ le niveau de l'unité documentaire, correspondant aux regroupements physiques manipulés,²⁴⁴ le niveau des entités logiques qui constituent le document,²⁴⁵ et enfin le niveau technique des fichiers du document numérique.²⁴⁶

- Métadonnées administratives

- Droit :

La gestion des droits et des accès, ainsi que la préservation des documents, sont absolument essentielles dans un contexte numérique. Pourtant, les métadonnées concernant ces aspects ne sont ni descriptives ni structurelles. Il s'agit de métadonnées administratives, qui comprennent les aspects techniques, de source, de droits et de

²³⁸ Voir Emmanuelle Bermès, « Les métadonnées à la Bibliothèque nationale de France : Une vue d'ensemble », Diaporama, 13/05/2005,

http://cdsweb.u-strasbg.fr/MDA/mda/ontologies/BNF_CDS20050513.ppt

²³⁹ En général, page après page.

²⁴⁰ Notamment au moyen de tables et d'index

²⁴¹ Ainsi le METS permet de superposer plusieurs cartes de structure dans la section Structmap, alors que le MPEG-21 DIDL a une structure unique, en « poupées russes ».

²⁴² Par exemple un titre de périodique, une monographie en plusieurs volumes, un document multimédia multisupports.

²⁴³ Par exemple une collection de périodiques.

²⁴⁴ Comme la monographie, le volume de monographie en plusieurs volumes, le fascicule de périodique.

²⁴⁵ Pages dans un livre, pistes d'un CD.

²⁴⁶ Dans le cadre de l'implémentation du METS à la BnF pour SPAR, ces quatre niveaux de granularités sont désignés sous les termes de « set », « group », « object », et « file », avec une carte de structure « structmap » intégrant les fichiers de structuration de production avec les éléments de renvois (dans refNum, avec des données de production, bibliographiques, et les informations pour l'archivage des données). Voir Marie-Elise Freon, Journée d'information AFNOR CG 46, 7 juin 2005, Les métadonnées de préservation, Journée d'information AFNOR CG 46,

<http://www.bnf.fr/pages/infopro/journeespro/pdf/AFNOR2005/Metadonneespreservation.pdf>

provenance du document. Les métadonnées techniques portant sur les fichiers peuvent être extraites de ces derniers à l'aide d'outils spécifiques.²⁴⁷

On regroupe en effet sous ce terme les données de gestion permettant de répertorier et de localiser un document source du fichier numérique, mais aussi de préciser des données juridiques, en particulier pour l'accès aux contenus sous droits dont les caractéristiques légales peuvent évoluer et sont intrinsèquement liées aux possibilités de lecture.²⁴⁸ Ces droits d'accès sont complétés par les droits liés à la manipulation du document,²⁴⁹ selon l'habilitation de l'utilisateur.²⁵⁰

- Préservation :

Enfin, on y trouve les données de préservation, qu'il s'agisse d'un document physique ou d'un document numérique. Elles sont bien sûr toutes essentielles pour les documents numériques, pour afficher, extraire et utiliser les collections de façon pérenne, puisque de nombreux paramètres doivent être connus pour pouvoir consulter le document.²⁵¹ La préservation implique donc, selon la terminologie OAIS,²⁵² l'information de représentation,²⁵³ et l'information descriptive de la pérennisation (PDI).²⁵⁴

4.1.2 Nouvelles règles de création des métadonnées

Une des premières difficultés de l'évolution des catalogues informatisés est que la notion de métadonnées est différente de celle de notices. En effet, le principe de la recherche tend à supprimer l'étape de la lecture d'une fiche entre le moteur et le document. Il s'agit donc d'adapter la description à des recherches directes.

- Décrire dans le document

Pour les ressources papier, y compris dans le contexte commercial, c'est la notice elle-même qui est l'objet de la recherche sur Internet, sauf dans le cas particulier des recherches en plein texte.²⁵⁵ Or, les métadonnées de droits peuvent être différentes, donnant accès ou pas au document. Que l'on recherche la notice en ligne d'un document physique, ou un document numérisé, l'évolution la plus logique est la même : intégrer la description à l'objet finalement consulté (notice d'ouvrage papier ou objet en ligne).

²⁴⁷ D'un point de vue technique, les formats peuvent être manipulés par l'« identification », où on détermine le format en détectant le type MIME du fichier (par exemple, "image/tiff" pour un fichier TIFF), et par l'opération de « caractérisation », où on extrait des informations précises sous la forme d'un fichier XML (par exemple avec JHOVE, programme développé par l'université de Harvard).

²⁴⁸ Avec le type de droits, la date de début, la durée, et la date de fin de droits.

²⁴⁹ Droit de modifier le document, de modifier ses métadonnées, ou de migrer ses fichiers existants.

²⁵⁰ Par exemple, la BnF utilise cinq contextes d'accès, du plus permissif au plus restrictif : Web, Etudes (sur place), Recherche (en rez-de-jardin), Hemicycle (sous accréditation spéciale), Préservation (professionnels de la BnF). La distinction des usages autorisés (consultation, copie, impression,...) peut être exprimée dans le format ODRL, ou en DIDL avec des items conditionnels.

Voir « Introduction à MPEG 21 », <http://www.figoblog.org/document1244.php>.

²⁵¹ Support matériel lisible, machines pour le lire, système d'exploitation, logiciel de lecture,...

Voir « les enjeux de la préservation numérique en 1 slide », <http://www.figoblog.org/node/1944>.

²⁵² http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf

²⁵³ Qui établit une correspondance entre le train de bits et l'information intelligible par un humain, en transformant une suite binaire en un document consultable.

²⁵⁴ Avec à la fois l'information de provenance (d'intégrité et d'authentification du fichier), et une information d'identification contextuelle liant vers les autres objets nécessaires à la préservation.

²⁵⁵ Par exemple dans Gallica 2, dont les ressources indexées en plein texte sont des documents numériques produits à partir de documents papier, avec des métadonnées de description bibliographique issues des notices.

Il convient donc de prendre en compte la relation univoque entre la description ou identification, et l'objet numérique. Les métadonnées font partie de l'objet. Au point qu'il est tentant d'automatiser tout ce qui peut l'être, avec l'équivalence de la production du document et de celle de ses métadonnées, renseignées dès l'étape de production.

Selon la façon dont elles lui sont intégrées, on distingue les métadonnées « encapsulées » (intégrées par balises dans le corps du document), « englobantes » (caractérisant l'ensemble et en général placées dans l'en-tête du code source), et « externes » (fournies dans un fichier séparé du document). La mise en place de ces schémas de description concerne toutes les ressources présentes sur le Web.²⁵⁶

Il est évident que les formats et les modes de description strictement bibliothéconomiques tendent à singulariser les ressources des bibliothèques. Or, même les ensembles de descripteurs utilisés en bibliothèque sont multiples, avec non seulement diverses DTD XML,²⁵⁷ mais aussi diverses versions de formats d'échange simplifiés comme le Dublin Core (Simple, Qualifié). La grammaire XML peut être en effet être exprimée sous forme de DTD ou de schéma. Il convient donc de répondre à l'explosion des formats ou « schémas de métadonnées » par des solutions permettant de les rendre interopérables. Pour des raisons pratiques liées aux multiples données déjà disponibles et au mode de production des descriptions, il est nécessaire de trouver des solutions compatibles avec l'utilisation des données déjà existantes.

En outre, la manipulation parallèle de documents papier et de ressources électroniques amène à utiliser à la fois un format de description comme un dérivé du MARC pour le papier, et un autre pour le numérique.²⁵⁸ Par exemple en bibliothèque Universitaire, on peut rencontrer parallèlement l'UNIMARC pour le papier, et des dérivés du Dublin Core pour le numérique.²⁵⁹ Les mêmes catalogueurs peuvent maîtriser des formats différents, mais il est difficile de faire apparaître les diverses ressources de façon satisfaisante.

- Les possibilités du RDA (Resource Description and Access)

Un des objectifs du RDA²⁶⁰ est donc, au travers de l'adaptation des FRBR et FRAD,²⁶¹ et grâce au langage XML, de décrire l'ensemble des ressources selon une logique de production de métadonnées, ce qui change la démarche de l'information bibliographique. Comme son nom l'indique, RDA vise à donner aussi bien la description que les accès des ressources, pour décrire numériquement les publications papier ou numériques. De cette façon, le successeur des AACR2 apporte plus qu'une simple reprise des règles de catalogage anglo-américaines, et est vraiment différent du fonctionnement des ISBD. Ainsi, d'un point de vue formel, l'emploi du XML permet de dissocier les contenus de la mise en forme.²⁶² Et la structuration basée sur les FRBR en Œuvre, Expression, Manifestation et Objet, est encore très nouvelle dans le monde des catalogues.

²⁵⁶ Voir Andy Powell: "Does Metadata Matter", diaporama sur

<http://www.slideshare.net/eduserfoundatio/does-metadata-matter?src=embed>

²⁵⁷ EAD, VRA, LOM,...

²⁵⁸ TENNANT, Roy, Library Journal, 8/15/2004, <http://www.libraryjournal.com/article/CA443949.html>

²⁵⁹ Les catalogueurs pouvant cataloguer en Unimarc sur le Sudoc (avec WinIBW, d'OCLC/PICA), et en Dublin Core pour les objets numériques

²⁶⁰ Resource Description and Access, Cf. Supra.

²⁶¹ FRBR (Functional Requirements for Bibliographic Records) et les FRAD (Functional Requirements for Authority Data)

²⁶² Pour l'interopérabilité avec des ressources documentaires non issues du monde des bibliothèques.

Il convient de rappeler que la réflexion sur le RDA correspond à des débats et travaux très récents, puisque ces règles seront publiées en 2009.²⁶³ Nous éviterons donc de trop entrer dans les aspects techniques, pour décrire surtout le contexte de développement.

Dans une perspective internationale, il n'est pas impossible que l'IFLA l'adopte, et qu'il soit utilisé en France, notamment par la BnF. Ce changement nécessitera une importante formation, pour la conversion des données existantes, et en raison du travail que suppose l'adaptation à un nouvel environnement. Le RDA, est actuellement étudié par la Library of Congress, la National Agricultural Library et la National Library of Medicine,²⁶⁴ ce qui correspond à une réflexion générale sur l'évolution des catalogues, au travers de l'organisation professionnelle, des changements de modèle de données et de structure, et même des modèles de production de l'information bibliographique. Cependant, à la fois pour des raisons d'économie financière et de respect de normes élaborées au niveau mondial, la BnF comme les autres établissements envisageant d'utiliser RDA comme code de catalogage s'il est adopté. En pratique, du fait de l'éclatement des bibliothèques publiques, il est probable que la BnF joue un rôle important de formation et de décision dans ce contexte. En ce sens, l'étude qui y est menée actuellement en interne sur la refonte du catalogue, s'intéressant aux produits possibles sur tous les formats, aura probablement une incidence sur les choix nationaux.

De cette façon, la réflexion de la BnF sur les logiciels de bases de données, au travers de nouveaux logiciels, essentiellement en XML, peut supposer des changements de format de description bibliographique. Diverses possibilités pourront alors être prises en compte. Par exemple, même si le MARC n'est pas manipulé et n'apparaît ni en entrée ni en sortie d'une base de données relationnelles, il n'est pas exclu de permettre d'extraire les données qui y correspondent, notamment pour les petites bibliothèques dérivant les notices de la BnF. De ce fait, le changement de la BnF vers un format en RDA n'exclurait pas que des produits puissent être livrés en UNIMARC ou en INTERMARC, voire en Marc 21 dans le cadre d'une coopération avec l'OCLC. Bien que l'utilisation de données en XML soit délicate pour le moment, il est important de conserver cette idée de création des données en récupérant des données créées ailleurs et assemblées dans un nouveau produit, en utilisant un *mash-up*. Il est en effet nécessaire de concilier des évolutions dans le contexte du Web, avec l'adaptation de schémas existants.

4.1.3 Adapter des schémas de données déjà existants

Le choix du format est essentiel pour tout projet de conversion rétrospective, de création ou d'évolution d'outils de description bibliographiques. Il faut concilier les particularités des collections décrites, les usages spécifiques que l'on prévoit, les possibilités d'exploitation par les systèmes informatiques et l'interopérabilité avec les outils auxquels les descriptions doivent pouvoir être intégrées.

Les schémas de métadonnées les plus employés pour les collections de bibliothèques sont tous conçus en XML, et il convient d'observer comment les outils peuvent évoluer vers une visibilité plus large et un désenclavement des domaines bibliothéconomiques.

²⁶³ En cours de relecture jusqu'au 2 février 2009, disponible sur le site : <http://www.collectionscanada.gc.ca/jsc/rda.html>

²⁶⁴ Voir Catalogblog, <http://catalogblog.blogspot.com/>, et les déclarations de la LOC notamment sur http://www.nlm.nih.gov/tsd/cataloging/RDA_Letter_050108.pdf

- Des schémas XML

Les ensembles de descripteurs de métadonnées les plus répandus sont issus du XML, mais leur multiplication a provoqué une relative confusion.²⁶⁵ Comme auparavant pour les formats MARC, c'est avant tout l'échange qui a été normalisé, avec la récente norme MarcXchange,²⁶⁶ proposant un modèle plus générique de MARC XML. Mais cette uniformisation n'a pas vocation à permettre plus que les échanges, et n'est pas un format exploitable en soi.

Les premiers schémas XML utilisés en bibliothèque correspondent à une simple transposition de l'information bibliographique produite auparavant selon les différents dérivés du MARC. Les données restent structurées en deux niveaux hiérarchiques seulement,²⁶⁷ et sont centrées sur la description de l'objet. Ces formats transposant l'information bibliographique en XML permettent une meilleure interopérabilité, et constituent sans doute une étape importante pour un nouveau traitement des données,²⁶⁸ mais ne résolvent pas les principaux problèmes que posaient les formats Marc. Ils favorisent en revanche l'uniformisation des données, avec l'hégémonie du MARCXML, issu du MARC 21. Pour observer comment l'information bibliographique peut évoluer vers des schémas de métadonnées applicables à tous types de documents, nous nous limiterons à l'étude des schémas répandus et polyvalents. Il s'agit principalement de MARCXML, MODS/MADS, METS et PREMIS, qui sont les plus importants à l'heure actuelle pour la description de ressources contemporaines.

En premier lieu, MARCXML est un schéma XML élaboré par la Bibliothèque du Congrès à partir de 1999 pour permettre une conversion réversible et sans perte d'information à partir du format MARC21, avec un outil de conversion²⁶⁹ et une structure simple. La Bibliothèque du Congrès en assure la maintenance, mais d'autres bibliothèques en utilisent des variantes ou formats comparables. Par exemple, la BnF utilise pour un usage interne un format semblable au MARC-XML.²⁷⁰ Ce qui permet à la fois de continuer à utiliser l'INTERMARC face à la faible pression des bibliothèques, et de mettre en place de nouveaux systèmes d'interrogation, avec notamment l'interrogation par « Mots-notices » dès la fin 2008, utilisant un algorithme de pondération, et une indexation de tous les mots de la notice. Il est pour ce faire nécessaire que les données soient prises dans un fichier XML, puisque les logiciels utilisés ne traitent que l'XML. En outre, l'information structurée permet l'application d'un algorithme de pondération, comme l'application d'une pondération forte pour l'auteur et titre dans Gallica 2. Cet atout de l'information structurée pour la pertinence des résultats est comparable quel que soit le type de ressources. Par conséquent, des

²⁶⁵ Cités par exemple sur <http://www.ifla.org/II/metadata.htm>, avec entre autres DC, VRA, MODS, ONIX, EAD, TEI Header, LOM, MPEG-7,...

²⁶⁶ MarcXchange, ISO 25577, 2008.

http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43005

²⁶⁷ Correspondant à la zone et sous-zone.

²⁶⁸ Comme nous l'avons vu, il s'agit notamment de MARCXML (format MARC21 en XML élaboré par la Bibliothèque du Congrès), de MODS/MADS (eux aussi élaborés par la Bibliothèque du Congrès, et basés sur MARC21), de XOBIS (un format original mis au point par la Bibliothèque Médicale de l'Université de Stanford), d'ONIX (par les éditeurs, américains notamment), et de BiblioML en France (UNIMARC en XML, élaboré par le Ministère de la Culture et intéressant principalement la France).

²⁶⁹ <http://www.loc.gov/standards/marcxml/marcxml-survey.php>

²⁷⁰ L'InterXMarc, élaboré par des informaticiens pour répondre au besoin technique d'appliquer une feuille de style XSL à la notice en InterMarc (et aux données locales) pour générer une notice dans un format XML donné (notamment les formats de métadonnées en XML exposés par les entrepôts OAI).

formats en XML visent à s'affranchir des contraintes du MARC pour s'adapter particulièrement à la diversité des supports.

De cette façon, la bibliothèque du Congrès a aussi élaboré, en 2002-2003, un schéma XML pour un MARC 21 simplifié. Il s'agit du format MODS, « Metadata Object Description Schema », qui se veut applicable à des ressources papier ou électroniques, et utilisable aussi par des personnes non spécialistes du catalogage. MODS traite l'information bibliographique en simplifiant les éléments du MARC 21. Ainsi, MODS utilise des balises en anglais à la place des balises numériques du MARC 21, en les rendant parfois plus précises, et regroupe des éléments du MARC 21 tout en ne reprenant qu'un sous-ensemble de ses champs (contrairement à MARCXML qui les reprend tous). MODS comporte environ 80 éléments, employés de façon fixe et non qualifiée, pour une meilleure interopérabilité. Bien que le format soit adapté à la description de ressources numériques, la reprise des éléments du MARC permet de réutiliser les données des catalogues existants. En pratique, l'application du format MODS est liée à celle de MADS, « Metadata Authority Description Schema », un schéma de description lui aussi développé par la bibliothèque du Congrès et publié en 2004, afin de compléter MODS pour la description des autorités. MADS emprunte à la fois au format MARC21 Authority data,²⁷¹ et au format MARC21 Bibliographic,²⁷² dont il utilise certains éléments spécifiques.

MARC XML et MODS sont les outils les plus répandus, en raison de la forte influence des bibliothèques et centres de catalogage américains qui les utilisent. Ils sont complétés par des schémas généraux de métadonnées.

De cette façon, on utilise des schémas généraux pour la description des objets numériques, avec les standards METS²⁷³ et DIDL²⁷⁴. METS, préconisé par la Bibliothèque du Congrès, est le plus employé dans le monde des bibliothèques. Il s'agit d'un format d'implémentation permettant de combiner des métadonnées descriptives en MODS ou en Dublin Core, des métadonnées techniques en TEI Header,²⁷⁵ MIX,²⁷⁶ TEXTMD,²⁷⁷ VIDEOMD,²⁷⁸ PBCORE,²⁷⁹ ou AUDIOMID²⁸⁰ par exemple, selon le type de contenus. Enfin, les métadonnées administratives sont gérées, pour les droits d'accès, avec METS Rights schema²⁸¹, ou avec d'autres plus complexes comme ODRL²⁸² ou XrML.²⁸³ Enfin, les métadonnées d'historique²⁸⁴ sont gérées en général avec PREMIS,²⁸⁵

²⁷¹ MARC21 format for Authority Data, <http://www.loc.gov/marc/authority>

²⁷² MARC21 format for Bibliographic data, <http://www.loc.gov/marc/bibliographic>

²⁷³ Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets>

²⁷⁴ Digital Item Declaration Language6, qui constitue en fait la seconde partie du standard MPEG-21, <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>

²⁷⁵ Text Encoding Initiative, <http://www.tei-c.org/>.

²⁷⁶ Metadata for Images in XML, pour la description de images fixes, <http://www.loc.gov/standards/mix/>

²⁷⁷ TEXTMD est un schéma XML élaboré par la LOC pour la description d'objets textuels. <http://www.loc.gov/standards/textMD>.

²⁷⁸ Pour les métadonnées techniques de vidéos numériques, http://www.loc.gov/rr/mopic/avprot/DD_VMD.html

²⁷⁹ PBCore est un standard de métadonnées et de catalogage utilisé pour le broadcast, <http://www.pbcore.org/>

²⁸⁰ Audio Technical Metadata Extension Schema, <http://lcweb2.loc.gov/mets/Schemas/AMD.xsd>, pour les Fichiers son.

²⁸¹ Mets rights : <http://www.loc.gov/standards/mets/mets-extend.html>

²⁸² Open Digital Rights Language, <http://odrl.net/>

²⁸³ eXtensible Rights Markup Language, <http://www.xrml.org/>, XrML est issu du secteur commercial.

un standard constitué de quatre schémas XML (Objects, Events, Agents, Rights). Il s'agit d'un dictionnaire de données, permettant de les organiser.²⁸⁶ Une rapport de veille technologique rendu en avril 2008²⁸⁷ par le JISC sur « les métadonnées pour les bibliothèques numériques : état de l'art et directions futures »²⁸⁸ reprend ces différents éléments dans une perspective réaliste.

On l'aura compris, au travers de l'emploi de formats comme MARCXML ou MODS et de la mise en place de schémas comme METS en respectant des standards comme PREMIS, une des difficultés est de conserver des éléments de description à la fois précis et interopérables, répondant à la diversité des sources.

- La diffusion du Dublin Core

Pour simplifier les descriptions, l'OCLC et le NCSA²⁸⁹ ont élaboré en 1995 à Dublin, dans l'Ohio, un schéma de métadonnées qui est devenu l'un des principaux formats utilisés.²⁹⁰ Il s'agit bien sûr du Dublin Core, qui a donné lieu en 2003 à la norme ISO 15836, et dont la maintenance est assurée par le DCMI (Dublin Core Metadata Initiative) sous le contrôle de l'OCLC – Office of Research & Special Projects. Ce schéma de métadonnées est composé de 15 éléments, répartis en trois groupes : contenu, propriété intellectuelle, instance de la ressource.²⁹¹ Ces éléments sont facultatifs et répétables, pour rendre l'emploi aussi simple que possible. Il s'agit à la fois d'une simplicité de création et de gestion des notices descriptives dans un environnement en réseau, d'une sémantique communément comprise, et d'une extensibilité vers des Dublin Core déclinés, qui restent du Dublin Core, au travers de qualificatifs normalisés permettant de l'enrichir, définis par le DCMI, pour le raffinement d'éléments ou pour l'interprétation des valeurs dans les schémas d'encodage. Le Dublin Core peut être encapsulé dans des fichiers HTML ou XML, et est donc applicable à tous les formats de fichiers que nous avons présentés, à condition que la syntaxe soit interprétable par les moteurs et par les humains.²⁹² Il s'est donc développé parallèlement et séparément de l'évolution de MARC en XML. On cherche maintenant à l'intégrer aux outils traditionnels de bibliothèques, au travers du RDA.

Son envergure internationale lui a permis de dépasser largement le modèle des bibliothèques pour lesquelles il a été conçu. En autorisant à la fois les recherches par champs, l'indexation d'objets non textuels et l'accès à des contenus de substitution, il se prête en effet à des usages très larges.

C'est pour cette raison que, lors de la création d'entrepôts de notices communs aux archives ouvertes, aux musées et aux bibliothèques, le Dublin Core est apparu comme l'outil le plus simple et efficace pour échanger les informations. Il s'est donc imposé

²⁸⁴ Incluant les métadonnées de préservation (de PDI).

²⁸⁵ PREservation Metadata Implementation Strategies, <http://www.oclc.org/research/projects/pmwg/>, et <http://www.loc.gov/standards/premis/>

²⁸⁶ PREMIS est utilisé notamment par la BnF dans SPAR.

²⁸⁷ http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf

²⁸⁸ Metadata for digital libraries: state of the art and future directions

²⁸⁹ National Center for Supercomputing Applications, <http://www.ncsa.uiuc.edu/>

²⁹⁰ <http://dublincore.org>

²⁹¹ Contenu: Title, Subject, Description, Type, Source, Relation, Coverage. Propriété intellectuelle : Creator, Publisher, Contributor, Rights. Instanciation: Date, Format, Identifiant, Language.

²⁹² Voir LE MOAL, Jean-Claude, et al. *La recherche d'information sur les réseaux*. Paris: ADBS Éd., 2002, 322 p., p. 106-107

comme un moyen d'assurer l'interopérabilité des métadonnées, dans le cadre de programmes en OAI-PMH.

- L'Open Archive Initiative

L'OAI-PMH (Open Archive Initiative- Protocol for Metadata Harvesting) est un protocole http et xml permettant d'exposer des données en différents formats, qui soient récupérables par des « moissonneurs ». ²⁹³ Ce protocole est issu des serveurs de *pre-print* utilisés pour le partage d'articles scientifiques entre pairs pour les améliorer avant publication. ²⁹⁴ L'évolution vers des « archives ouvertes » plus larges au travers de publications en ligne donna lieu à une importante réflexion sur le signalement des ressources, avec en 1999 la conférence de Santa-Fé visant à mettre en place un protocole de fédération des contenus de ressources électroniques. ²⁹⁵ Le choix fut fait de ne pas conserver un protocole synchrone d'interrogation simultanée de bases de données (comme le Z 39-50), mais de prévoir la constitution d'entrepôts de données normalisées, les « DP » (Data Providers), qui contiennent les métadonnées répondant aux requêtes des « SP » (Service Providers, ou moissonneurs). Schématiquement, le lecteur (humain) utilise donc une interface permettant au SP d'adresser une requête aux DP, qui fournissent ensuite le lien vers les ressources décrites. Il s'agit donc d'un outil adapté aux recherches bibliographiques. ²⁹⁶

Outre l'économie de ressources liée à l'interrogation asynchrone, cette constitution d'entrepôts permet que les métadonnées initialement « cachées » soient exposées sur le Web, et visibles par les moteurs de recherche qui sont techniquement capables d'effectuer des requêtes prédéfinies par « verbes » pour identifier les entrepôts OAI et utiliser leurs notices. ²⁹⁷ En revanche, ils ne souhaitent pas forcément les utiliser. Si Yahoo ! et In-Extensio utilisent de façon systématique les notices en OAI, Google leur fait moins confiance et ne les utilise que dans le cadre de projets spécifiques.

Pour cette raison, l'utilisation de l'OAI-PMH permet une plus grande visibilité dans le cadre de numérisations de masse et de collections importantes que pour des bibliothèques de taille restreinte, puisqu'un nombre important de notices permettra d'intéresser les moteurs et d'obtenir un meilleur classement dans les résultats de recherche. ²⁹⁸ Ainsi OAI-NUM, entrepôt OAI mis en place par la BnF pour Gallica à partir de 2005 avec 79 000 notices environ, ²⁹⁹ sera mieux repéré que les portails spécialisés signalant les ressources en ligne d'un domaine particulier. On peut citer des

²⁹³ <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

²⁹⁴ Il s'agit du Peer-Review, développé notamment en sciences dures.

²⁹⁵ OAI Open Archive Initiative Journée d'information, Voir la présentation d'Emmanuelle Bermès, Journée d'information AFNOR CG46, 7 juin 2005, <http://www.bnf.fr/PAGES/infopro/journeespro/pdf/AFNOR2005/OAI.pdf>.

²⁹⁶ Voir NAWROCKI, François. Le protocole OAI et ses usages en bibliothèque, Paris, Ministère de la culture et de la communication, 2005. Mise à jour le 15 février 2005.

<http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>

²⁹⁷ Voir PEYRARD, Sébastien, *L'accès aux collections numériques...*, Op. Cit., p.15-16

²⁹⁸ *Ibid.*, p.16

²⁹⁹ La liste des notices en Dublin Core de l'entrepôt OAI-NUM est accessible à l'adresse http://oai.bnf.fr/oai2/OAIHandler?verb=ListRecords&metadataPrefix=oai_dc

bases comme TELMA³⁰⁰, l'entrepôt OAI-PMH du CRHST³⁰¹, et des catalogues de lecture publique exposés en OAI comme Lectura en Rhône-Alpes³⁰².

Au-delà de la spécialisation thématique, la partie de l'offre publique qui vise à s'intégrer dans une fédération des recherches doit donc prendre en compte le Dublin Core et l'OAI-PMH au travers des outils développés : outils techniques pour l'emploi de mots-clés avec des clusters, indicateurs, et vedettes autorité ; mais aussi critères de pertinence pour pondérer les algorithmes de recherche en plein texte (comme la valorisation du titre et de l'auteur dans Gallica 2), et intégration dans les recherches des moteurs déjà existants. Il est en particulier possible dans un entrepôt OAI de créer des « sets », ensembles qui peuvent être récupérés indépendamment du reste par un moissonneur spécialisé.³⁰³

Outre ces objectifs, les notices OAI peuvent être intégrées dans des fonctions de gestion, en particulier pour la création de fichiers de sauvegarde pérenne. Par exemple à la BnF, SPAR³⁰⁴ utilise la notice OAI des documents pour créer le fichier METS. Le système SOLON de la BnF qui gère les droits d'accès aux documents numérisés utilise lui aussi les données OAI pour la gestion des notices d'autorité.

De plus, certaines normes de métadonnées permettent l'exposition en OAI de descriptions qu'il serait sinon difficile d'exploiter. Notamment, le MPEG 21 est déjà utilisé en entrepôt OAI (en particulier pour la presse en Diddle MPEG 21)³⁰⁵ Ce qui peut répondre à l'empilement de différents niveaux pour certains objets, comme par exemple un coffret de disques d'intégrale ou de compilation (comportant les disques eux-mêmes) avec au moins trois niveaux de description. On peut alors utiliser une notice générale en Dublin Core, et Diddle, pour répondre à l'empilement des niveaux. Ces procédés sont déjà mis en place à la BnF, mais fonctionnent mal en MARC (avec des sous-notices « ANL » analytiques et des avec notices-mères). Il est donc essentiel d'utiliser des solutions XML pour pouvoir utiliser efficacement les différents niveaux de description.

- Rompre l'isolement des bibliothèques

Pour exposer les données de façon efficace dans le Web, il est nécessaire de traiter les éléments de notices comme des métadonnées à part entière. Les anciens modèles de *Search and Retrieve* avec les protocoles de recherche synchrone dans des bases de données fonctionnant différemment, comme le Z39-50, sont encore utilisés.³⁰⁶ Mais ils ne sont efficaces qu'en connaissant déjà les outils et l'objet recherché, par exemple pour un catalogueur sous WinIBW.

La recherche en temps réel pourrait sembler supérieure aux moissonnages à intervalles fixes du PMH, ne permettant que d'accéder de façon asynchrone à des états figés. Ce qui

³⁰⁰ Centre de ressources numériques de l'Ecole nationale des Chartes et de l'Institut de Recherche en Histoire des Textes, <http://www.cn-telma.fr>

³⁰¹ Centre de Recherche en Histoire des Sciences et Techniques, <http://www.crhst.cnrs.fr>

³⁰² <http://www.lectura.fr/fr/catalogues>

³⁰³ Par exemple dans Gallica3, un set « manuscrits », un set « cartes et plans », ...

³⁰⁴ Système de Préservation et d'Archivage Réparti, <http://bibnum.bnf.fr/spar/index.html>.

Voir Thomas Ledoux, Le Système de Préservation et d'Archivage Réparti de la BnF :

De la production à la pérennisation des objets numériques, présentation lors des 11e journées des pôles associés de la BnF, 27 - 28 mars 2008,

http://www.bnf.fr/pages/infopro/journeespro/pdf/poles_pdf/poles2008/spar.pdf

³⁰⁵ Diddle MPEG 21 est une norme internationale ISO/CEI développée par MPEG pour le multimédia numérique, et utilisée à la BnF en plus du Dublin Core pour l'exposition des données OAI. Voir « Les entrepôts OAI de la BnF », <http://bibnum.bnf.fr/oai/index.html>.

³⁰⁶ Z39-50 fait aussi partie des protocoles étudiée en TC46, comme d'autres protocoles, dont http.

pose une question d'usages, le protocole asynchrone n'étant problématique qu'en cas de différence notable entre l'état moissonné et celui de l'entrepôt d'origine.

Pourtant, l'évolution vers des métadonnées utilisables (notamment en OAI) est générale, avec par exemple l'élaboration de « dictionnaires de métadonnées » comme ILL, Z39-50, pour harmoniser les descriptions.

Ce « *metadata movement* » dépasse les formats de l'information bibliographique traditionnels, puisqu'une grande partie de l'information bibliographique sur le Web est produite hors du monde des bibliothèques. Le basculement vers un nouvel univers de recherche des ressources doit garantir une interopérabilité des descriptions pour éviter que les bibliothèques restent isolées, en leur permettant d'échanger les données avec d'autres communautés. Il s'agit de permettre à des applications extérieures d'utiliser l'information bibliographique produite par les bibliothèques, et inversement, d'utiliser des données extérieures pour signaler les ressources des bibliothèques.

En particulier, la place des nombreux acteurs tend à se redéfinir, en impliquant des difficultés nouvelles, puisque le degré de disponibilité des ressources sur le Web est comparable malgré une visibilité différente. Il serait donc inutile d'y rendre deux fois le même service, en revanche le signalement est crucial, y compris pour des approches diverses d'une ressource identique. De plus, jusqu'à présent, l'ensemble des personnes avaient leur place, avec des règles et formats normalisés dans les différentes professions (Archives, musées, bibliothèques). A l'inverse, on évolue vers un univers où les standards sont transverses entre diverses communautés qui se les ou approprient, comme dans les cas de l'EAD, du Dublin Core ou de la TEI XML,³⁰⁷ en archives et bibliothèques. De cette façon, les décisions séparées d'acteurs comme l'ALA, l'ISO et l'IFLA semblent révolues.

4.2 NOUVELLE ARCHITECTURE POUR LE CATALOGUE DE DEMAIN

La toute récente parution d'une version en ligne de RDA amène à s'interroger sur le passage à une nouvelle étape dans les modes de description des ressources des bibliothèques. De cette façon, on peut espérer qu'après la normalisation de structures autour du MARC jusqu'aux années 1990 puis l'évolution vers des technologies Web (XML) avec une conceptualisation des descriptions (FRBR), on franchisse une autre étape.³⁰⁸ Il s'agit à la fois de pouvoir développer, adapter et utiliser parallèlement les différents produits nécessaires, et de prévoir l'évolution possible vers de nouveaux modes d'organisation de l'information, en particulier avec la structure RDF³⁰⁹.

4.2.1 Une spécialisation nécessaire

Avec l'existence de nombreuses normes permettant de décrire les documents, le problème fondamental n'est pas d'en créer de nouvelles, mais de choisir lesquelles utiliser, en combinant plusieurs ensemble si besoin.³¹⁰

³⁰⁷ Text Encoding Initiative, <http://www.tei-c.org/>

³⁰⁸ Voir Figoblog, <http://www.figoblog.org/node/1945>, Cité en Annexe 3, « une hypothèse d'avenir ».

³⁰⁹ Voir *Infrq*, 4.2.2.

³¹⁰ Voir par exemple Figoblog, <http://www.figoblog.org/document278.php>.

- Une standardisation mondiale

Les formats utilisant le langage XML, emploient différents standards. Une des difficultés est de mettre en oeuvre efficacement un format autodocumenté, pour les machines et les humains. En effet, les descriptions invisibles en elles-mêmes pour les utilisateurs sont néanmoins essentielles pour la visibilité des ressources. En particulier, l'idée de s'intégrer aux outils et aux modes de recherche existants suppose de suivre le modèle de métadonnées du W3C, en se conformant aux préconisations effectuées hors du monde des bibliothèques avec la diffusion du XML.

L'interopérabilité des formats de métadonnées est donc sous-tendue par le respect d'outils généraux. Face à la « palette » de normes existantes, il n'est pas tant nécessaire de normaliser encore, que de ce qui correspond aux besoins. Comme l'expliquait notamment Andy Powell,³¹¹ repris par Lorcan Dempsey,³¹² on arrive à un besoin de complémentarité entre des schémas de description strictement bibliothéconomiques, et les outils habituels de description du Web, eux-même diversement exploités par les moteurs. De fait, XML est employé partout (contrairement à l'ISO 2709), avec une standardisation par le W3C. Mais il est évidemment nécessaire de s'adapter aux usages spécialisés. C'est ainsi que l'ISO 23083 préconise une spécialisation des métadonnées selon les secteurs, pour répondre à la diversité des besoins.³¹³

A titre d'exemple de format développé pour un contenu particulier de ressources, on peut citer le LOM (Learning object metadata), décrit depuis 2002 sur le site du W3C.³¹⁴ Ce schéma de métadonnées permet la description de ressources pédagogiques, notamment pour l'interopérabilité des bases de données afin de trouver et échanger les ressources. Le LOM n'est pas normalisé au niveau mondial, bien que la version « Lom.fr » soit normalisée AFNOR. Il est notamment utilisé par Formist.³¹⁵ Au travers de cet exemple, l'aspect important est la standardisation mondiale par le W3C, y compris dans des domaines spécialisés intéressant tout à fait les bibliothèques, par exemple sur les requêtes en XML, qui est déjà utilisé entre autres par Open Content Alliance³¹⁶ pour l'édition en ligne des ouvrages.³¹⁷

- Un besoin particulier de pérennité

Même en se conformant aux mouvements de standardisation mondiale pour garantir l'interopérabilité des données, diverses solutions sont élaborées. Il s'agit en particulier de sauvegarder durablement les données en garantissant la pérennité des métadonnées. Il est nécessaire d'identifier de façon unique les objets dans l'archive, car les liens doivent être pérennes. Ce qui permet notamment de n'établir qu'une description bibliographique minimale pour les sauvegardes de documents continuant à être repérables à partir du

³¹¹ Voir Andy Powell: "Does Metadata Matter", diaporama sur <http://www.slideshare.net/eduservfoundation/does-metadata-matter?src=embed>

³¹² http://orweblog.oclc.org/archives/cat_metadata.html

³¹³ CLAVEL, Gilles. *XML, la synthèse : intégrez XML dans vos architectures*. Paris: Dunod, 2002, p.124.

³¹⁴ <http://Itsc.ieee.org/wg12>

³¹⁵ Réseau de FORMation à l'Information Scientifique et Technique, <http://formist.enssib.fr/>, mais pas sous la forme Lom.fr, car elle n'était pas encore parue au moment de la mise en place ; Voir http://formist.enssib.fr/documents/LOM_et_acces_FORMIST_document-n-4121-r-18-t-typdoc.html.

³¹⁶ Le consortium OCA rassemble des associations dont avant tout Internet Archive, le Research Library Group, et des acteurs publics et privés prestigieux dont Yahoo et MSN.

<http://www.opencontentalliance.org>

³¹⁷ Notamment pour les langages de requêtes par le XML, comme XQuery, voir compte-rendu sur Telma, <http://www.cn-telma.fr/veille/recommandations-XSLT-xquery>

catalogue habituel. L'archivage de données numériques suppose donc un autre mode de description des ressources, adapté aux diverses étapes successives (Versement, stockage en « silo », et gestion des droits et accès adaptée à l'évolution des usages). Des solutions techniques sont actuellement recherchées.³¹⁸ En outre, un travail de normalisation des formats de stockage est mené au niveau mondial, avec l'action du W3C, de l'IETF³¹⁹ et de l'ISO. On peut mentionner par exemple le groupe WARC de la TC 46 (TC 46 C4) sur les formats de stockage des données Web.

Avec la mise en place d'une standardisation utilisée hors du monde des bibliothèques, se pose la question de l'utilisation des formats de l'information bibliographique par les professions voisines. En effet, la nécessité de passer du MARC à des schémas de métadonnées utilisant XML ne pose pas seulement des problèmes de choix des formats, de coût de conversion des millions de notices, et d'adaptation des professionnels (catalogueurs). Il s'agit aussi de savoir quel modèle de création de l'information bibliographique est mis en place. En effet, l'interopérabilité avec les descriptions utilisées par les organismes producteurs et diffuseurs des ressources pourrait impliquer un changement important des habitudes de travail.

- Un double travail ?

L'actuel travail de catalogage en bibliothèque consiste essentiellement, outre la dérivation, l'exemplarisation et le bulletinage, à recopier en les ordonnant de façon structurée des informations prélevées sur la page de titre des publications, et autres sources prescrites.

Par ailleurs, l'actuel travail de publication, y compris pour des ressources papier, implique l'élaboration de fichiers informatiques, avec éventuellement la production de métadonnées exprimées en XML, utilisant par exemple la DTD ONIX, peu répandue en France mais employée notamment aux Etats-Unis, et utilisée pour les catalogues de diffusion commerciale des ouvrages.

Cette double création de l'information bibliographique semble redondante, du fait de l'interopérabilité possible entre les données. Il serait donc tentant d'imaginer par exemple que les éditeurs déposent leurs publications au format électronique (au régime du dépôt légal),³²⁰ et permettent d'utiliser directement les données de la page de titre, voire de reprendre les métadonnées déjà établies. Ce modèle de production de l'information bibliographique supposerait que les anciennes agences bibliographiques nationales centralisent différemment l'information : dans le cadre des activités d'édition, l'éditeur déposerait les ouvrages en fournissant des métadonnées en XML, que les catalogueurs transformeraient pour obtenir une notice bibliographique intégrant les divers types de métadonnées, et associée aux divers renvois d'autorités. Ces données seraient ensuite fournies aux autres bibliothèques. D'un point de vue technique, la Library Of Congress propose déjà des passerelles entre l'ONIX et MARC-XML. Pourtant, dans le monde des bibliothèques comme dans celui de l'édition, l'emploi de données XML n'est pas encore suffisamment répandu en France pour que ces échanges soient simples.

³¹⁸ Comme SPAR à la BnF, voir Thomas Ledoux, Le Système de Préservation et d'Archivage Réparti, http://www.bnf.fr/pages/infopro/journeespro/pdf/poles_pdf/poles2008/spar.pdf

³¹⁹ L'*Internet Engineering Task Force* est une organisation interprofessionnelle des métiers de l'informatique et des réseaux, établissant des standards, <http://www.ietf.org/>.

³²⁰ Ce dépôt électronique existe pour le moment en France sous la forme limitée de l'« exception handicap », article 122-5 alinéa 7 du code de la propriété intellectuelle, ajouté par la « loi DADVSI ». Chapitre 1, article 1, alinéa 3, <http://www.legifrance.gouv.fr/>

La simplification des procédures de description semble pourtant souhaitable, comme le souligne le rapport PATINO,³²¹ qui formule des recommandations pour l'association de la BnF, du Centre National du Livre³²² et de Dilicom,³²³ afin de mettre en commun les données bibliographiques. Dilicom maintient en effet un fichier d'ouvrages disponibles pour la librairie, tandis que le Cercle de Librairie publie *Livres hebdo* avec le signalement des nouveautés parues, et que la BnF élabore la Bibliographie Nationale (intégrant aussi les documents audiovisuels, et les cartes et plans). Ces trois organisations fournissent donc un travail semblable, en élaborant des descriptions bibliographiques. Ce travail est cependant effectué à des niveaux variables. Ainsi, le emploi des données de Dilicom pose un problème informatique, car la description est proche d'un pré-catalogage très général. Cependant d'autres données seraient essentielles, comme le fichier des éditeurs du SNE,³²⁴ utile pour connaître les nouveaux éditeurs, qui est réclamé pour le Dépôt Légal. Réciproquement, le service du Dépôt légal de la BnF recense plus de publications que le Cercle de la librairie, car on y trouve des ouvrages issus de l'auto-édition, et de dépôt volontaire avec des tirages limités.³²⁵ Une production autonome devrait donc rester possible malgré le partage des données. La fusion des données poserait aussi un important problème économique, puisqu'un reformatage technique et organisationnel serait nécessaire pour que les différentes entités puissent diffuser les données.

La réflexion sur l'avenir des catalogues, en s'intéressant à la refonte des données, doit prendre en compte le contexte de production, avec l'éventuelle intégration des éditeurs dans la production de l'information bibliographique. Il pourrait s'agir de données alphanumériques, voire d'images (couverture ou 4^e de couverture). On pourrait imaginer par exemple une bibliographie de France illustrée qui serait peu complexe à concevoir, mais ne correspond pas aux possibilités actuelles des métiers de l'édition, dont les situations sont variables, et éventuellement menacées. Surtout, ces projets se limitent aux publications papier. Ils n'excluent donc pas de continuer à réfléchir aussi sur les modes de description des autres ressources.

- S'adapter aux supports

Sur les supports autres que l'écrit imprimé, les éléments de description sont très liés aux modes de production, du moins pour les objets numériques natifs.

Qu'il s'agisse par exemple de métadonnées décrivant des images photographiques ou des films de *broadcast*,³²⁶ les éléments de données sont prévus par des corpus de normes fermés, avec la prise en charge automatique d'une partie des descriptions. Il s'agit souvent d'applications fermées, utilisant des standards où chaque profession est un

³²¹ Rapport sur le livre numérique remis à Christine Albanel le 30 juin dernier, mettant l'accent sur les changements d'usages et les probables évolutions du commerce du livre. PATINO, Bruno, rapport remis à madame Christine Albanel, ministre de la culture, sur le livre numérique, 30 juin 2008 :

<http://www.culture.gouv.fr/culture/actualites/conferen/albanel/rapportpatino.pdf>

³²² <http://www.centrenationaldulivre.fr/>

³²³ Dilicom (anciennement Edilectre), <http://www.dilicom.net/>, société interprofessionnelle développant les échanges de données informatiques entre les éditeurs, les distributeurs et les libraires.

³²⁴ Syndicat National de l'Édition, <http://www.sne.fr/>

³²⁵ Ainsi, sur environ 60000 titres de livres répertoriés par an en France, environ 50000 sont actuellement communs entre la BnF, Dilicom et le CNL.

³²⁶ Voir GOUYET, Jean-Noël; GERVAIS, Jean-François. *Gestion des médias numériques : digital media asset management*. Paris : Dunod ; 2006, p. 50 à 121.

« marché » trop petit pour une norme spécifique, d'où l'emploi de structures d'application communes. Celles qui sont concernées par l'exposition sur le Web peuvent être encadrées par l'ISO avec la TC46, mais ce travail lui échappe souvent, avec notamment l'action du W3C et de l'IETF qui établissent des standards prenant souvent une valeur de norme en s'intégrant dans le cycle de production des matériels et, indirectement, des métadonnées associées aux contenus produits. Cette question s'étend, pour les documents textuels nés-numériques, au codage des caractères. Et à l'inverse, pour les documents manuscrits, l'utilisation de DTD spécifiques comme l'EAD suppose un type de traitement particulier, ou l'intervention du catalogueur est essentielle.³²⁷ Pour répondre à cette diversité des types de documents, on admet généralement qu'il reste nécessaire de dissocier les ressources des bibliothèques avec, de façon schématique, trois grands ensembles correspondant à trois formats :

- le Marc XML pour la notice bibliographique, éventuellement issue d'une conversion rétrospective,
- la DTD Dublin Core pour les photographies et documents iconographiques,
- la DTD EAD pour les notices de manuscrits, et de fonds d'archives.

Dans le cadre du Web, les 3 formats se complètent pour la mise en ligne et la structuration de l'information.

Mais comme le fait remarquer Diane Hillmann dans ses articles *Facing Forward: The Challenges Facing Cataloging and Catalogers*³²⁸ et *The Future of Cataloging and Catalogers*,³²⁹ on peut envisager que les futurs instruments de recherche en ligne soient bien plus qu'une mise en application du langage XML et des outils orientés sur les utilisateurs. Au travers du développement du Web sémantique, il est possible que les modalités de description du contenu des documents changent radicalement.

4.2.2 Les possibilités du RDF

Le RDF³³⁰ est un des chantiers majeurs du W3C, lancé en 1997 dans le cadre de projets de gestion des métadonnées pour permettre une « sémantique des métadonnées » lisible en machine, interopérable et utilisable pour des informations de nature hétérogène.³³¹

- Structurer l'information en triplets

En adoptant une structure sémantique calquée sur le langage humain, le RDF adopte pour toute description une structure en triplet, « Sujet/predicat/objet ».³³² Ce qui permet d'établir des assertions associant les ressources et les autorités, pour établir des graphes descriptifs, formant des nœuds entre eux.³³³ En effet, les arcs³³⁴ des graphes peuvent être étiquetés : on obtient donc des énoncés élémentaires avec un triplet (propriété, objet,

³²⁷ Voir L'accès en ligne aux manuscrits : Possibilités de XML, systèmes de navigation, référencement, Présentation par Raphaële Mouren et Isabelle Westeel,

http://www.bm-orleans.fr/userfiles/file/portail/manuscrit_4.pdf

³²⁸ <http://hdl.handle.net/1813/11536>

³²⁹ <http://www.slideshare.net/kramsey/the-future-of-cataloging-and-catalogers-presentation?type=powerpoint>

³³⁰ *Ressource Description Framework*, Cadre de description des ressources, <http://www.w3.org/RDF/>

³³¹ Voir LE MOAL, Jean-Claude, et al. *Bibliothèques numériques*. Paris: ADBS Ed., 2000, 246 p., p. 159.

³³² Subject, Predicate, Object. Voir la « fiche pratique » réalisée par Sébastien Peyrard, <http://www.enssib.fr/bibliotheque-numerique/document-1836>

³³³ Voir <http://www.lespetitescases.net/>, en particulier <http://www.lespetitescases.net/rdf-pour-les-nuls>

³³⁴ Les arcs sont des prédicats reliant une ressource=un sujet à une valeur=un objet.

valeur), où la « propriété » constitue l'étiquette de l'arc, où le nœud origine de l'arc est l'objet, avec une référence directe (ressource sur le Web), et où la « valeur » est le nœud destination, qui peut être un autre nœud. Ces triplets peuvent être attachés à d'autres nœuds (avec des sous-ensembles, des exemples et des équivalences) et « réifiés » avec l'intégration de propriétés et de valeurs à un autre nœud.³³⁵

Sans entrer dans le détail de ce modèle, on peut observer qu'il s'agit de schémas élaborés pour les ressources en ligne, renvoyant à des URL pérennes et non à des formes papier comme source première. Il s'agit d'évolutions peu réalistes à très court terme dans le monde des bibliothèques, qui se soucie pour le moment d'adapter XML aux MARC. Pourtant, le modèle RDF n'exclut absolument pas d'utiliser le Dublin Core, l'EAD, ou tout autre modèle de métadonnées. Il clôt ainsi le débat sur la distinction entre données et métadonnées, puisque la « réification » permet de rattacher les contenus des ressources aux descriptions. Il s'agit vraiment d'un modèle de description des ressources, compatible avec les divers formats XML existants.³³⁶ C'est donc une des applications les plus ambitieuses de XML, cherchant à décrire le contenu des documents dans un contexte de développement possible du « Web sémantique ». Au-delà des indexations déjà existantes, l'objectif est de s'en servir pour des requêtes d'exploitation du contenu sémantique³³⁷, avec SPARQL³³⁸ qui est le langage conçu pour interroger les entrepôts RDF. Ce protocole de requêtes est normalisé par le W3C³³⁹ (contrairement au SQL pour les bases de données relationnelles). Avec SPARQL, il n'est pas nécessaire de connaître la structure de l'entrepôt pour pouvoir l'interroger.

Encore jeune, le RDF est en évolution. Surtout, il est délicat de savoir dans quelle mesure son utilisation serait pertinente en bibliothèques,³⁴⁰ puisque sa mise en place nécessiterait une implication forte de l'établissement, et une certitude de la pertinence de tels choix par rapport aux développements généraux du Web.

- Un travail hors des bibliothèques

Les difficultés posées par les catalogues sont souvent liées aux moyens technologiques, qui supportent mal les volumes traités.³⁴¹ L'évolution vers de nouvelles technologies, avec de nouveaux systèmes, pose donc un problème de choix, d'adaptation et de moyens. Or, les moyens déployés par les bibliothèques peuvent difficilement atteindre la puissance déployée par les groupes de travail du W3C sur les évolutions informatiques fondamentales. Ainsi, un groupe réunissant plusieurs acteurs principaux³⁴² a défini le DOM,³⁴³ une application RDF permettant de gérer les documents et d'y naviguer

³³⁵ PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique... *Op. Cit.*, p.204. Mais la réification peut poser des problèmes d'exploitation des données, et est parfois considérée comme un défaut.

³³⁶ Voir LE MOAL, Jean-Cl., et al. *La recherche d'information sur les réseaux*. Paris: ADBS, 2002, p. 221-222.

³³⁷ PRAX, Jean-Yves; LARCHER, Simon. La gestion..., *Op. Cit.*, p.204.

³³⁸ Norme disponible à l'adresse <http://www.w3.org/TR/rdf-sparql-query>

³³⁹ <http://www.w3.org/2007/12/sparql-release.html> et <http://www.lespetitescases.net/sparql-maillon-essentiel-web-of-data> pour une introduction.

³⁴⁰ Son utilisation sur des ressources comme celles de la BnF pourrait permettre de donner des réponses avec une structure souple. L'actuelle réflexion sur la refonte du catalogue implique en effet un groupe de travail sur les possibilités de ce type d'outils.

³⁴¹ Bien que le RDF soit âgé de dix ans, les technologies qui le mettent en œuvre sont récentes, et on est pas encore totalement sûr de sa fiabilité (voir le Billion Triples Challenge en 2008).

³⁴² Notamment Microsoft, IBM, NetScape, Sun, Oracle, Novell et Softquad.

³⁴³ Document Object Model, recommandation du W3C pour des interfaces permettant à des programmes d'accéder aux contenus. <http://www.w3.org/DOM/>

indépendamment des plateformes et des langages de programmation.³⁴⁴ Ces évolutions, doivent être prises en compte pour que les instruments de description des ressources (physiques ou numériques) restent pertinents.

- Le « Web sémantique » : un grand catalogue ?

Au travers de l'intégration des éléments de description dans les documents eux-mêmes, on peut espérer la mise en place de modes de recherche utilisant la structure sémantique des documents.³⁴⁵ On pourrait envisager une évolution du catalogue liée aux documents numériques. Mais pour le moment, les éléments de données permettant l'indexation des documents numériques viennent des catalogues traditionnels, et il est impossible pour les moteurs de recherche d'indexer des catalogues sous la forme de bases de données.

De cette façon, le fossé risque de continuer à se creuser entre les documents directement disponibles en ligne en mode texte, et ceux dont la bibliothèque dispose sous une forme matérielle.³⁴⁶ On peut distinguer les documents numériques libres de droits, les documents numériques sous droits, et les documents physiques libres de droits ou non. Les documents numériques peuvent être visibles parallèlement en mode texte pour les recherches. Google Books propose même un accès à des extraits et à une notice bibliographique d'ouvrages sous droits, pour un mode de recherche de masse à partir des contenus.

Mais dans le cas des descriptions des ressources proposées par les bibliothèques, la visibilité sur le Web suppose l'exposition de données descriptives. Face aux évolutions technologiques, deux moyens principaux peuvent être retenus. D'une part, la constitution d'entrepôts OAI, avec des notices bibliographiques sur des documents numériques ou non. D'autre part, la création de pages html générées à partir du catalogue, et offertes aux moteurs, avec des présentations de l'œuvre et de ses manifestations pouvant reprendre les nouveaux modèles théoriques de description.

Sans se risquer à des prédictions, on peut observer que la mise en place de schémas standardisés de description sémantique favorise la mise en valeur des ressources en mode texte, au détriment des diverses autres ressources, notamment disponibles en mode image. Ces évolutions sémantiques amènent à s'interroger sur le fonctionnement des bibliothèques numériques, dont les collections risquent d'apparaître d'une façon dissociée des institutions.³⁴⁷

³⁴⁴ PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique..., *Op. Cit.*, p.205 : « DOM définit une API (Application Programming Interface) spécifiée en utilisant le langage IDL (Interface Definition Language) de l'OMG (Object Management Group), lui-même défini dans la norme CORBA Common Object Request Broker) de l'OMG. Le langage IDL a été choisi précisément pour rester indépendant d'un langage de programmation, mais pour faciliter la mise en œuvre, le standard DOM contient en annexe des définitions équivalentes d'interfaces en Java (de Sun) et ECMAScript, langage de spécification de script de l'ECMA (European Computer Manufacturers Association), lui-même basé sur JavaScript (de Netscape) et Jscript (de microsoft). »

³⁴⁵ Voir <http://www.lespetitescases.net/>, en particulier <http://www.lespetitescases.net/a-la-decouverte-du-web-semantique-et-de-ses-technologies>.

³⁴⁶ L'AULA (utilisateurs d'Archimed) a organisé une journée sur le Web sémantique le 27 juin 2008, <http://listes.adbs.fr/sympa/arc/adbs-info/2008-05/msg00124.html>

³⁴⁷ Gallica, reflet de la BnF, intègre des pôles associés (BIUM, CNAM); avec Internum (<http://data.internum.org>), le Centre de Conservation du Livre à Arles, numérise des fonds de nombreuses institutions en les mettant en ligne sur une bibliothèque numérique.

5. Catalogue et bibliothèque numérique

Malgré l'évolution des modes de description dans le contexte du Web, les ressources des bibliothèques continuent à devoir être décrites. Il faut donc que le catalogue comporte les éléments nécessaires à la gestion des documents, de tous types.

5.1 UN OUTIL POUR TOUS LES SUPPORTS

Le catalogue est un fichier de métadonnées qui décrit tous les documents de la bibliothèque, par unités, que ceux-ci soient analogiques (papier, VHS) ou numériques (indépendamment du support).

5.1.1 Un fichier de métadonnées : identifier et localiser

Le catalogue reste un élément essentiel de la bibliothèque numérique en constituant la source documentaire qui sert à construire le système d'accès au document selon des évolutions constantes.

- Des changements communs.

Les bibliothèques numériques impliquent un fort lien entre les métadonnées et les objets auxquels elles sont associées, qui sont de même nature, avec un catalogue d'informations secondaires qui donne accès à un document numérique. Ce lien entre les métadonnées et les objets influence la façon d'aborder les catalogues, y compris pour la description des ressources papier. Le document numérisé doit être trouvé quand on le cherche sous forme électronique, mais il devrait tout autant être signalé quand on le cherche sous la forme matérielle, (sans savoir s'il est en ligne). La réflexion actuelle sur la refonte des catalogues, par exemple à la BnF avec l'intégration des ouvrages numérisés en masse, implique donc une rupture avec ce qui a été fait auparavant.

En effet, le catalogue répondait jusqu'à présent à la localisation d'objets physiques, recherchés par des humains dans des magasins matériels. Or, dans l'environnement numérique, ce ne sont pas des humains qui utilisent les éléments de localisation. Le type de localisation change donc, en entraînant une modification importante : là où une partie des informations pouvait être implicite, tout doit désormais être explicite.

La première informatisation, en place jusqu'à maintenant, reproduisait le travail du catalogueur manuel, avec une structure finalement comparable à celle de fiches. Or les structures changent, avec de nouveaux modes de production et de nouveaux objectifs. Dans le cadre de la numérisation, les métadonnées sont saisies à la source, avec d'éventuelles extractions automatiques. On retrouve donc des éléments de données déjà existants, avec des métadonnées descriptives issues des catalogues, des métadonnées techniques correspondant à l'en-tête des fichiers, et des métadonnées de

structure utilisant des fichiers de structuration de production.³⁴⁸ De cette façon, le catalogage ne limite plus à l'action de spécialistes *a posteriori*, mais s'intègre dans l'activité de production, de prestataires ou de l'atelier de numérisation.

De cette façon, les métadonnées vont au-delà d'une description rétrospective des ressources, pour impliquer les structures et leur utilisation. C'est une des différences entre la « bibliothèque numérique » et la « bibliothèque numérisée » : la présence en ligne d'une institution à part entière implique de garantir des ressources authentiques et fiables, avec à la fois des éléments structurels permettant la gestion des données du document, et des éléments de conservation et de critique des sources, retraçant la relation entre le document et l'activité qui le produit au travers d'identifiants (date, personnes impliquées, liens entre documents).³⁴⁹

- Un autre univers sémantique.

Avec l'éventuelle adaptation à un autre univers sémantique, la reprise des données suppose une nouvelle organisation des éléments, selon leur pertinence, plus ou moins nombreux qu'auparavant, la structure sémantique amenant à sophistiquer la structure et les éléments de données, par exemple pour la gestion des droits. Plus encore, le traitement catalographique, en s'adaptant aux documents numériques, fait évoluer les catalogues de documents physiques. De cette façon, les modes de navigation qui changent la façon de percevoir l'ouvrage.³⁵⁰

En particulier, la possibilité d'effectuer des recherches en plein texte conduit à considérer les ouvrages au même titre que des pages Web, comme des publications informatives dont on ne cherche pas à connaître l'intégralité, et que l'on n'aborde pas forcément comme une œuvre cohérente. Une des difficultés est donc de faire apparaître à un niveau comparable les ressources papier de la bibliothèque, et les ressources numériques qui en sont l'extension.

Surtout, il est délicat de combiner les modes de recherche en plein texte avec les éléments de description traditionnels. De cette façon, le fonctionnement de Gallica 1 impliquait un index séparé matériellement sur le serveur, contrairement à son évolution actuelle où tout passe par le catalogue, avec une recherche dans le plein texte des documents et les notices des documents de Gallica 2 combinée avec des critères de pondération selon la nature des termes trouvés (avec une pondération plus forte si le terme ramené est en champ titre ou auteur).

En ce sens, le catalogue s'adapte au catalogue numérique, en alimentant le moteur de recherche externe. La possibilité de rechercher dans le texte des ouvrages numérisés peut être intéressante même pour trouver des ouvrages papier. C'est la logique extrême de Google Books, avec la possibilité de rechercher dans le texte d'ouvrages sous droits dont on ne voit ensuite en ligne que des extraits, avec a nouveau la question de la granularité de l'information.

³⁴⁸ Comme refNum à la BnF.

³⁴⁹ Voir Les métadonnées du Records Management, Intervention de Geneviève Drouhet, CG46- 7 juin 2005, <http://www.bnf.fr/pages/infopro/journeespro/pdf/AFNOR2005/MetadonneesRecMan.pdf>.

³⁵⁰ LUPOVICI, Catherine, « Du catalogue informatisé à la navigation dans l'information numérique », In Revue de la Bibliothèque nationale de France, N°9 : Le catalogue, 2002, p. 80-83.

- Des choix importants

Le choix de nouveaux modes de description des ressources, même en reprenant des éléments de données adaptés, suppose la formulation d'hypothèses sur les évolutions plus générales des techniques de traitement de l'information, puisque les établissements et leurs systèmes d'information s'intègrent dans des problématiques plus générales. Quel que soit l'établissement, il s'agit donc de concilier l'activité du comité de pilotage spécialisé ou des groupes travaillant sur le sujet du catalogue, les pistes retenues par la direction de la bibliothèque, et les évolutions et normalisations plus générales que l'établissement va prendre en compte, ou influencer par ses choix dans le cas de grandes bibliothèques comme la BnF réfléchissant à l'évolution de son catalogue dans les années à venir.

De cette façon, l'évolution d'un système informatisé complexe suppose une refonte par phases sur plusieurs années. Il convient de composer avec l'existant ; l'OCLC établit des standards de fait au travers de ses décisions. Notamment, l'évolution du MARC vers des bases de données en XML correspond à des solutions techniques choisies aux Etats-Unis, et exportées pour le reste du monde. De même, l'absorption de RLG (archives et musées) par l'OCLC (bibliothèques) depuis 2006 amène à décrire des ressources diverses.

5.1.2 Les supports multiples des médiathèques virtuelles

L'évolution vers des supports non imprimés correspondait à un élargissement de service des bibliothèques physiques qui se satisfaisait des outils de description des publications. En revanche, face aux habitudes de recherche en ligne, il devient très délicat de refléter la diversité des supports en permettant un accès pertinent et efficace.

- Analogiques et numériques, multiples et uniques.

Si la complémentarité de ressources papier et numériques est communément acceptée,³⁵¹ il est plus délicat de voir comment la bibliothèque peut aussi être « hybride » en ce qui concerne le type de ressources numériques proposées. Ainsi, on observe peu à peu une évolution de l'accès aux contenus, notamment dans le contexte universitaire, où une porosité croissante entre le bureau virtuel et la documentation électronique en général (thèses, périodiques électroniques et articles en ligne, bases de données) entraîne la mise en place d'une bibliothèque numérique dont les documents numérisés ne sont qu'un élément. Plus encore, avec la mise en place d'actions de numérisation sur des documents anciens, rares ou précieux, la diversité des supports rejoint celle des types de documents, en les rendant paradoxalement accessibles sur un seul écran.

De fait, les livres anciens et manuscrits sont par définition des objets uniques, du fait de leur reliure, des commentaires et marques de possession, de leur état et de leur histoire éventuelle. Même dans des programmes de numérisation de masse, tout en restant des

³⁵¹ LAHARY, Dominique «L'accès aux catalogues et ressources des bibliothèques à l'âge des bibliothèques numériques et des moteurs de recherche : écarts, perturbation, mutation ? » In Les bibliothèques numériques, Dir. Fabrice Papy, Hermes-Lavoisier, 2005.

livres, ils sont également des objets uniques, pour lesquels le catalogue est à la fois un moyen de retrouver un ouvrage, et un instrument de connaissance des fonds.

Or, la mise en ligne en fait des objets accessibles de partout, ce qui n'en amoindrit pas la valeur mais change la perspective de description. On peut citer par exemple le catalogue en ligne Calames³⁵² (abondamment commenté sur le site de Lorcan Dempsey³⁵³), où des documents manuscrits uniques deviennent accessibles de partout en plein texte.

- Généraux et spécialisés.

Si les catalogues commerciaux distinguent toujours les différents types de supports,³⁵⁴ les bibliothèques sont habituées à la logique du mélange des ressources. La consultation des objets numériques pose donc un problème d'intégration dans les autres ressources de la bibliothèque. Dans le cadre de collections strictement numériques, la gestion peut être comparable à celle de collections en libre accès : la première maquette d'Europeana prévoyait ainsi une reprise de la Dewey, avec un découpage thématique permettant de naviguer dans l'arborescence.³⁵⁵ Des modes de recherches comparables peuvent être mises en place pour des bibliothèques hybrides, avec des études sur l'ergonomie des interfaces, comme à Paris VIII, avec le « Visual catalog ».³⁵⁶ Plusieurs sites de B.U. proposent à la fois des ressources électroniques natives et des documents numérisés.³⁵⁷ Il s'agit en ce cas d'un portail unifié utilisant un méta moteur pour interroger les diverses ressources (catalogue, ressources électroniques, documents numérisés). L'accueil prévu pour le lecteur ne dissocie pas la bibliothèque numérique des autres ressources, avec éventuellement une intégration complète à l'Environnement Numérique de Travail présentant une interface d'accueil commune, en tout cas dans le contexte universitaire³⁵⁸.

Des outils comparables peuvent exister dans des bibliothèques de lecture publique. Par exemple, le « Catalog+ » de la Bibliothèque Municipale de Lyon³⁵⁹ permet d'interroger à la fois le catalogue, les réponses du « Guichet du savoir », et les bases de données. Outre les fonctionnalités de vérification d'orthographe et de recherche sur des déclinaisons du terme (lemmatisation), on trouve une présentation graphique arborescente de suggestions de sous-sujets, permettant une navigation intuitive qui complète la logique de recherche par mots-clés.

- Quelle place pour l'audiovisuel ?

Après la diversification des supports disponibles en bibliothèque, la numérisation revient paradoxalement à rendre accessibles des contenus de diverses natures sous une seule forme, électronique. Il faut que le mode d'accès, uniformisé au travers d'un instrument de recherche unique, permette de faire face à la multiplication des contenus.

³⁵² <http://calames.abes.fr>

³⁵³ DEMPSEY, Lorcan, Lorcan Dempsey's weblog : On libraries, services and networks, <http://orweblog.oclc.org>

³⁵⁴ Par exemple sur <http://www.fnac.com>, il n'est pas possible d'effectuer une recherche multi-supports.

³⁵⁵ Maquette d'Europeana élaborée par la BnF :

http://maquette.bnf.fr/labs/scenario/Europeana_demo.html

³⁵⁶ Voir PAPY, Fabrice, « Au-delà de la transfiguration du catalogue : Le « visual... catalog », *BBF*, 2005, n° 4, p. 5-12 <http://bbf.enssib.fr>.

³⁵⁷ Comme Jubil pour Paris 6 (Jussieu) avec la Jubilothèque, <http://jubil.upmc.fr>, ou le site de Rennes 1, <http://www.univ-rennes1.fr/themes/cultureEtDocumentation>

³⁵⁸ par exemple à Rennes 1, avec des ouvrages sous droits, <http://www.scd.univ-rennes1/ebook/>, dans le programme de l'Université Numérique en Région Bretagne, <http://unrb.univ-bretagne.fr>

³⁵⁹ <http://autonomy.bm-lyon.fr/retina/public/login.do?username=guest&password=guest>

On peut ainsi imaginer une évolution vers des « médiathèques numériques », selon le modèle d'Europeana dans sa version Européenne mise en ligne en Novembre 2008,³⁶⁰ permettant une recherche fédérée d'images, textes, vidéos et sons, avec un tri simple selon les types de documents.

La difficulté majeure dans ce cas est de se positionner sur l'accès aux contenus sous droits, lorsqu'ils ne sont pas empruntables sur un support matériel. Par exemple, si la Médiathèque de l'agglomération Troyenne fait le choix de limiter l'accès aux contenus audiovisuels numériques à une consultation sur place, la BPI réfléchit actuellement à des modalités de prêts de fichiers, par exemple pour des baladeurs MP3, qui supposeraient l'élaboration d'instruments de recherche pertinents. En effet, le mode de recherche habituel des usagers pour ce type de ressources est le « butinage » dans les bacs de boîtiers de disques : on pourrait imaginer de le transposer directement avec un téléchargement après le choix du boîtier, mais pas si ce service existe déjà en ligne.

Dans ce domaine, les bibliothèques souffrent d'une concurrence telle qu'il est peut-être inutile de développer des services de ce type. Non seulement les mécanismes de *Peer-to-peer*³⁶¹ permettent un accès immédiat et gratuit à presque toute la production récente, mais même des sites légaux proposent en fait des contenus sous droits selon un mode avec lequel il est difficile de rivaliser. Ainsi Youtube,³⁶² racheté par Google en 2006, mène pour une part une politique de rétribution *a posteriori* des auteurs, et pour une part de dédommagements dans un contexte judiciaire, jugé plus rentable que le tri, qui nécessiterait des moyens permanents et importants, et entraînerait un appauvrissement des ressources en ligne. Chaque année, 200 millions de dollars sont donc mis de côté pour des procès éventuels.³⁶³ Ces pratiques correspondant à la législation américaine de l'« *Opt-out* » sont illégales et actuellement impensables en France, ce qui rend impossible de les utiliser en bibliothèque.

5.2 GERER LES DOCUMENTS

La prise en compte commune de documents matériels localisés et de documents numériques complètement pris en charge, entraîne une modification de l'approche de la gestion des documents.

5.2.1 Classer l'immatériel

De fait, envisager une collection repérable seulement par des langages de machines suppose des modes de description adaptés.

- Tout doit être dit.

De nombreuses informations implicites dans le domaine physique, deviennent inexistantes dans le domaine numérique. Par conséquent, l'accès normal au document numérique suppose l'ajout d'informations de gestion, comme l'accès restreint, ou le retrait de la vente. Ainsi, les notes manuscrites ou pastilles des documents physiques³⁶⁴ correspondent à des modes de gestion qui doivent trouver un équivalent dans

³⁶⁰ <http://www.europeana.eu>

³⁶¹ Avec des clients de partage de fichiers peer-to-peer comme Emule, BitTorrent ou Azureus.

³⁶² Youtube, <http://fr.youtube.com/>

³⁶³ GERVAIS, Jean-François. *Web 2.0 : les internautes au pouvoir ...*, Op. Cit.

³⁶⁴ Comme à la suite du rapport Otto, n'ayant pas toujours entraîné la destruction des ouvrages.

l'environnement numérique, de façon comparable aux fonctionnalités de la GED, avec en particulier la rétroaction. Par exemple, dans le cas de livres interdits, contrôlés ou sous droits, il est possible de mettre en place des processus de levée ou d'aménagement des restrictions. Ces évolutions font partie des perspectives de développement des bibliothèques.

De même, alors que le désherbage est un des points principaux de gestion des collections matérielles, on ne l'évoque jamais pour les ouvrages numériques. Certes, pour des questions de coûts, libérer de l'espace de stockage ne justifie pas de perdre les données. On considère donc par défaut comme rentable la conservation des données numériques. Il reste cependant la question de la pondération des recherches selon des critères d'obsolescence, qui pourrait faire partie de développements futurs pour des ouvrages récents.

- Quels modes de recherche ?

Dans un contexte général de réflexion sur l'apport des bibliothèques face à la puissance des moteurs de recherche commerciaux,³⁶⁵ les possibilités de l'OCR avec les recherches en plein texte induisent un nouveau rapport au livre numérisé, qui devient proche d'une page Web classique. En effet, l'indexation en *full text* ne correspond pas à l'utilisation d'un vocabulaire contrôlé. La question de la pondération des résultats dans les recherches en plein texte se pose donc, avec à la fois le calcul de pondération faisant aussi intervenir les points d'accès normés, et l'identité de la bibliothèque au travers des documents consultés. Par exemple, Gallica 2³⁶⁶ d'insérer les documents dans une interface particulière signifiant la présence de la bibliothèque numérique comme institution, et de pondérer les résultats des recherches en faisant intervenir le titre et l'auteur plus que les occurrences dans le texte dans le calcul des résultats. L'emploi combiné de la notice et du texte intégral invite en effet à privilégier les métadonnées, en ajustant l'algorithme pour ne pas survaloriser les mots du texte lui-même.

La pondération des métadonnées suppose dans ce cas d'adopter un référentiel, complétant par exemple les classifications Rameau et Dewey, pour trouver un compromis entre la visibilité des ressources les mieux décrites et celle des ressources dont le travail de description a été moins approfondi. La diversité des qualités des types de description est en effet un problème essentiel pour la recherche fédérée dans des ressources numériques. C'est ainsi que le carrousel thématique proche de la Dewey proposé pour le prototype d'Europeana correspondait moins à l'évolution vers des ressources diversifiées.

En ce sens, la mise en place d'outils de recherche est souvent alourdie par la présence d'un catalogue historique dont la conversion rétrospective amène à hériter d'une information thématique peu découpée, éventuellement sans Rameau ni Dewey, d'informations bibliographiques ne reprenant que le contenu des fiches, voire moins, et surtout de doublons parasitant les résultats des recherches.

Dans le contexte d'une navigation en ligne, ces doublons parasitent considérablement la recherche lorsqu'il s'agit de doublons d'autorités. Leur nettoyage est donc souvent un

³⁶⁵ BARTHET, Emilie ; CHURCH, Amélie, CATANESE-PALANCHE, Véronique, « Qui a besoin des bibliothèques à l'heure de Google ? », BBF, 2007, t.52, n°5, p. 92-93.

³⁶⁶ <http://gallica2.bnf.fr/>, Voir De Gallica à Gallica 2, l'évolution de la bibliothèque numérique de la BnF, In *Chroniques*, Collections, n°43, mars-avril 2008, pages 12-13.
<http://www.bnf.fr/pages/chroniques/chroniques43.pdf>

travail important,³⁶⁷ valable aussi bien pour les documents physiques que numériques, mais qui prend une importance particulière pour ces derniers.

En effet, selon les produits et services, ces données d'autorité peuvent être liées aux informations de gestion, par exemple pour la communication de documents numériques avec le réservoir Spar à la BnF, où le module Gestion de droits calcule les droits avec un arbre de décision (modulable selon les évolutions de la législation), pour prendre en compte des éléments comme la date de mort des auteurs morts pour la France (sous droits 30 ans de plus). Cela nécessite un travail sur tous les morts du XXe siècle.³⁶⁸

Or, ces techniques de gestion peuvent être étendues à la communication des documents physiques, les fonctionnalités développées pour les documents numériques étant dans ce cas transposées au reste du catalogue.

- Quelle indexation ?

Il semble évident que les habitudes de recherche ont évolué, depuis une dizaine d'années, avec l'utilisation des moteurs de recherche³⁶⁹, à la fois en eux-mêmes et par les logiques de formulation des questions et d'analyse des résultats qu'ils induisent pour le lecteur. Intellectuellement, ce type d'interrogation suppose une démarche différente, avec la perspective de limiter la demande en quelques mots synthétiques, suivant une autre approche que les modes de recherche traditionnels, même si la question de l'indexation reste essentielle. La logique d'interrogation, qui pour le lecteur peut s'assimiler à l'utilisation d'un index,³⁷⁰ peut en fait aussi être conçue comme une interrogation proche du langage naturel que le moteur est conçu pour traiter avec le plus de pertinence possible. Par conséquent, l'interrogation d'un ouvrage en plein texte ne correspond pas forcément pour l'utilisateur à une démarche différente de l'interrogation par sujet, et il peut être tout à fait légitime de combiner le plein texte avec une indexation contrôlée.

De fait, l'indexation en plein texte sous sa forme actuelle n'identifie pas les articles, et ne permet pas encore de structuration de l'information intrinsèquement supérieure à l'emploi d'un vocabulaire contrôlé. Si, pour des questions de volume de travail, l'indexation en plein texte est la seule solution réaliste pour la presse par exemple,³⁷¹ on peut envisager en complément un avenir pour les thésaurus d'autorités, comme Rameau pour les autorités matières, et les autres thésaurus avec des tables de traduction des concepts.³⁷²

Elles peuvent correspondre à des évolutions ergonomiques de catalogue, comme le proposent notamment les projets *Worldcat identities* (Pour des affichages graphiques)³⁷³ ou VIAF de l'IFLA.³⁷⁴

³⁶⁷ Par exemple pour les autorités, à la BnF, le chantier « lagarde et michard », avec auteurs par siècle pour les XVIe et XIXe, et pour Galica 2, les oeuvres emblématiques avec le chantier « académiciens », sur les homonymes et vedettes, dans le cadre de manifestations, par exemple en 2009 sur Händel, et pour des anniversaires comme celui de Berlioz, avec une exposition

³⁶⁸ A partir de Solon, entrepôt où sont collectées, complétées et conservées les métadonnées de droit.

³⁶⁹ LELOUP, Catherine, Moteurs d'indexation et de recherche : environnements client-serveur, Internet et intranet, Eyrolles, 1998.

³⁷⁰ AMAR Muriel, Les fondements théoriques de l'indexation, Ed. ADBS, 2000.

³⁷¹ Voir « La conservation et la numérisation de la presse », In *Chroniques de la Bibliothèque nationale de France*. Paris: Bibliothèque nationale de France, 1998-, n°47, janvier- février 2009, p. 5 à 9.

³⁷² On peut citer par exemple l'intérêt d'Orange pour les fichiers d'autorités géographiques de la BnF.

³⁷³ Worldcat identities, <http://orlabs.oclc.org/identities/>.

³⁷⁴ Virtual International Authority Files, <http://orlabs.oclc.org/viaf/>.

Ces projets peuvent correspondre au développement d'algorithmes sophistiqués, s'inscrivant éventuellement dans le futur développement du Web sémantique. On peut citer le projet réunissant l'OCLC, la DNB (Francfort), la Library Of Congress et la BnF, pour l'établissement de renvois d'autorité pour les auteurs physiques (et les sites géographiques plus tard), avec un fichier central des autorités entre catalogues, gérant les homonymies. Il s'agira d'un algorithme de l'OCLC, avec des renvois aux notices bibliographiques liées à l'autorité, pour désambiguïser les personnes.

On peut aussi citer le projet MACS de la Communauté Européenne,³⁷⁵ cherchant à mettre en place des autorités pour un vocabulaire en constante évolution. TEL³⁷⁶ par exemple souhaite utiliser MACS pour les matières, mais un travail important reste à faire sur la traduction des termes et les équivalences de concepts. Notamment, STITCH³⁷⁷ est un projet sur la traduction automatique entre vocabulaires (en particulier, les vocabulaires iconographiques). Les bibliothèques françaises se posent notamment la question du passage de Rameau en SKOS³⁷⁸ actuellement, contrairement à l'ancien emploi de l'ISO 2709.

5.2.2 Contrôler l'accès

La mise à disposition de documents numérisés pour les internautes correspond à l'idée de gratuité et d'ubiquité d'un document accessible de partout³⁷⁹. C'est pourquoi les ouvrages édités en ligne dans des bibliothèques numériques peuvent correspondre à un statut d'édition d'ouvrages en ligne, différente de la simple mise à disposition en salle et du prêt temporaire.

- L'ubiquité.

La présence parallèle d'ouvrages acquis et manipulés matériellement, et d'ouvrages numérisés gérés sur des serveurs, peut provoquer une évolution de la notion de collection. La distinction des supports rejoint en ce sens celle des types de documents, déjà définie auparavant, par exemple dans la norme Z 44.077 pour les images dans le catalogue multimédia. Cette norme distingue les collections « factices » et « non factices », selon les bouquets, portails, collections d'éditions, si elles sont prévues ensemble, par exemple les collections d'un livre, selon le collectionneur, avec l'origine, les formats, et les supports divers.³⁸⁰ Outre la diversité des ressources, les modes d'accès eux aussi changent, face à des programmes de numérisation de masse.³⁸¹ Ce changement majeur a entraîné un vif débat lié à l'action de Google, avec une réflexion de la BnF³⁸²,

³⁷⁵ Multilingual ACcess to Subjects, impliquant la Suisse, la France et l'Allemagne, Voir <http://rameau.bnf.fr/informations/pdf/MACS-bnf-2007.pdf>.

³⁷⁶ <http://www.theeuropeanlibrary.org>.

³⁷⁷ Semantic Interoperability To access Cultural Heritage, voir la présentation de la Journée d'étude BnF / AFNOR CG46, 28 Mars 2008, <http://www.bnf.fr/PAGES/infopro/journeespro/no-Afnor2008.htm>, en particulier l'intervention d'Antoine Isaac, Bibliothèque nationale des Pays Bas, <http://www.bnf.fr/PAGES/infopro/journeespro/pdf/AFNOR2008/Isaac.pdf>

³⁷⁸ SKOS : Simple Knowledge Organisation System, langage permettant la représentation formalisée de thésaurus, développé par le W3C pour le « Web sémantique », adapté aux vocabulaires contrôlés

³⁷⁹ LUPOVICI, Christian, « L'évolution du rôle des bibliothécaires et documentalistes dans le cadre de la bibliothèque numérique », In Les bibliothèques numériques, Dir. Fabrice Papy, Hermes-Lavoisier, 2005.

³⁸⁰ MELOT, Michel. *Une brève histoire de l'image*. Paris: l'œil neuf, 2007.

³⁸¹ Mémoire d'études de PEYRARD, Sébastien « L'accès aux collections numériques de masse », ENSSIB 2007.

³⁸² BALLEY, Noëlle, « Numérique et bibliothèques : le deuxième choc : Entretiens professionnels de la BnF », Bulletin des Bibliothèques de France, 2007, n° 3, p. 98-99. <http://bbf.enssib.fr>

et un écho au niveau mondial encore actuellement³⁸³. Il s'agit d'abord d'une augmentation importante des volumes de documents numérisés, avec des évolutions pour la recherche par OCR (reconnaissance optique de caractère) permettant une recherche en plein texte, mise en place dans Google book search et Gallica 2 par exemple.

Ces nouveaux outils impliquent la recherche de solutions,³⁸⁴ mais aussi un positionnement théorique lié à l'offre commerciale d'ouvrages en ligne. Au travers des possibilités d'interactivité, de lecture navigationnelle (hypertexte) et de rétroaction, l'édition électronique proprement dite correspond à des pratiques de lecture différentes de la simple consultation d'ouvrages numérisés, même en mode texte.³⁸⁵ Au travers de la question de l'ubiquité inhérente à la mise en ligne, les ressources numérisées peuvent donc prendre un statut hybride, entre les ressources papier et l'édition électronique proprement dite, à la fois exemplaire et notice. L'éventuelle intégration d'une offre éditoriale correspond donc à une réflexion sur le signalement des ressources.

- Les collections à accès restreint

Au départ, les bibliothèques numériques mettaient en ligne uniquement des documents du domaine public. Pour autant, l'évolution les a naturellement porté vers une offre éditoriale sous droits, dans l'idée de permettre un accès à des documents récents. Critiqué dans le cadre de Gallica 2³⁸⁶ avec l'idée que les bibliothèques y perdraient leur esprit de gratuité en renvoyant à des accès payants, cette intégration a un aspect politique lié à l'évolution de l'édition sous une forme numérique.

Dans le contexte universitaire, l'accès à des périodiques en ligne à accès payant est une pratique habituelle, avec une authentification de l'utilisateur pour lui permettre d'y accéder. On peut se référer aux bibliothèques de l'université Concordia,³⁸⁷ dont la « réserve électronique » est intégrée au catalogue sous la forme d'un module de gestion de contenu depuis 2004. Il s'agit à la fois de permettre aux étudiants d'accéder aux ressources, et de garantir un accès sécurisé, pour le dépôt par les auteurs ou la négociation avec les éditeurs. Des accès comparables existent dans diverses bibliothèques universitaires.³⁸⁸ L'accès restreint à des ressources payantes, comme le renvoi vers des sites de diffuseurs proposant des ouvrages sous droits, supposent un positionnement dans le contexte de gestion et de production des ressources.

³⁸³ Par exemple la conférence d'Alex Byrne, président de l'IFLA, <http://babel.enssib.fr/document.phd?id=990>

³⁸⁴ DUTREMEE, Régis, « La mise en ligne de documents numérisés, quels besoins et quelle offre en logiciels ? », BBF, 2007, t.52, n°5, p. 100-101.

³⁸⁵ PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique documentaire. Paris: Dunod, 2004, 342 p., p.249

³⁸⁶ Gallica 2 intègre une offre électronique de prestataires privés : http://www.bnf.fr/PAGES/catalog/gallica2_distributeurs.htm

³⁸⁷ Olivier Charbonneau, The virtual customer : A new paradigm for improving customer relations in library and information services, Satellite meeting, Sao Paulo, Brazil, August 18-20, 2004, Edited by Sueli Mara Soares Pinto Ferreira and Rejean Savard IFLA, publications 117, K.G. Saur Verlag, 385p., p.217-218.

³⁸⁸ Par exemple, la bibliothèque numérique régionale du Nord- Pas de Calais, <http://nordnum.univ-lille3.fr/nordnum2.html>.

5.2.3 Cycle de vie et de mort

La création des métadonnées administratives en vue de gérer le cycle de vie du document numérique amène à développer des fonctionnalités sur l'ensemble du catalogue.

- Gérer des contenus

Le rapport à l'unité d'une collection « dématérialisée » stockée sous forme numérique est très différent de la gestion de ressources tangibles.³⁸⁹ La perspective ancienne d'entrée par un portail virtuel reste une manifestation de la bibliothèque, et surtout, le repérage par des outils de recherche de bibliothèque correspond à une pratique possible, en particulier l'utilisation de notices bibliographiques sur les documents.

Or, l'utilisation de métadonnées comparables à celles des Content Management Systems permettent de contenir et formaliser toutes les informations nécessaires à une gestion électronique à proprement parler. Nature du contenu, auteur, lecteurs autorisés, éditeurs autorisés emplacement dans le *workflow*, utilisations et opérations possibles sont non seulement des éléments essentiels pour de futurs développements sémantiques du Web,³⁹⁰ mais aussi une couche de gestion utile en interne pour l'institution.³⁹¹

Ces répondent avant tout, dans le domaine des bibliothèques, à la gestion des droits liés aux auteurs et éditeurs, en employant un algorithme pour les documents sous droits ou libres, et une barrière mobile prenant en compte l'auteur et la publication.

- S'associer à la production ?

L'intégration des bibliothèques dans la production de métadonnées des documents numériques peut revêtir deux aspects : d'une part l'intégration adaptée d'un système informatique souple et de techniques modernes, et d'autre part la possibilité d'importer des données originales déjà produites, par un des acteurs de la chaîne de diffusion des ouvrages, ou par une autre institution. Par exemple, le Sudoc identifie fréquemment des auteurs de thèses, pour lesquels des notices d'autorité sont créées, ce qui pose un problème lorsqu'ils publient ensuite des ouvrages déposés au régime du dépôt légal.

L'intégration de la production d'information bibliographique dans le circuit de fabrication et de diffusion des ouvrages pose problème en elle-même, puisque les bibliothèques ont par nature des intérêts différents de ceux des éditeurs. En ce sens, les programmes de numérisation ont longtemps pesé sur les relations entre bibliothèques et éditeurs, ce qui tend à s'améliorer avec l'action du SNE, et les financements liés au CNL. De cette façon, la récupération de données pourrait s'effectuer avec des *mashups*, pour l'interrogation de données d'éditeurs ou du cercle de la librairie, intégrées dans la bibliographie nationale, avec éventuellement des données d'autorité en partie ouvertes sur un principe de Wikis entre les différents acteurs.

³⁸⁹ LAHARY, Dominique, «L'accès aux catalogues et ressources des bibliothèques à l'âge des bibliothèques numériques et des moteurs de recherche : écarts, perturbation, mutation ? » In Les bibliothèques numériques, Dir. Fabrice Papy, Hermes-Lavoisier, 2005.

³⁹⁰ par exemple pour naviguer dans les documents dans le cas d'applications RDF, PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique... *Op. Cit.*, p.328.

³⁹¹ pouvant par exemple figurer sous forme de source XML, comme index dans une base de données relationnelle, ou dans un autre CMS spécifique, Ibid.

Réciproquement, les données élaborées ou centralisées par les bibliothèques peuvent aussi intéresser les acteurs de la diffusion des ouvrages. Ainsi, Amazon achète des données établies par les bibliothèques publiques, même si ce n'est pas le cas en France avec la restructuration d'amazon.fr. Ces aspects pourraient évoluer rapidement avec une offre d'ouvrages numériques sous droits, pour la diffusion de livres électroniques.³⁹² Après les échecs précédents, les progrès des produits pourraient amener leurs diffuseurs à s'intéresser au signalement des ressources numérisées des bibliothèques, qui font partie des arguments de vente de ces machines.

5.3 ACCEDER AUX COLLECTIONS

La diversification des supports de lecture ne doit pas faire oublier que, pour le moment, la machine utilisée pour accéder au signalement des collections est presque uniquement l'ordinateur personnel. Pour autant, les collections numériques amènent à aborder différemment le catalogue.

5.3.1 Depuis où ?

On l'aura compris, une des ambivalences des données des catalogues dans le Web est de devoir servir à la fois pour le fonctionnement du catalogue lui-même, et pour trouver les informations ou renvois aux ressources à partir d'autres catalogues, d'entrepôts ou de moteurs.

- Depuis les catalogues ?

La présence de ressources accessibles directement depuis le catalogue amène à inverser la logique de recherche. Il s'agit de passer par l'interface de la bibliothèque pour trouver des ressources du Web. Comme le montre notamment l'enquête de la société Ourouk effectuée en 2007 sur la maquette d'Europeana³⁹³, une des attentes principales du public est que les documents soient présentés dans un contexte sérieux, la bibliothèque apportant une forme de caution. Ceci peut concerner aussi bien le contenu (établissement de la collection) que le format numérique choisi.

On peut aussi supposer que la valeur ajoutée de l'offre des bibliothèques repose sur la mise à disposition avec des instruments de recherche performants,³⁹⁴ si possible unifiés pour simplifier l'utilisation. Ainsi, le CCFR³⁹⁵ conduit aux documents de Gallica, qui renvoie non seulement aux ouvrages numérisés par la BnF, mais aussi la BIUM, la Jubilothèque (Paris 6), la Bibliothèque virtuelle Humaniste, et d'autres. On y retrouve aussi bien Bn Opale+ que le Sudoc³⁹⁶ qui signale lui aussi les documents repérés sous une forme numérisée.³⁹⁷ On observe donc une réelle complémentarité entre les ressources papier et numériques, et entre bibliothèques, pour éviter de numériser

³⁹² Kindle, Sony Reader (FNAC), et Cybook, de Bookeen (Gibert) par exemple.

³⁹³ BOUVIER-AJAM, Laurent ; LESQUINS, Noémie ; RIPON, Romuald. Europeana : étude sur les usages et les attentes relatifs à l'interface de consultation de la future Bibliothèque numérique Européenne : rapport final. 21 mars 2007, 53 p. http://bibnum.bnf.fr/usages/BnF_BNuE_Ourouk_Rapport_final.pdf.

³⁹⁴ BALLEY, Noëlle, « Numérique et bibliothèques : le deuxième choc : Entretiens professionnels de la BnF », Bulletin des Bibliothèques de France, 2007, n° 3, p. 98-99, <http://bbf.enssib.fr>

³⁹⁵ <http://www.ccf.fr/portailccfr/servlet/LoginServlet>

³⁹⁶ <http://www.sudoc.abes.fr>

³⁹⁷ Merci à Mme Frédérique Joannic-Seta (BnF) pour sa présentation, ENSSIB 2008.

plusieurs fois les mêmes ouvrages. Cette complémentarité s'étend même désormais aussi à la BnF vis-à-vis de Google, puisque pour Gallica 2 l'on évite expressément de numériser les ouvrages déjà mis en ligne par Google.

Malgré l'efficacité de ces portails, on peut supposer que la plupart des internautes cherchent les documents dont ils ont besoin au travers d'un moteur de recherche classique, en se souciant assez peu de savoir où il est hébergé. Ce besoin d'accès au document lui-même peut correspondre aussi bien à une recherche d'un document particulier et déjà repéré, qu'à une recherche sur un domaine vague sur lequel on espère trouver un document en « ayant de la chance ».

Ces deux approches sont compatibles si la recherche peut être effectuée sans passer par une interface intermédiaire, avec un document repérable aussi bien par le portail que par un moteur commercial. Ensuite, pour permettre au lecteur de découvrir la bibliothèque, l'adresse du lien peut apparaître au cours d'une recherche, par exemple avec un renvoi à la page d'accueil du document à partir d'un document texte trouvé au cours de la recherche, et surtout par un encadrement du document permettant de rebondir intuitivement sur une autre interrogation de la bibliothèque numérique (par exemple sur Gallica 2), ce qui permet au lecteur de trouver le portail d'interrogation au hasard de sa navigation. On sort ici de l'opposition simple entre les concepts de *pull* et *push*,³⁹⁸ puisque c'est à partir d'une information explicitement recherchée que l'on permet de rebondir sur un site institutionnel ou fédérateur, éventuellement inconnu. Ceci exclut la solution technique d'un Opac classique, supposant de formuler des interrogations impossibles pour un robot.

- Par des entrepôts moissonnables ?

La mise en ligne d'entrepôts de données exposées en OAI-PMH a été effectuée au départ par des « bibliothèques numériques parallèles » où les auteurs eux-mêmes déposent leurs écrits. Avec l'emploi du Dublin Core pour les métadonnées en OAI-PMH,³⁹⁹ ces techniques correspondent à des réflexions sur la gestion d'information dans le monde des bibliothèques : on peut citer notamment Patrick Bazin et son idée du « méta-langage ».⁴⁰⁰ En ce sens, on peut parler de « bibliothèques numériques natives »⁴⁰¹ pour qualifier les dépôts d'archives ouvertes structurés en OAI-PMH,⁴⁰² comme OAIster,⁴⁰³ et HAL⁴⁰⁴ en France.

Les catalogues actuels pouvant mal se prêter au modèle FRBR, une des possibilités est l'exposition des données en OAI. Par exemple, à la BnF, la fusion des comités de pilotage bibliographique (catalogue) et numérique (bibliothèque numérique) devrait

³⁹⁸ « c'est au système que revient la charge de forcer mon attention pour m'informer de quelque nouvelle que je ne cherchais pas. C'est le concept Push, », PRAX, Jean-Yves; LARCHER, Simon. La gestion électronique documentaire. Paris: Dunod, 2004, 342 p. p. 218

³⁹⁹ Par exemple : Guide d'utilisation du Dublin Core non qualifié à la BnF. http://bibnum.bnf.fr/oai/20060808_GuideOAI.pdf

⁴⁰⁰ BAZIN, Patrick, "Après l'ordre du livre", In Medium, Dir. Régis Debray, Ed. Babylone, 2005 n°4.

⁴⁰¹ A. FOX, Edward; SULEMAN, Hussein; LUO, Ming, "Building digital libraries made easy : toward open digital libraries", In Digital libraries: people, knowledge and technology, Lecture notes in computer science n°2555, 2002.

⁴⁰² NAWROCKI, François. Le protocole OAI et ses usages en bibliothèque. Paris : Ministère de la culture et de la communication, 2005. <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>

⁴⁰³ <http://www.oaister.org>

⁴⁰⁴ <http://hal.archives-ouvertes.fr>

permettre la création d'entrepôts OAI, avec une dissociation entre les catalogues et les documents numériques.⁴⁰⁵

- Par des pages Html ?

Les documents numériques ne sont en pratique accessibles pour l'utilisateur qu'au travers de moteurs de recherche, commerciaux ou internes. En ce sens l'évolution des catalogues vers un modèle FRBR crée un besoin de générer des pages Html qui soient simples à indexer.

L'information bibliographique peut dans ce cas être intégrée sous la forme de balises, pour améliorer à la fois le référencement du site et la pertinence des résultats, comme le permet le protocole Sitemap⁴⁰⁶, en déclarant la structure du site dans un fichier XML, qui indique au moteur de recherche les pages à indexer.⁴⁰⁷ Les en-têtes de documents Html en Dublin Core ou autres formats de l'information bibliographique, correspondent donc à la mise en valeur des anciens éléments de notices dans le code source. Cette technique entraîne pour le moment une méfiance des moteurs.⁴⁰⁸

L'intégration d'informations bibliographiques dans les codes sources des pages, facilitée par la diffusion d'XML sur Web, reste cependant efficace pour les moteurs de recherche propres (avec les algorithmes de pondération), et répond à sa fonction première de faciliter l'indexation par les moteurs.

5.3.2 Jusqu'où ?

La présence parallèle des instruments de recherche et des ressources dans le Web amène à s'interroger sur les limites souhaitables de l'action des bibliothèques.

- Des services payants ?

La négociation de ressources en ligne, notamment avec les éditeurs de périodiques électroniques incontournables en sciences dures,⁴⁰⁹ l'accès avec login ou depuis une adresse IP identifiée est admise dans le contexte universitaire.⁴¹⁰ L'expérience d'accès à des objets sous droits depuis des catalogues de bibliothèques publiques est comparable. Ainsi, Gallica 2 propose les références de 5000 ouvrages électroniques vendus en ligne, par Numilog⁴¹¹ et d'autres.⁴¹² Ce travail visant à renvoyer aussi à des ouvrages récents sur une question posée, est en fait largement perçu comme un portail d'appel, critiqué par des internautes qui y voient une compromission des bibliothèques avec le secteur commercial, effectivement parfois surprenante en l'absence de charte documentaire.

⁴⁰⁵ A la BnF, l'OAI-CAT expose BNOpale + et OAI-NUM contient les notices des documents numérisés.

⁴⁰⁶ Présentation du protocole Sitemap à l'adresse <http://www.sitemaps.org>

⁴⁰⁷ Voir PEYRARD, Sébastien, *L'accès aux collections numériques ...*, Op. Cit..

⁴⁰⁸ Par crainte du « méta-spam » (informations fausses pour induire en erreur) et du « méta-crap » (informations introduites par erreur, par exemple instructions de réalisation laissées par le catalogueur).

⁴⁰⁹ Elsevier notamment, <http://www.elsevier.com>.

⁴¹⁰ Avec l'action de l'ABES, et la BIMP (liste collaborative, par Archimed, gérée par Ebsco).

⁴¹¹ Agrégateur de ressources électroniques, qui négocie avec les éditeurs (Masson, Hachette, EDP sciences). <http://www.numilog.com/accueil.asp>

⁴¹² De plus en plus nombreux, car les éditeurs veulent plusieurs distributeurs : On trouve en liste sur http://www.bnf.fr/PAGES/catalog/gallica2_distributeurs.htm.

- Un catalogue de lecteurs ?

A l'issue des essais de la maquette Europeana, une enquête d'usages⁴¹³ effectuée auprès des utilisateurs et s'intéressant notamment à la pertinence de mettre en place un espace collaboratif avait montré un désaccord profond. Les utilisateurs consultés formulaient principalement trois souhaits : la possibilité d'usages collaboratifs (sur les fonds traités), l'authentification des informations, et surtout que la bibliothèque numérique reste une bibliothèque publique, gratuite.

Ces attentes correspondent à des évolutions des pratiques des lecteurs, comme le montre l'étude effectuée par un groupe d'enseignants-chercheurs de l'association Paragraphe de l'université de Paris VIII,⁴¹⁴ avec une demande forte des chercheurs de participer à des publications en ligne, à la fois dans la construction de contenu, de méta-données et dans l'utilisation. En France, cette démarche ne s'inscrit pas vraiment dans l'action des bibliothèques, mais elle semble parfois, notamment aux Etats-Unis, très liée aux bibliothèques numériques.⁴¹⁵ Elle est aussi liée aux modes d'évaluation des chercheurs : en particulier, la notation « bibliométrique » des chercheurs à partir de leurs publications conduit les membres de laboratoires à signer leurs publications sous une forme uniformisée, avec des éléments standardisés d'identification. Cet aspect d'authentification des informations peut justifier l'intégration par la bibliothèque de documents qui pourraient être publiés plus simplement, comme le propose le projet Manum⁴¹⁶ pour les étudiants de 1^{er} et 2^e cycles en SHS, avec des documents pédagogiques mis en ligne à la demande, accessibles par un guichet unique,⁴¹⁷ avec un modèle d'unification inspiré des revues payantes.⁴¹⁸ En se limitant aux catalogues de bibliothèques, le modèle américain de Worldcat fonctionne de façon inverse, puisqu'il est rattaché à Google Book Search, permettant d'interroger des catalogues depuis des ressources numériques.

On observe ainsi une dissociation entre les attentes des bibliothèques de recherche et de lecture publique. En effet, la recherche universitaire tend bien sûr à la mondialisation, avec une unité des modes d'interrogation (Par exemple avec Worldcat), avec un besoin d'utiliser des ressources en ligne quelles que soient les disciplines. A l'inverse, les bibliothèques de lecture publique semblent tendre à la spécialisation, le lecteur se souciant peu de savoir si un ouvrage intéressant se trouve ailleurs que là où il va ; c'est en tout cas une des conclusions du groupe de Fairfield travaillant sur les OPAC⁴¹⁹, qui préconise une modernisation sans fédération systématique, car le catalogue se limite en

⁴¹³ BOUVIER-AJAM, Laurent ; LESQUINS, Noémie ; RIPON, Romuald. Europeana : étude sur les usages et les attentes relatifs à l'interface de consultation de la future Bibliothèque numérique Européenne : rapport final. 21 mars 2007, 53 p.

http://bibnum.bnf.fr/usages/BnF_BNuE_Ourouk_Rapport_final.pdf.

⁴¹⁴ Compte rendu par LE SAUX, Annie, « Les bibliothèques à l'heure du numérique : nouveaux lieux, nouveaux usages ? », BBF, 2005, n°5, p. 76-77.

⁴¹⁵ UNSWORTH, John, "The emergence of digital scholarship: new models for librarians, scholars, and publishers", In Digital library development: the view from Kanazawa, Ed. Daenna B. Marcum and Gerald George, 2006.

⁴¹⁶ <http://manum.enssib.fr>, Minon, Marc, Manum, projet de bibliothèque numérique pour les SHS, 2003.

⁴¹⁷ Ghislaine Chartron et Marc Minon, Quell, « Quelle offre numérique pour l'utilisateur « virtuel » des universités françaises », In The virtual customer : A new paradigm for improving customer relations in library and information services, Satellite meeting, Sao Paulo, Brazil, August 18-20, 2004, IFLA, publications 117, K.G. Saur Verlag, 385p., p.316-330.

⁴¹⁸ M. GUTHRIE, Kevin: "JSTOR: Offering access to an archive of scholarly journal literature", In Digital library development: the view from Kanazawa, Ed. Daenna B. Marcum and Gerald George, 2006.

⁴¹⁹ MARCIN, Susan; MORRIS, P., « Opac : the next generation », In Computers in libraries, Mai 2008.

ce cas à pour recherche locale des ouvrages déjà connus mais non trouvés en libre accès.⁴²⁰ Cette conclusion peut remettre en cause la pertinence d'associer directement les ressources numériques aux résultats des OPACs de lecture publique. Françoise Benhamou remarque que l'augmentation des pratiques immatérielles s'accompagne paradoxalement de la demande de lieux ancrés physiquement,⁴²¹ face à laquelle les bibliothèques recherchent parfois des solutions mixtes,⁴²² conciliant l'offre numérique et l'utilisation de documents sur place.

- La bibliothèque partout ?

L'« *age of access* » évoqué par Jeremy Rifkin⁴²³ est liée à l'idée de « *browsing* », où les ressources doivent être intégrées à un parcours de navigation. Par exemple, les actuels pourparlers entre la BnF et Wikipédia devaient permettre l'élaboration de liens croisés, surtout sur Gallica 2, augmentant la popularité des sites, et permettant la labellisation des articles liés aux ouvrages de la BnF. En un sens, l'établissement de liens vers des ressources de bibliothèques depuis Wikipédia constitue une nouvelle forme de catalogage, avec des descriptions libres et collaboratives des ressources, le seul élément normalisé par l'institution étant l'URL. Dans ce but, l'adaptation des structures administratives des bibliothèques sera sans doute un des points majeurs de l'évolution, pour permettre des changements. Au travers de l'implication dans des sites extérieurs, se posent la question technique des possibilités d'influence efficace, et la question éthique du prosélytisme : jusqu'où faut-il chercher à attirer vers bibliothèques physiques ? Ces choix sont liés à l'évolution de l'offre éditoriale, avec les changements du marché du livre, y compris avec le développement d'une offre électronique où s'affirme l'hégémonie de Google.

Comme on a pu l'observer récemment dans le rapport Isaac⁴²⁴ remis à madame Valérie Pécresse sur « l'université numérique », qui montre une grande confiance dans l'efficacité des ressources numériques, la légitimité et l'intérêt de l'offre semblent ne jamais être remis en cause, même au travers de développement parfois anarchiques,⁴²⁵ alors qu'« une bibliothèque numérique n'est pas une bibliothèque numérisée ».⁴²⁶ Cependant, outre l'aspect idéologique de la croyance en un passage forcément souhaitable à un monde numérique, on peut observer que le mode d'accueil du lecteur est radicalement changé, dans l'idée que même des sites nationaux importants comme Revues.org et Cairn sont peu connus par leurs utilisateurs potentiels. Le signalement des ressources numériques suppose donc une adaptation au Web plus profonde que la simple intégration de liens depuis les outils de recherche traditionnels.

⁴²⁰ DUCHEMIN, Pierre-Yv., « L'enrichissement des catalogues ? Et après ? », *BBF*, 2005, n°4, p. 21-27.

⁴²¹ VERNEUIL, Anne, « L'avenir des bibliothèques à l'ère du Web 2.0 et face aux évolutions d'Internet », *BBF*, 2007, t.52, n°5, p. 93-94.

⁴²² PAPY, Fabrice ; CHAUVIN, Sophie, « Au-delà de la transfiguration du catalogue : le Visual...Catalog », *BBF*, 2005, n°4, p. 5-12

⁴²³ RIFKIN, Jeremy ; *L'âge de l'accès : la nouvelle culture du capitalisme*. Paris: La Découverte, 2005.

⁴²⁴ ISAAC, Henri, rapport sur l'Université numérique remis à madame Valérie Pécresse, ministre de l'Enseignement supérieur et de la recherche :

http://www.nouvelleuniversite.gouv.fr/IMG/pdf/Rapport_univ_num.pdf

⁴²⁵ Selon FRESSARD, Olivier, BU de Paris VIII St. Denis, « L'esprit du numérique : bibliothèques numériques et démocratie », In PAPY, Fabrice. *Usages et pratiques dans les bibliothèques numériques*. Paris : Hermès science publications: Lavoisier, 2007, p. 123-143.

⁴²⁶ Patrick Bazin, en réponse à Jean-Noël Jeanneney sur numérisation de 300000 documents sans collection ni corpus, voir BAZIN, Patrick, «Après l'ordre du livre», In DEBRAY, Régis. *Médium: transmettre pour innover*. Paris: Éd. Babylone, 2004-, 2005 n°4.

6. Conclusion

L'évolution des catalogues doit répondre à un changement de problématique, avec le passage d'une présentation des collections à une réutilisation des données dans un univers où tout est structuré *a posteriori*. En effet, les changements d'usages induits par l'ergonomie du Web vont bien au-delà de simples possibilités techniques. Il s'agit vraiment d'un changement des rapports aux contenus recherchés en ligne. Si les outils ne s'adaptent pas, ils peuvent devenir de simples instruments de localisation, pour trouver dans un lieu des ouvrages déjà connus, à l'inverse d'une participation aux réseaux d'une information mondialisée.

Les idées et perspectives de solutions correspondent donc à des attentes diverses, selon les acteurs impliqués et les types d'usages auxquels ils souhaitent répondre.

On a pu évoquer au travers des différents projets développés une partie des actions des principaux protagonistes que sont l'IFLA, l'ALA, l'OCLC, la Bibliothèque du Congrès, le JSC, la BnF, l'AFNOR et l'ABES.

Elles visent, de façon générale, à la fédération des ressources et à l'unification des interfaces de recherche, au travers de l'interopérabilité des données. Or, pour les bibliothèques de lecture publique, le catalogue est parfois perçu comme un simple outil de gestion, le mode de contact normal entre les lecteurs et les ouvrages étant le libre accès, ce qui correspond d'ailleurs largement aux usages commerciaux. Il ne s'agit pas seulement de répondre à la demande des personnes non reliées à Internet, ou ne désirant pas l'utiliser, mais aussi de fournir une offre sur place où le catalogue puisse être considéré comme un simple outil de professionnels, comme dans les circuits commerciaux de grande et petite distribution. Ce qui n'exclut pas, justement, la fédération des interrogations, puisque l'éventuelle disponibilité de l'ouvrage dans une bibliothèque voisine est un élément de réponse important. On pourrait donc envisager une scission des outils pour les bibliothèques locales, entre des outils Web, et des instruments de recherche strictement locaux.

L'action de chaque institution sur son catalogue doit donc se situer dans le contexte national et mondial. En particulier, l'usage d'Internet entraîne une évolution des pratiques de recherche, et des changements des modes de consommation, auxquels peut correspondre une évolution souhaitable des catalogues.⁴²⁷ Avec d'une part les progrès du travail d'uniformisation des données effectué depuis les années 1960, et d'autre part le souhait des établissements de consacrer une part des crédits de moins en moins grande au catalogage, il est probable que la production bibliographique en elle-même se limite de plus en plus à la description des *unica*,⁴²⁸ éventuellement en lien avec les politiques de numérisation.

Malgré un principe de bon sens d'unifier des données et de ne pas reproduire le même travail, il est délicat de modifier des habitudes anciennes, qu'il s'agisse des diverses

⁴²⁷ Voir par exemple <http://www.bibliobsession.net/2007/04/16/naviguer-dans-la-longue-traine-necessite-des-mediateurs/>

⁴²⁸ En particulier des livres manuscrits, des manuscrits d'auteurs, des archives.

bibliothèques, ou des autres professions du commerce du livre utilisant elles aussi des données bibliographiques.⁴²⁹

Pourtant, les évolutions du Web à court terme rendront probablement nécessaires des évolutions allant bien au-delà des fonctionnalités intégrant les usagers et des formats échangeables en XML. En effet, il est possible que les changements des modes de description liées aux innovations sémantiques rendent les métadonnées et traitements secondaires des informations plus nécessaires que jamais, tout en posant une seconde fois la question du lien entre documents physiques et numériques, qui échapperont alors aux institutions, quels que soient les possesseurs des fichiers. Mais par la masse documentaire qu'elles représentent, les bibliothèques peuvent être un des acteurs majeurs de la réalisation du Web Sémantique.⁴³⁰

⁴²⁹ Voir PATINO, Bruno, rapport remis à madame Christine Albanel, ministre de la culture, sur le livre numérique, 30 juin 2008 :

<http://www.culture.gouv.fr/culture/actualites/conferen/albanel/rapportpatino.pdf>

⁴³⁰ C'est en tout cas la vision optimiste de Christian Fauré, <http://www.christian-faure.net/2006/04/22/puissance-des-bibliothcaires/>.

Bibliographie

BIBLIOGRAPHIE.....	93
MONOGRAPHIES	93
<i>Bibliothèques</i>	93
- Bibliothèques et usages documentaires.....	93
- Bibliothèques numériques et catalogues collectifs	94
- Schémas de description des bibliothèques	95
<i>Informatique</i>	96
- Traitement des contenus et métadonnées	96
- Outils et réseaux.....	97
- Normes	98
PERIODIQUES SPECIALISES.....	98
- <i>Articles de périodiques papier</i>	98
- <i>Articles de périodiques en ligne</i>	99
SITES INTERNET.....	99
- <i>Sites institutionnels</i>	100
- <i>Blogs</i>	100
- <i>Littérature grise en ligne</i>	100
- <i>Ressources de la BnF</i>	100
- <i>Portails et catalogues utilisés pour ce travail</i>	101

MONOGRAPHIES

Bibliothèques

- Bibliothèques et usages documentaires

ACCART, Jean-Philippe ; BAZIN, Patrick. *Les services de référence. : Du présentiel au virtuel*. Paris: Ed. du Cercle de la Librairie, 2008.

Association des archivistes français. *Abrégé d'archivistique : principes et pratiques du métier d'archiviste*. Paris: Association des archivistes français, 2007.

BENAYACHE, Ahcène; ABEL, Marie-Hélène. *Construction d'une mémoire organisationnelle de formation et Évaluation dans un contexte e-learning : Le projet MEMORAe*. Compiègne: 2005.

BOULOGNE, Arlette; *Vocabulaire de la documentation*. Paris: ADBS Editions, 2004.

BERISHA-BOHE, Suela. *Système d'aide à l'utilisateur pour la conception de documents scientifiques adaptés à la recherche d'information par le contenu : cas des bibliothèques numériques de thèses scientifiques*. Lyon, CNAM: 2005.

DE ROSA, Cathy, et al. *The 2003 OCLC environmental scan : pattern recognition : a report to the OCLC membership*. Dublin, Ohio: OCLC Online Computer Library Center, Inc., 2000.

GALAUP, Xavier; LAHARY, Dominique. *L'utilisateur co-créateur des services en bibliothèque publique : l'exemple des services non- documentaires*. Villeurbanne, ENSSIB: 2007.

HUDRISIER, Henri, L'iconothèque, documentation audiovisuelle et banques d'images, La Documentation française, collection « audiovisuel et Communication », 1982.

LEBRETON, Claire, Bibliothèques, tags et folksonomies : L'indexation des bibliothèques à l'ère sociale, Mémoire ENSSIB, 2007, 117 p.

LUPOVICI, Christian, « L'évolution du rôle des bibliothécaires et documentalistes dans le cadre de la bibliothèque numérique », In Les bibliothèques numériques, Dir. Fabrice Papy, Hermes-Lavoisier, 2005.

MELOT, Michel. *Une brève histoire de l'image*. Paris: l'œil neuf, 2007.

YEE, Martha M.; LAYNE, Sara S. *Improving online public access catalogs*. Chicago: American Library Association, 1998.

- Bibliothèques numériques et catalogues collectifs

BROUDOUX, Evelyne; CHARTRON, Ghislaine. Document numérique et société. *Traitements et pratiques documentaires : vers un changement de paradigme ? : actes de la deuxième conférence Document numérique et société, Paris, CNAM, 17-18 novembre 2008*. Paris: ADBS Éditions, 2008.

BURESI, Charlette; CÉDELLE-JOUBERT, Laure. *Conduire un projet de numérisation*. Villeurbanne : ENSSIB ; Paris : Éd. Tec et Doc: Lavoisier, 2002.

FESQUET, Guillaume; COUSINS, Jill. *Restructuration et mise à jour du registre des métadonnées de The European Library*. Villeurbanne, ENSSIB: 2007.

FOULONNEAU, Muriel; RILEY, Jenn. *Metadata for digital resources : implementation, systems design and interoperability*. Oxford: Chandos Publishing, cop. 2008.

GUERRE, Louise; LUPOVICI, Christian. *Guide pratique : les publications Électroniques*. Paris: Archimag, DL 2003.

LAHARY, Dominique «L'accès aux catalogues et ressources des bibliothèques à l'âge des bibliothèques numériques et des moteurs de recherche : écarts, perturbation, mutation ? » In Les bibliothèques numériques, Dir. Fabrice Papy, Hermes-Lavoisier, 2005.

- LE MOAL, Jean-Claude, et al. *Bibliothèques numériques*. Paris: ADBS Ed., 2000, 246 p.
- LIU, Jia. *Metadata and its applications in the digital library : approaches and practices*. Westport, Conn.: Libraries Unlimited, 2007.
- LUPOVICI, Catherine. *La pérennité des documents numériques : Grenoble 21 septembre 2000*.
- MARCUM, Deanna B., et al. *Digital library development : the view from Kanazawa*. Westport (Conn.): Libraries Unlimited, 2006.
- MASANES, Julien; Dir. LUPOVICI, Catherine, *L'archivage des sites Internet : rapport de stage. Diplôme de conservateur de bibliothèques*. Villeurbanne: ENSSIB, 2000.
- MESTRE, Stéphane. *Publication et gestion de documents s'intégrant dans un environnement sécurisé de type portail : mise en oeuvre des processus de publication*. s.l. , 2005.
- PAPY, Fabrice; EUVRARD, Gilles François. *Les bibliothèques numériques*. Paris : Hermès science: Lavoisier, 2005.
- PEDAUQUE, Roger T.; SALAÛN, Jean-Michel, MELOT, Michel. *Le document à la lumière du numérique*. Caen: C&F Éditions, impr. 2006.
- PEYRARD, Sébastien, *L'accès aux collections numériques de masse*, Mémoire ENSSIB, 2007, 85 p.
- SOARES PINTO FERREIRA, Sueli Mara. Éditeur scientifique; SAVARD, Réjean *The virtual customer : a new paradigm for improving customer relations in libraries and information services : satellite meeting, Sao Paulo, Brasil, August 18-20, 2004* München: KG Saur, 2005

- Schémas de description des bibliothèques

- CHANTERAUD, Annabel, et al. *Une adaptation de la norme XML-EAD : l'informatisation des inventaires archivistiques du Département des manuscrits de la Bibliothèque publique universitaire de Genève*. Villeurbanne, ENSSIB: 2003.
- CHAPLIN, A. et al. *International Conference on Cataloguing Principles, Paris, 9th-18th October, 1961 : report*. London: IFLA International Office for UBC, 1963.
- COLE, Timothy W.; FOULONNEAU, Muriel. *Using the Open Archives Initiative protocol for metadata harvesting*. Westport, Conn.: Libraries Unlimited, 2007
- DELACOUR, Christiane E. s., et al. *Cataloguer : mode d'emploi : initiation aux techniques du catalogue*. Paris : ABF: diff. ABIS, 2006.
- DOOLEY, Jackie M. *Encoded Archival Description : context, theory, and case studies*. Chicago: Society of American Archivists, 1998.
- DUSSERT-CARBONE, Isabelle, et al. *Le catalogage : méthode et pratiques*. Paris: Ed. du Cercle de la librairie, 1991-1992 (2 vol.).

Fédération internationale des associations de bibliothécaires et des bibliothèques;
Comité français UNIMARC. *Manuel Unimarc: format bibliographique*. München: KG
Saur, 2007.

LE BOEUF, Patrick É s. *Functional requirements for bibliographic records (FRBR) : hype or
cure-all ?* Binghamton (NY): The Haworth information press, 2005.

SCOPSI, Claire ; SOUAL, Laurent ; FERRAILLE, Jean-François. *Mener un projet Open
Source en bibliothèque, documentation et archives*. Paris: Éd. du Cercle de la librairie,
2007.

TILLET, Barbara B. É s. *IFLA cataloguing principles : steps towards an international
cataloguing code, 2 : report from the 2nd IFLA Meeting of Experts on an International
Cataloguing Code, Buenos Aires, Argentina, 2004*, München: K. G. Saur, 2005.

Informatique

- Traitement des contenus et métadonnées

ALLEMANG, Dean; HENDLER, James A. *Semantic Web for the working ontologist :
effective modeling in RDFS and OWL*. San Francisco, Calif.: Morgan Kaufmann, 2008.

AMAR, Muriel; LE GUERN, Michel. *Les fondements théoriques de l'indexation : une
approche linguistique*. Lyon 2 : 1997.

AUBRY, Stéphane, et al. *Annotations et gestion des connaissances en environnement virtuel
collaboratif*. Compiègne: 2007.

BALDE, Abdourahamane ; DIDAY, Win. *Utilisation de métadonnées pour l'aide à
l'interprétation des résultats de classification automatique*. Université Paris-Dauphine: 2007.

CLAVEL, Gilles. *XML, la synthèse : intégrez XML dans vos architectures*. Paris: Dunod,
2002.

DAILLY, Romain. *Web sémantique et e-learning : Ontologie pédagogique d'une plate-forme
d'apprentissage*. Nantes, CNAM: 2005.

HAROLD, Elliotte R., et al. *XML en concentré : manuel de référence*. Paris: O'Reilly, 2005.

HASSANALY, Parina., et al. *Eighth international conference on the design of cooperative
systems : proceedings, Carry-le-Rouet, Provence, France, May 20th-23th, 2008*. Aix en
Provence: Institut d'Etudes Politiques, 2008.

HIDOINE, Bernard, et al. *Métadonnées : mutations et perspectives : séminaire INRIA, 29
septembre - 3 octobre 2008, Dijon*. Paris: ABDS Éditions, 2008.

IHADJADENE, Madjid. *Les systèmes de recherche d'informations : modèles conceptuels*.
Paris : Hermès science publ.: Lavoisier, 2004.

INTNER, Sheila S.; *Metadata and its impact on libraries*. Westport, Conn.: Libraries Unlimited, 2006.

KEFI, Hassen; REYNAUD, Chantaland. *Ontologies et aide à l'utilisateur pour l'interrogation de sources multiples et hétérogènes*. s.l.: 2006.

MARCO, David; and JENNINGS, Michael. *Universal meta data models*. Indianapolis, Ind.: Wiley, 2004.

POWERS, Shelley. *Practical RDF*. Beijing ; Cambridge ; Paris: O'Reilly, 2003.

PRAX, Jean-Yves; LARCHER, Simon. *La gestion électronique documentaire*. Paris: Dunod, 2004, 342 p.

PETITGAND, Gérald. *Mise en place d'une stratégie de gestion Électronique de documents*. Versailles: CNAM, 2007.

RIETSCH, Jean-Marc, et al. *Dématérialisation et archivage Électronique : mise en oeuvre de l'ILM (Information Lifecycle Management)*. Paris : Dunod: 01 informatique, 2006.

ROXIN, Ioan ; MERCIER, Daniel ; JOLION, Jean-Michel. *Multimédia : les fondamentaux : introduction à la représentation numérique*. Paris: Vuibert, 2004.

SEDES, Florence. *Métadonnées et nouveaux SI*. Paris : Hermès science publ.: Lavoisier, 2007.

TANNENBAUM, Adrienne. *Metadata solutions : using metamodels, repositories, XML, and enterprise portals to generate information on demand*. Boston ; San Francisco ; New York: Addison-Wesley, 2002.

WARIN, Bruno. *L'algorithmique : votre passeport informatique pour la programmation*. Paris: Ellipses, 2002.

WOOTTON, Cliff. *Developing quality metadata : building innovative tools and workflow solutions*. Amsterdam ; Boston : Elsevier: Focal Press, op. 2007.

- Outils et réseaux

ÉLIE, Jean-Philippe. *Tout sur les outils Google*. Paris : Dunod, 2008, 212 p.

GERVAIS, Jean-François. *Web 2.0 : les internautes au pouvoir : blogs, réseaux sociaux, partage de vidéos, mashups..* Paris: Dunod, impr. 2007, impr. 2007, 216 p.

GOUYET, Jean-Noël; GERVAIS, Jean-François. *Gestion des médias numériques : digital media asset management*. Paris : Dunod ; 2006.

GREGOIRE, Gilles. *Le référencement sur Google*. Paris: Micro application, impr. 2008, impr. 2008.

JEANNENEY, Jean-Noël. *Quand Google défie l'Europe : plaidoyer pour un sursaut*. Paris: Mille et une nuits, 2005.

LANGVILLE, Amy N.; MEYER, Carl D. *Google's PageRank and beyond : the science of search engine rankings*. Princeton (N.J.) ; Oxford: Princeton University Press, cop. 2006.

LAUDON, Kenneth C., et al. *Management des systèmes d'information*. Paris: Pearson Education, DL 2006, DL 2006.

LELOUP, Catherine. *Moteurs d'indexation et de recherche : Environnements client-serveur, Internet et Intranet*. Paris: Eyrolles, 1997.

LE MOAL, Jean-Claude, et al. *La recherche d'information sur les réseaux*. Paris: ADBS Éd., 2002.

PILLOU, Jean-François. *Tout sur les systèmes d'information*. Paris: Dunod, 2006.

WISE, David A.; MALSEED, Mark. *The Google story*. New York: Delacorte Press, 2005.

- Normes

Association française de normalisation. *Technologies de l'information pour l'Éducation, la formation et l'apprentissage : profil français d'application du LOM (LOMFR) : Métadonnées pour l'enseignement = Information technologies for education, training and learning : french LOM application profile (LOMPR) : learning metadata = Informationstechnik für Erziehung, Ausbildung und Lehre : Französisches Profil für die LOM-Anwendung (LOMFR) : Metadaten für das Unterrichtswesen*. La Plaine Saint-Denis: AFNOR, 2006.

Fédération internationale des associations de bibliothécaires et des bibliothèques. Groupe de travail sur les spécificités fonctionnelles des notices bibliographiques. Paris: Bibliothèque nationale de France, 2001. <http://www.bnf.fr/pages/infopro/normes/pdf/FRBR.pdf> pour la traduction du texte des FRBR

Organisation internationale de normalisation. *Records management processes : metadata for records. Part 1, Principles = Processus de gestion des enregistrements : mÉta donnÉes pour les enregistrements. Partie 1, Principes*. Geneva: ISO, 2006.

PERIODIQUES SPECIALISES

- Articles de périodiques papier

BAZIN, Patrick, "Après l'ordre du livre", In DEBRAY, Régis. *Médium: transmettre pour innover*. Paris: Éd. Babel, 2004-, 2005 n°4.

FRESSARD, Olivier, BU de Paris VIII St. Denis, « L'esprit du numérique : bibliothèques numériques et démocratie », In PAPY, Fabrice. *Usages et pratiques dans les bibliothèques numériques*. Paris : Hermès science publications: Lavoisier, 2007, p. 123-143.

LUPOVICI, C. – « Du catalogue informatisé à la navigation dans l'information numérique ». – *Revue de la Bibliothèque nationale de France*, no 9, 2002.

MARCIN, Susan; MORRIS, Peter. OPAC: the next generation: placing an encore front end onto a SirsiDynix ILS. In *Computers in librairies [Ressource Électronique]*. Westport, CT: Meckler, 1989-, Mai 2008.

Articles de périodiques en ligne

Computers in libraries [Ressource Électronique]. Westport, CT: Meckler, 1989-. : *Passim*.

Bulletin des Bibliothèques de France, <http://bbf.enssib.fr> (Consultés en Décembre 2009), et particulièrement :

BERMES, Emmanuelle : « Les moteurs de recherche : petit précis de mécanique à l'usage des bibliothèques numériques ». Bulletin des bibliothèques de France, 2007, n°6, p. 5-10.

BROPHY, Peter, « La bibliothèque hybride », *BBF*, 2002, n° 4, p. 14-20.

DUTREMEE, Régis, « La mise en ligne de documents numérisés, quels besoins et quelle offre en logiciels ? », *BBF*, 2007, t.52, n°5, p. 100-101.

LUPOVICI, Catherine, « L'information bibliographique des documents électroniques », *BBF*, 1998, n° 4, p. 42-47

LUPOVICI, Catherine ; CLOAREC, Thierry ; CHARENTENAY, France de, « Les usages de Gallica », *BBF*, 2003, n° 4, p. 40-44.

PAPY, Fabrice, « Au-delà de la transfiguration du catalogue : Le « visual... catalog », *BBF*, 2005, n° 4, p. 5-12.

SITES INTERNET

Nous ne citons ici qu'une partie des multiples ressources utilisées pour ce travail, qui ne sont elles-mêmes qu'une partie des ressources existantes.

TOUTES CES ADRESSES ONT ÉTÉ VÉRIFIÉES EN DÉCEMBRE 2008

- Sites institutionnels

IFLA : <http://www.ifla.org>
ALA : <http://www.ala.org/>
OCLC : <http://www.oclc.org>
LOC : <http://www.loc.gov/>
JSC : <http://www.collectionscanada.gc.ca/jsc/index.html>
BnF : <http://www.bnf.fr>
ABES : <http://www.abes.fr>

- Blogs

Lorcan Dempsey's weblog : <http://orweblog.oclc.org>
Catalogablog, <http://catalogablog.blogspot.com/>,
Figoblog : <http://www.figoblog.org/>
Les petites cases : <http://www.lespetitescases.net/>
Bibliobsession : <http://www.bibliobsession.net/>
Affordance : <http://affordance.typepad.com/>

- Littérature grise en ligne

Il ne s'agit là que de quelques-uns des documents cités dans ce mémoire.

BOUVIER-AJAM, Laurent ; LESQUINS, Noémie ; RIPON, Romuald. Europeana : étude sur les usages et les attentes relatifs à l'interface de consultation de la future Bibliothèque numérique Européenne : rapport final. 21 mars 2007, 53 p. http://bibnum.bnf.fr/usages/BnF_BNuE_Ourouk_Rapport_final.pdf.

HILLMANN, Diane, Facing Forward: The Challenges Facing Cataloging and Catalogers, <http://hdl.handle.net/1813/11536>

HILLMANN, Diane, The Future of Cataloging and Catalogers, <http://www.slideshare.net/kramsey/the-future-of-cataloging-and-catalogers-presentation?type=powerpoint>

- Ressources de la BnF

De nombreux documents ont été utilisés, notamment :
« Des métadonnées pour bien utiliser les ressources électroniques »,
Journée d'information AFNOR/CG46, 7 juin 2005
<http://www.bnf.fr/pages/infopro/journeespro/no-Afnor2005.htm>

10e journée des pôles associés, Convergence des politiques pour l'accès aux documents, 6 décembre 2006 http://www.bnf.fr/pages/infopro/journeespro/po_2006.htm

11e journées des pôles associés, Politiques et réseaux de coopération : enjeux régionaux, nationaux et internationaux 27 - 28 mars 2008

http://www.bnf.fr/pages/infopro/journeespro/po_2008.htm

Réunion d'information-débat des

utilisateurs des produits et services bibliographiques,

29 mai 2008, http://www.bnf.fr/pages/infopro/journeespro/jp_produits2008.htm

Cinquième Journées d'information RAMEAU, 30 mai 2008,

<http://rameau.bnf.fr/informations/journees.htm#mai>

Déployer des solutions Open source dans les bibliothèques, centre de documentation et services d'archives, 13 novembre 2007,

http://www.bnf.fr/pages/infopro/journeespro/jp_opensource.htm.

BnF : Guide d'utilisation du Dublin Core non qualifié à la BnF.

http://bibnum.bnf.fr/oai/20060808_GuideOAI.pdf

De Gallica à Gallica 2, l'évolution de la bibliothèque numérique de la BnF, in *Chroniques*, Collections, n°43, mars-avril 2008, pages 12-13.

En ligne sur <http://www.bnf.fr/pages/chroniques/chroniques43.pdf>

Les entrepôts OAI de la Bibliothèque nationale de France

<http://bibnum.bnf.fr/oai/index.html>

LEDOUX, Thomas, Le Système de Préservation et d'Archivage Réparti de la BnF :

De la production à la pérennisation des objets numériques, présentation lors des 11e journées des pôles associés de la BnF, 27 - 28 mars 2008, En ligne sur

http://www.bnf.fr/pages/infopro/journeespro/pdf/poles_pdf/poles2008/spar.pdf

- Portails et catalogues utilisés pour ce travail

Archives Numérisées (Base du Ministère de la culture) : <http://sdx.culture.fr/sdx/anum/>

Bibliothèque de la Cité des Sciences et de l'Industrie, <http://www.cite-sciences.fr>

Bibliothèque numérique régionale du Nord- Pas de Calais, <http://nordnum.univ-lille3.fr/nordnum2.html>

Calames (Catalogue) : <http://calames.abes.fr>

CCFR : <http://www.ccf.fr/bnf/fr/portailccfr/servlet/LoginServlet>

Centre d'Etude d'Histoire Juridique, disponibles à partir du site des Archives Nationales, <http://www.archivesnationales.culture.gouv.fr>

Centre de Recherche en Histoire des Sciences et Techniques, <http://www.crhst.cnrs.fr>

Criminocorpus : <http://www.criminocorpus.cnrs.fr>

Ecole nationale des Chartes : <http://www.enc.sorbonne.fr>

Enluminures (Base) : <http://www.enluminures.culture.fr>

Enssib : <http://www.enssib.fr>

Europeana : <http://www.europeana.eu> (Et un historique :

http://www.bnf.fr/pages/europeana/bnue_hist.htm)

Hathi trust : <http://www.hathitrust.org>

Gallica: <http://www.gallica.bnf.fr>

Google Books : <http://books.google.com> (version française : <http://books.google.fr>);

réflexions par exemple sur <http://booksearch.blogspot.com>

Le Guichet du savoir, service de la bibliothèque municipale de Lyon, www.guichetdusavoir.org
HAL : <http://hal.archives-ouvertes.fr>
Jubilothèque (Paris 6) : <http://www.jubil.upmc.fr>
Lectura (bibliothèques municipales de Rhône-Alpes): <http://www.lectura.fr/fr/catalogues>
Lewatson.com : Congrès Iber, « learning center » sur Web, <http://www.lewatson.com>
Liber Floridus (Base) : <http://liberfloridus.cines.fr>
Mandragore (Base) : <http://mandragore.bnf.fr>
OAister : <http://www.oaister.org>
Open Content Alliance : <http://www.opencontentalliance.org>
Open library : <http://www.openlibrary.org>.
Rennes 1 (SCD) <http://www.univ-rennes1.fr/themes/cultureEtDocumentation>
Rennes 1, avec des ouvrages sous droits, <http://www.scd.univ-rennes1/ebook/>
Sudoc : <http://www.sudoc.abes.fr>
Telma, Centre de ressources numériques de l'École nationale des Chartes et de l'Institut de Recherche en Histoire des Textes, <http://www.cn-telma.fr>
The European Library, <http://www.theeuropeanlibrary.org/portal/idex.html>
Université Numérique en Région Bretagne, <http://unrb.univ-bretagne.fr>
Worldcat : <http://www.worldcat.org>
W3C: <http://www.w3c.org>.

Table des annexes

ANNEXE 1 : PRECISIONS SUR LES ISBD.....	105
ANNEXE 2 : PRECISIONS SUR LES FRBR	106
ANNEXE 3 : UNE HYPOTHESE D'AVENIR	107

Annexe 1 : précisions sur les ISBD

Eléments d'identification, répartis en 8 zones :

1- Titre +responsabilité.-2- Edition.-3- Document spécial (Cartographique, musique, Electronique).-4- Adresse bibliographique.-5- Collation (Description physique).-6- Collection (Edition) et titre de monographie.-7- Notes.-8- Numéro international normalisé (ISBN) +Prix

Structure des zones :

Sources d'information (Zones 1 à 4) ; Zones 5 à 7 : Document lui-même, Source extérieure ; Zone 8 : Verso de page de titre, Bas de dos de 4^e de couverture ; Une information tirée d'autres sources est transcrite entre crochets ([]). La ponctuation a un rôle structurant, pour l'ordre des éléments (zones séparées par des tirets, intérieur des zones structuré, avec « / » pour la première mention de responsabilité, « : » pour un complément, « [crochets] » pour information ajoutée par le catalogueur.

ISBD relatifs à des documents spécifiques adaptés par des normes AFNOR correspondantes :

ISBD(A) : Z 44-074 (livres anciens). ISBD(M) : Z 44-050, Z 44-073 (impr. modernes). ISBD(CR) : Z 44-063 (périodiques, collections). ISBD(ER) : Z 44-082 (ressources électroniques). ISBD(PM) : Z 44-069 (musique imprimée). ISBD(CM) : Z 44-067 (ressources cartographiques). ISBD(NBM) : Z 44-066 (enregistrements sonores), Z 44-065 (vidéogrammes), Z 44-077 (images fixes)

L'intégration de l'ISBD en UNIMARC :

Par exemple en Unimarc : « 200 » est Zone 1 d'ISBD avec en « 1# » la sous-zone « \$a » titre, « \$b » type ([texte imprimé] par exemple), ... « \$f » mention de responsabilité.

Les blocs fonctionnels de l'UNIMARC :

0xx Identification.-1xx Informations codées (101 : langue).-2xx Informations descriptives.-3xx Notes.-4xx Liens entre notices.-5xx Accès titres (points d'accès titres, titres parallèles).-6xx Accès matières.-7xx Accès de responsabilité.-8xx Usage international (Par exemple 856 : lien à ressources électroniques).-9xx Usage national ou local (Liberté de chaque pays).

Annexe 2 : Précisions sur les FRBR

Exemple de modélisation :

Ce mémoire d'étude est une œuvre dont la version remise en janvier 2009 est une expression ; le tirage papier en Français en est une manifestation dont vous lisez un item. Ainsi, un autre tirage papier sera un autre item, la traduction en Breton sera une autre manifestation, et l'hypothétique adaptation cinématographique une autre expression, mais il s'agira toujours de la même œuvre. Quant à sa version numérisée, c'est une autre manifestation de la même expression.

Il faudra donc une notice pour chaque expression.

Exemples d'applications :

L'OCLC travaille sur un catalogue en FRBR, avec la même notice pour les diverses éditions d'une œuvre. Ce degré le plus simple de l'application des FRBR est déjà impossible en UNIMARC. L'application complète du modèle est donc encore loin d'être possible. On peut pourtant mentionner des tentatives comme VTLS, un système Américain d'affichage du FRBR basé sur le Marc 21, et acheté notamment par Paris VII, ce qui pose problème avec l'UNIMARC.

Annexe 3 : Une hypothèse d'avenir

Figoblog, <http://www.figoblog.org/node/1945>, propose une hypothèse...

« La première période (jusqu'aux années 1990) se caractérise par une approche d'informatisation pragmatique : on a mis l'accent sur la structure (MARC) au détriment des modèles. La 2e période fait l'objet d'un effort de modélisation conceptuelle (FRBR). La structure MARC évolue vers XML, tandis que le Dublin Core se développe en parallèle (à la fois pour la structure, le modèle et le format). Dans la 3e période, celle qui est devant nous, le modèle conceptuel reste valable. Le Dublin Core converge avec les structures et des bibliothèques (RDA vocabulary) et peut être exprimé en RDF. RDA remplace les anciennes règles de catalogage. L'utilisation de RDF rend inutile le besoin d'un format pour les notices (on exprime directement les données). Au niveau de l'encodage, le passage par XML facilite la conversion vers RDF. »