

Vers une révolution dans la conception des catalogues... et bien au-delà ?

À propos de la concertation sur l'information bibliographique enrichie www.abf.asso.fr/enrichi

par Pierre-Yves Duchemin* et Dominique Lahary**

Naissance d'une concertation

Au printemps 1999, quelques conversations informelles entre bibliothécaires et fournisseurs de données bibliographiques ont tourné autour de ce constat : parmi les seconds, certains sont prêts à fournir des images de couverture et des tables des matières. Des bibliothèques ont anticipé le mouvement, elles-mêmes ou avec le concours de fournisseurs de systèmes, en donnant accès à de tels éléments dans leurs catalogues en ligne. Mais nous ne savons pas comment échanger ces données. Y avait-il là matière à délibérer ?

Telle fut la question à laquelle étaient invités à répondre les bibliothécaires, fournisseurs de données et fournisseurs de systèmes présents au congrès de l'ABF de La Rochelle en mai 1999, au cours d'une réunion de deux heures. La réponse fut positive, et rendez-vous fut pris pour une séance d'une journée en octobre 1999.

Or cette séance, absolument décisive, a lancé un processus qui non seulement n'a cessé depuis de s'approfondir et d'étendre ses objectifs, mais qui pose des questions dépassant de loin la question d'origine. Elle a placé ses participants au cœur de ce qui est peut-être une révolution culturelle naissante, et qui pourrait bien remettre en cause tout l'édifice normatif sur lequel s'était construite la bibliothéconomie informatisée.

L'objet du présent article est donc de donner un aperçu du travail de la « concertation pour une information bibliographique enrichie » (en abrégé, Enrichi), mais aussi de présenter quelques-uns des aspects de

cette révolution que la concertation a rencontrée et à laquelle, indubitablement, elle est amenée à participer.

La commande initiale

Elle consistait à examiner comment enrichir un catalogue de données que l'on peut classer en trois types :

- extraits du document (par exemple, image de la couverture d'un livre ou du boîtier d'un disque compact, quatrième de couverture, table des matières) ;
- ressources associées ;
- document lui-même, s'il s'agit d'un document électronique accessible en ligne.

Quant à l'usage de ces données, il se décline comme suit :

- affichage, soit en même temps que la notice, soit par lien hypertexte à partir de celle-ci ;
- recherche, pour autant qu'il s'agisse de texte ou de données non textuelles mais décrites et/ou indexées (des images par exemple).

De tels besoins font apparaître des problèmes que la tradition catalographique n'avait jamais posés. Comment nommer les ressources associées ? Quelle typologie définir (table des matières, couverture...) ? Comment décrire techniquement les fichiers joints ?

En matière d'échange, se pose la question de la fourniture conjointe de données associées figurant dans plusieurs fichiers. On sait que, dans le cadre des échanges en format MARC, on n'a jamais résolu la question de la fourniture de notices liées entre elles : on n'importe jamais que des notices isolées, certains fournisseurs concentrant dans une même notice des éléments issus de plusieurs afin que le système importateur procède éventuellement à leur éclatement¹.

* Bibliothèque nationale de France

** Bibliothèque départementale du Val d'Oise

Deux niveaux, trois groupes

Il est rapidement apparu que le cadre normatif actuel des échanges bibliographiques ne permettait pas de résoudre de façon satisfaisante la plupart de ces besoins. Un autre cadre est apparu nécessaire, et il fut nommé avec insistance au cours de la séance du 19 octobre : XML. Cet intrus sera présenté plus bas. Il apparut à tous comme une perspective sérieuse, mais à moyen terme.

C'est pourquoi il fut décidé d'étudier l'éventualité de mesures immédiates, dans le cadre du format d'échange UNIMARC et de la norme ISO 2709 auquel il est conforme. Ont été mis en place trois groupes de travail qui se réunissent indépendamment. Des réunions plénières font périodiquement le point sur l'ensemble des travaux, et des séances thématiques transversales sont également envisagées.

- **Le groupe UNIMARC** (coordination Françoise Bourdon) avait pour mission de rechercher et de proposer des solutions pour la gestion de l'information bibliographique enrichie dans le cadre normatif actuel, notamment pour l'échange de données. Cet objectif a été confié au comité français UNIMARC, dont la constitution a été décidée le 31 mars 2000 lors d'une réunion du groupe.

- **Le groupe XML** (coordination Pierre-Yves Duchemin) a pour mission d'explorer les possibilités de cette structure de format. Il s'agit d'une démarche à moyen terme qui passe par une phase de familiarisation et d'apprentissage.

- **Le groupe Signalement partagé des ressources en ligne** a pour mission de favoriser la coopération entre bibliothèques dans la description et l'organisation de l'accès aux ressources d'internet, mais aussi d'explorer les techniques de signalement de ces ressources (simples pages de liens, intégration dans un catalogue classique de bibliothèque, gestion par des bases de données spécialisées).

Du document physique au document électronique, et retour

Né d'un besoin lié aux documents physiques, notamment aux livres, Enrichi traite de plus en plus de documentation électronique. Outre que cet aspect n'était dès le départ nullement exclu, il y a une bonne raison à cela. Le développement de l'information numérique en réseau oblige à inventer de nouvelles façons de décrire des ressources et d'y accéder. Cela constitue incontestablement le secteur de pointe et, dans un premier temps, les solutions divergent.

Des participants ont expliqué par exemple qu'il était inutile de décrire des ressources électroniques dans un

catalogue en format MARC. Même si ce principe souffre d'utiles exceptions, il montre qu'une nouvelle bibliothéconomie est en train de naître. Dans un premier temps, cela se traduit par une séparation au moins partielle entre le traitement des ressources physiques et celui des ressources électroniques.

L'hypothèse qui justifie le travail du groupe est que cette dichotomie est provisoire, et que le traitement des ressources classiques bénéficiera tôt ou tard des acquis du traitement de l'information électronique.

Les limites des formats MARC

C'est en butant sur les limites imposées par les formats MARC et la norme ISO 2709 qu'Enrichi s'est tourné vers XML. Il convient de revenir sur les limites d'un cadre qui, rappelons-le, a été conçu dans les années 1960.

Le cadre conceptuel

Les formats MARC ont pour fondement les concepts bibliographiques traditionnels tels qu'ils sont formalisés dans les ISBD. Ils sont à la fois représentatifs de l'âge pré-informatique (description textuelle concise, accès limités) et d'une gestion informatique que l'on dira aujourd'hui primitive, réduite au mode texte et au traitement de l'information secondaire.

On peut certes introduire des notions nouvelles, ainsi que le fait la zone 856 (voir tableau ci-joint) qui gère le lien à une ressource électronique. Mais on ne saurait indéfiniment alourdir le format de notions pour lesquelles il n'a pas été conçu.

Le mode de gestion des formats

Les formats MARC sont gérés par des bibliothécaires pour les bibliothécaires. Ils n'ont jamais réussi de vraie percée en dehors de ce monde. Ils ont donc laissé au-dehors les autres professions du livre et de l'information, et même une partie des bibliothèques et des centres de documentation. Dans un monde ouvert, le monde MARC est un ensemble clos.

En outre, sans justification aucune, il s'est fractionné en sous-ensembles nationaux ou internationaux qui, s'ils n'empêchent pas les échanges par conversion, les compliquent et les renchérissent.

Enfin, une partie des formats MARC évolue lentement. C'est le cas d'UNIMARC, administré par un comité international, le PUC (Permanent UNIMARC Committee de l'IFLA), qui ne se réunit qu'une ou deux fois par an.

Le site Enrichi www.abf.asso.fr/enrichi

Hébergé sur le site Web de l'ABF, le site Enrichi constitue la plate-forme de travail en ligne des participants à la « concertation pour une initiative enrichie ». Vous y trouverez les convocations aux réunions et le compte rendu de celles-ci, ainsi que des documents de travail et des liens vers des sites et ressources accessibles en ligne, comme les liens proposés par le groupe XML (<http://www.abf.asso.fr/enrichi/xmldoc.htm>) qui constituent une invitation à l'autoformation.

Complété par l'usage de la messagerie, ce site permet à Enrichi de fonctionner sans moyens financiers, sans secrétariat, sans papier à en-tête ni timbres-poste. Mais il permet également à quiconque y trouve intérêt de suivre les travaux qui se déroulent, en toute transparence, sous les yeux de tous.

La structure

Nous touchons là à l'essentiel : la norme ISO 2709, à laquelle tous les formats MARC sont conformes, n'organise qu'une arborescence à deux niveaux (la zone, la sous-zone), éventuellement augmentée d'indicateurs. Cette limitation structurelle est une contrainte majeure. Il n'y a aucune raison fonctionnelle de se limiter à deux niveaux, et les formats émergents ne présentent pas cette limite.

Les promesses de XML

- XML (eXtensible Markup Language = langage de balisage extensible ou eXtended Markup Language = langage de balisage étendu) est le résultat des réflexions d'un groupe de travail du World Wide Web Consortium (W3C) de 1996 à 1998.

XML est un sous-ensemble de la norme SGML qui a été conçu dans le dessein de concevoir un nouveau langage d'utilisation moins lourde que celle de la norme SGML générique, mais aussi simple que HTML et offrant une puissance supérieure capable de décrire des bases de données, des documents volumineux ou à la structure complexe et multimédia. XML a également été conçu pour être interopérable avec SGML et HTML, et accroître les fonctionnalités du World Wide Web. XML est ainsi à la croisée de deux univers : d'une part, le monde documentaire qui a produit la norme SGML ; d'autre part, le monde d'internet et du World Wide Web qui a produit HTML.

En absorbant la substantifique moelle et en gommant les défauts de ses deux prédécesseurs, XML est ainsi « presque parfait » dès sa naissance : il simplifie la structuration électronique de la sémantique du document électronique de SGML, conserve la séparation établie par SGML entre le fond et la forme, entre le contenu et la présentation, et emprunte à HTML sa facilité d'emploi, sa possibilité d'utilisation de liens hypertextuels et du multimédia, et son interopérabilité sur internet.

- SGML (Standard Generalized Markup Language = langage normalisé de balisage généralisé) a été créé au cours des années 1980 dans le monde de l'édition électronique. C'est à la fois une technique de production électronique de documents papier et un outil qui permet la saisie de documents électroniques en bases de données structurées grâce à des logiciels éditeurs. SGML est devenu la norme ISO 8879, c'est-à-dire la norme universelle de production de documents électroniques, et a trouvé sa place dans la production d'applications documentaires.

SGML définit la structure logique d'un document électronique (en-têtes, titres, chapitres, paragraphes, notes, etc.) et fonctionne grâce à des DTD (Document Type Definitions = descriptions de type de document) qui définissent les **éléments** à coder avec leur syntaxe d'utilisation (obligation, répétitivité, etc.), des balises qui identifient ces éléments ou ces groupes d'éléments, des logiciels éditeurs qui permettent de poser les balises et des *parsers* ou analyseurs qui permettent de vérifier la conformité du document à la DTD. SGML s'est surtout développé dans le monde de la production de documents électroniques et n'a eu que quelques applications strictement documentaires dans les bibliothèques ou les centres de documentation.

- HTML (HyperText Markup Language = langage de balisage hypertextuel) est en quelque sorte une DTD SGML simplifiée, utilisée pour concevoir des pages Web : titres, textes, listes, points d'ancrage hypertextuels, multimédia, fonctions de navigation à l'intérieur d'un document ou vers un autre document sur le même site, ou encore vers un autre document sur un autre site grâce à l'utilisation de pointeurs spécifiques, les URL (Uniform Resource Locators), qui sont les adresses des sites accessibles sur internet.

HTML définit plus la **structure physique** (retraits, marges, justification, couleurs, corps, conventions typographiques, etc.) que la **structure logique** du document électronique (codée par un ensemble de balises prédéfinies qui ne donnent pas d'informations sur ce qu'elles contiennent, mais peuvent être comprises par tous les

Hommage à un visionnaire

La scène se situe au début des années 1980, durant les épreuves orales de l'examen professionnel d'employé de bibliothèque organisé par le syndicat de communes pour le personnel de la Petite Couronne d'Île-de-France. Une des questions à tirer au sort porte sur « Les catalogues en bibliothèque ».

Un candidat répond à cette question en décrivant une sorte de brochure illustrée, présentant les photographies des livres avec une courte description : telle était son idée du catalogue d'une bibliothèque. Dès qu'il eut tourné les talons, consternation à la fois apitoyée et amusée des deux examinateurs, dont j'étais : le pauvre nous a décrit le catalogue de La Redoute ! Nous lui avons attribué une mauvaise note.

Je constate aujourd'hui que ma condescendance était bien mal venue : son idée du catalogue, c'est très exactement celle que nous sommes en train de mettre en œuvre. C'est le profane qui avait raison, et les professionnels qui avaient tort. Pouvons-nous entendre cette leçon ?

Dominique Lahary

navigateurs). HTML est devenu le standard de fait pour toutes les publications sur internet, mais montre ses limites dès qu'il s'agit de traiter un document volumineux : les *frames* ne sont rien d'autre qu'une tentative palliative pour arriver à gérer des documents HTML de grande taille.

Qu'est-ce que XML ? XML n'est pas un format ; XML n'est pas un logiciel ; XML n'est pas un langage de programmation ; XML n'est pas un protocole d'échange ; XML est un peu tout cela à la fois, c'est aussi une infrastructure, une approche, voire une méthodologie. C'est un « métaformat », c'est-à-dire une structure générique qui a été conçue pour être transmise, lue et traitée sur le World Wide Web. C'est également un « métalangage », c'est-à-dire un ensemble de balises utilisées pour coder des documents ou des parties de documents, et qui permet de traiter le document électronique indépendamment des outils utilisés pour sa mise en forme.

XML reprend les grands principes de SGML, c'est-à-dire le concept de balisage structural, les balises et les DTD. Les DTD SGML utilisées dans le monde des bibliothèques ont été traduites en XML : ISO 12083 pour les monographies, les périodiques, les articles de périodiques et les formules mathématiques, TEI (Text Encoding Initiative) et EAD (Encoded Archival Description), traduite en décembre 1999 (version 1.1) et bien adaptée au traitement des fonds d'archives et des documents dits « spécialisés ».

D'autres DTD existent : la bibliothèque du Congrès a écrit une DTD permettant le balisage structural de notices bibliographiques rédigées en format USMARC ; et la Mission de la recherche et de la technologie au ministère de la Culture a écrit et est en train d'implémenter la DTD BiblioML, qui permet le balisage structural de notices bibliographiques rédigées en format UNIMARC. D'autres DTD permettant de traiter les documents sonores, l'image fixe et l'image animée restent à créer.

Alors que le langage HTML, DTD simplifiée et hypertextuelle de SGML, met l'accent sur la présentation plus que sur la structure, XML corrige ce défaut et ouvre d'autres possibilités, notamment la compatibilité avec les métadonnées et le Dublin Core². XML permet, comme SGML, de produire des documents électroniques structurés pouvant être édités sur des supports divers (livre, disque optique compact, DVD ou Web) ; il permet également, dans un même enregistrement, de juxtaposer des données bibliographiques, l'image numérique de la couverture et une structuration du sommaire interrogeable en texte intégral et qui respecte l'arborescence du document original, ainsi que des liens hypertextes et hypermédias et des indexations liées à des fichiers d'autorité ou des thesauri pour gérer les accès.

En ce qui concerne la codification des caractères, y compris des caractères non latins, XML est basé sur le système UNICODE UCS-4 (norme ISO 10646), qui permet le codage des caractères sur 4 octets et offre ainsi près de 4,3 milliards de possibilités.

XML autorise l'utilisation simultanée de plusieurs DTD dans le même document (document « valide »), rend facultatif l'emploi d'une DTD, en permettant aussi de produire des documents non rattachés à une DTD (document « bien formé ») et comportant eux-mêmes les informations nécessaires pour interpréter ses balises et comprendre sa structure : on parle alors de documents « autodocumentés ». Offrant une arborescence *a priori* illimitée, XML permet le traitement « à niveaux » dans une perspective à la fois plus simple et plus structurée que ce qui était possible auparavant : c'est la fin du fameux syndrome « Carmen-Mozart³ » puisque chaque « notice analytique » de niveau inférieur est une véritable notice et non plus un fragment de la notice principale.

La structure d'un document XML se présente comme une suite d'éléments enchâssés les uns dans les autres : chaque élément possède une balise de début (<balise >) et une balise de fin (</balise >). Afin de mieux visualiser la structure du document, les éléments sont présentés avec des indentations successives :

```

<biblio>
  <auteur>
    <nom> Martin </nom> <prénom> Jean </pré-
    nom> </auteur>
  <titre>
    <titrepropre> XML </titrepropre>
    <complémenttitre> à quoi ça sert ? </complé-
    menttitre </titre>
</biblio>

```

À la différence d'un document HTML, un document XML ne contient pas l'ensemble des balises qui conditionnent sa présentation : ainsi, par exemple, la mise en page ou la police de caractères utilisées ne sont pas spécifiées explicitement dans le document source, pas plus que d'éventuelles illustrations présentes dans le texte. XML permet d'utiliser, à côté des balises marquant la structure du document textuel lui-même, un autre document, la « feuille de style » normalisée XSL (eXtended Stylesheet Language), adaptation de la norme ISO 10179 sur les feuilles de style de SGML qui spécifie toutes les données absentes du texte lui-même, à laquelle on peut adjoindre l'adresse logique d'un fichier graphique.

Grâce à cette structuration « éclatée », qui sépare d'une part la description sémantique de la structure du document et d'autre part les attributs physiques permettant leur réalisation et leur affichage, il est possible de spécifier des feuilles de style qui définissent les éléments, les attributs, les contenus des éléments et les valeurs des attributs nécessaires au traitement du document, ainsi que d'accroître les possibilités d'échanges de données. On peut donc réaliser n'importe quel type de sortie : une impression sur papier ou une sortie en braille, un affichage sur écran en intranet ou sur le Web, une édition électronique, voire une lecture par logiciel de synthèse vocale.

XML offre la possibilité de décrire les droits de propriété intellectuelle d'un document électronique ou les droits d'accès à des pages Web. Il permet également le codage d'une signature numérique qui valide et protège les documents électroniques placés sur internet. Indépendant comme SGML de tout logiciel, XML offre une pérennité à long terme des documents structurés, et son formalisme à la fois simple et puissant facilite le traitement automatique tout en permettant l'échange de données entre des systèmes informatiques différents. Dans quelques dizaines d'années, les technologies ayant évolué, il ne sera plus guère possible de lire sans transcodages compliqués et générateurs de pertes d'informations des documents HTML remplis de balises et d'attributs en tous genres. Les documents XML, libres de tout attribut de mise en page, devraient normalement

être toujours lisibles puisque le langage XML n'autorise aucun caractère non affichable à l'écran.

XML en bibliothèque

L'apport essentiel de XML est qu'il permet au document électronique d'échapper à la fois aux contraintes d'un support quelconque et à celles des logiciels nécessaires pour le créer, le rechercher, le visualiser, l'imprimer, l'éditer.

Il est à prévoir que XML soit reconnu rapidement au niveau international comme norme ISO. Il est clair que XML est appelé à un développement important dans le monde des bibliothèques. 1998, date de naissance de XML, sera une date aussi importante que 1964, date de naissance du premier format MARC.

XML est l'instrument de la révolution qui va secouer les bibliothèques dans les années à venir, bien qu'il ne soit pas sur le même plan que les formats MARC, et plutôt sur celui de la norme ISO 2709. En effet, XML ne propose pas de définition des éléments bibliographiques (pour lesquels nos bons vieux ISBD demeurent une référence utile) mais propose une structuration de ces données. L'ère de l'« après-MARC » approche et, au bout de trente-cinq années de bons et loyaux services, les données stockées ou échangées en formats MARC pourraient être converties dans des structures XML.

Cela ne signifie pas que toutes les notices MARC seront à jeter aux orties dans un avenir proche ! Les formats MARC vont continuer à jouer leur rôle pendant un certain temps comme format d'échange, notamment dans le cadre de la norme ISO 2709, au même titre qu'une page Web ou une édition imprimée, tandis que, à l'inverse, des bases structurées en référence à un format MARC pourront importer et exporter des données structurées en XML : le format d'échange ISO 2709 n'étant pas directement utilisable sur internet, il est possible d'envisager à l'avenir la conversion de notices MARC en XML, via une structuration en RDF. Le format RDF (Resource Description Framework) est une syntaxe particulière qui permet l'implémentation de XML et qui permet également la structuration des métadonnées sous forme de « propriété d'objets ».

Les formats MARC distinguent deux aspects : l'aspect sémantique (données, zones et sous-zones, indicateurs, etc.) et l'aspect logique (étiquettes, codage des zones et sous-zones, valeur des indicateurs, caractères de contrôle, etc.). Rien n'empêche *a priori* de conserver l'aspect sémantique des formats MARC et d'utiliser la struc-

À lire, à consulter en ligne

L'ouvrage essentiel à consulter est :

ALAIN MICHARD : **XML : langage et applications**. Paris, Eyrolles, 1999. ISBN 2-212-09052-8.

Pour s'informer en ligne, voici un extrait de la page de liens proposée par le groupe XML d'Enrichi : <http://www.abf.asso.fr/enrichi/xmldoc.htm>

Le standard

- Extensible Markup Language (XML) 1.0, original en anglais : <http://www.w3.org/TR/REC-xml>
- Traduction française : http://babel.alis.com/web_ml/xml

Présentations de XML

- Pages XML du World Wide Web Consortium : <http://www.w3.org>
- The SGML/XML Web Page : <http://www.openoasis.org/cover/sgml-xml.html>
- Initiez-vous au XML/par Arnaud Petitjean : <http://www.citeweb.net/apetitje/xml>
- XML/par Christophe Reubrez : <http://www.chez.com/xml>
- Introduction au XML/Guy Teasdale in : La Lettre du bibliothécaire québécois :

<http://www.sciencepresse.qc.ca/lbq/lbq10.4.html>

DTD XML et autres applications normalisées de XML

- Norme ISO 12083 : <http://www.xmlxperts.com/12083xml.htm>
- EAD (Encoded Archival Description) : <http://lcweb.loc.gov/ead>
- TEI (Text Encoding Initiative) : <http://www.uic.edu/orgs/tei>
- Norme GED 1.1 de l'APROGED : <http://www.aproged.org/Norme/NORME1.HTM>
- DTD UNIMARC XML du ministère de la Culture : www.culture.gouv.fr/BiblioML et aussi :
- La DTD MARC de la bibliothèque du Congrès : <http://lcweb.loc.gov/marc/marcdtd/marcdtd-back.html>

Voir aussi :

- Les liens XML : <http://www.chez.com/xml/liens/liens.htm>
- Bath Profile de Z 39.50 : http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/bath
- XML and XML/MARC Related Links : <http://xmlmarc.stanford.edu/webliography.xhtml>

ture logique de XML : après tout, quelle importance y a-t-il à baptiser un auteur 700\$a ou < auteur >, pourvu que, d'une manière ou d'une autre, l'auteur puisse être identifié comme tel ? Du point de vue du catalogueur ou de l'utilisateur du catalogue, cela ne change strictement rien au problème !

Ainsi, le format MARC, avec son formalisme (étiquettes, zones obligatoires, zones répétables, etc.), devient une DTD... c'est ce qu'ont fait la bibliothèque du Congrès et le ministère de la Culture. Une telle DTD, utilisée avec une feuille de style XSL (zones affichables, ponctuation restituée, libellé en clair des zones, etc.), aurait l'avantage de permettre une plus grande standardisation : en effet, ces standards communs pourraient être **interprétés** par des systèmes différents au lieu d'être, comme aujourd'hui, **paramétrés** différemment pour chacun de ces systèmes. Par ailleurs, les querelles stériles entre INTERMARC, UNIMARC, USMARC, ETCMARC appartiendraient à un passé révolu. Enfin, toute publication sous forme électronique ou papier ne serait qu'une application particulière de la feuille de style, ce qui simplifierait grandement le paramétrage des éditions.

Certains systèmes intégrés de gestion de bibliothèque ou de centre de documentation évoluent actuellement vers l'intégration et la gestion des procédures XML, déjà utilisées par des fournisseurs d'intergiciels⁴ répandus dans le monde des bibliothèques. Le développement de convertisseurs MARC-ISO 2709 vers des convertisseurs MARC-XML est l'amorce d'une phase transitoire. En France, l'APROGED (Association des professionnels de la gestion électronique de documents) a mis au point la norme GED 1.1, qui est une norme d'échange basée sur XML et qui peut gérer texte et images grâce à un fichier placé dans un dossier faisant référence au document primaire. Cette norme est actuellement en cours d'implémentation.

On sait que la norme Z 39.50 (ISO 239.50) permet qu'un serveur fournisse des données bibliographiques en format MARC à un logiciel client, qui peut éventuellement générer un affichage en format HTML. Or le Bath Profile, qui vise à simplifier l'application de cette norme pour en permettre une plus large utilisation, prévoit que ces données puissent, en réponse à une interrogation, être fournies en XML.

Appelons cela un péché de jeunesse : il n'existe encore aucun outil de production d'information bibliographique ou documentaire en XML. Nul doute que ces outils apparaîtront bientôt sur le marché. En attendant, certains établissements précurseurs « bricolent » des solutions grâce à des macros Word afin de produire directement des documents électroniques en XML sans passer par l'étape MARC, toujours obligatoire dans le monde des bibliothèques.

L'utilisation de XML dans les bibliothèques peut être envisagée comme format de stockage des données numériques, puisque c'est un format « générique » qui peut s'appliquer à toutes sortes de documents, comme format d'échange en importation et comme format d'échange en exportation, l'échange dans ce dernier cas pouvant aller jusqu'à la mise en ligne de documents électroniques sur des pages Web. Comme format de stockage, XML apporte une solution puissante et souple de structuration indépendante de toute édition et exportation, permettant des arborescences profondes et le lien à des fichiers externes multimédias. Comme format d'échange, XML offre des possibilités plus nombreuses de liens à des fichiers externes, notamment des données multimédias, ainsi qu'une structure plus riche que la norme ISO 2709, laquelle ne permet qu'une arborescence à deux niveaux (zones, sous-zones) avec étiquettes et indicateurs.

Mais il doit être bien clair que XML n'est jamais qu'une structure générale, au même titre que la norme ISO 2709, et ne dit rien des éléments... lesquels, en matière d'information secondaire, sont notamment définis par les ISBD ! C'est pourquoi devront être poursuivis les efforts de définition de DTD adaptés aux données que manipulent les bibliothèques, les centres de documentation et les autres professions du livre.

Leçons pour une histoire qui ne fait que commencer

Ce n'est sans doute pas un hasard si Enrichi est né en 1999, quelque cinq ans après l'explosion d'internet, dans une période à la fois de remise en cause et de construction d'un nouveau paysage de l'information. Nous esquissons ci-dessous quelques traits essentiels de ce nouveau paysage dont Enrichi est devenu, avec modestie et ambition, un acteur.

Une nouvelle conception des catalogues

Les catalogues de bibliothèque évoluent :

- dans leur aspect, la figure du Web tendant à une hégémonie irréversible, sur internet comme en intranet, et apportant deux fonctionnalités essentielles : l'image

et le lien (lien vers une ressource affichable ou téléchargeable, lien relançant une recherche) ;

- dans leur nature, avec l'apparition de nouvelles notions qui lui étaient jusque-là étrangères (gestion des droits, images de couverture, tables des matières) ;

- dans leur fonction puisqu'ils permettent désormais d'avoir accès au document ou au moins à des extraits significatifs de celui-ci.

Des normes pour bibliothécaires aux standards communs

Dans un monde de l'information et de la communication qui se construit rapidement sur des standards communs, il est permis de se demander quel avenir ont des formats compris et pratiqués par une seule profession. Il apparaît désormais probable que ces standards communs prendront progressivement le pas.

Le rôle des bibliothécaires est alors de se préparer à les comprendre, à les adopter, à les enrichir, et finalement à y transférer tout l'acquis de leur propre normalisation. Il ne s'agit pas de jeter le bébé avec l'eau du bain. Bien au contraire, il leur revient de dégager les éléments fondamentaux qu'ils ont su définir de leur gangue normative, de sauver les concepts au prix du sacrifice du code.

Les ISBD mêlaient indissolublement les concepts et les codes, la structure logique et la présentation. Les formats MARC ont constitué une première dissociation de la structure générale (la norme ISO 2709) et des éléments, énumérés dans chacun d'eux. Si XML est appelé à prendre la relève, ce n'est que comme structure générale, les éléments étant définis dans des DTD.

Structure	ISBD	ISO 2709	XML
Éléments		MARC	DTD

Remarquons enfin que les standards émergents sont pour une part gérés d'une façon nouvelle. L'ISO (International Standard Organization) joue certes toujours un rôle, tandis qu'on peut s'interroger sur l'AFNOR, ou du moins sur la pertinence de normes françaises dans un contexte évident de globalisation. Mais nombre de standards sont aujourd'hui gérés par des groupes à la fois internationaux, informels et mixtes, composés de représentants de sociétés privées et d'institutions publiques.

C'est le cas par exemple du W3C (World Wide Web Consortium), qui définit un grand nombre de standards d'internet, dont HTML et XML. C'est aussi celui du ZIG (Z 39.50 Implementor's Group). Malgré son caractère national, ou du moins francophone, Enrichi, dans sa constitution et son mode de fonctionnement, est conforme à ce nouveau contexte.

Sources :

USMARC :

- <http://www.loc.gov/marc/bibliographic/ecbdhold.html#mrcb856>
(avec commentaire sur chaque valeur ou sous-champ)

UNIMARC :

- <http://ifla.inist.fr/VI/3/p1996-1/concise.htm> (en anglais, sans commentaire)
- <http://www.bnf.fr/web-bnf/infopro/normes/unimarc2.htm> (en français, document RTF à télécharger)
- Manuel UNIMARC version française 3^e éd. Saur, 1999

UNIMARC	USMARC
Indicator 1: Access Method	Indicator 1: Access method
# - No information provided	# - No information provided
0 - Email	0 - Email
1 - FTP	1 - FTP
2- Remote login (Telnet)	2- Remote login (Telnet)
3- Dial-up	3- Dial-up
4 - HTTP	4 - HTTP
7 - Method specified in subfield \$y	7 - Method specified in subfield \$2
Indicator 2: Blank (not defined)	Second - Relationship
	# - No information provided
	0 - Resource
	1 - Version of resource
	2- Related resource
	8 - No display constant generated
Subfield Codes	Subfield Codes
\$a Host name (Repeatable)	\$a - Host name (R)
\$b Access number (Repeatable)	\$b - Access number (R)
\$c Compression information (Repeatable)	\$c - Compression information (R)
\$d Path (Repeatable)	\$d - Path (R)
\$f Electronic name (Repeatable)	\$f - Electronic name (R)

L'échange, levier du changement

Quel sera le moteur du changement ? Il est permis de supposer que ce sera le format d'échange. Souvenons-nous. C'est par l'échange que les formats MARC se sont imposés en France, avec une prédominance d'UNIMARC, et ont fini par peser sur la structure des bases de don-

nées bibliographiques elles-mêmes. Ce pourrait bien être par l'échange que XML s'introduira progressivement dans le paysage avant d'influencer les bases elles-mêmes.

L'information bibliographique enrichie et le traitement des ressources électroniques auront alors joué le rôle de ver dans le fruit, ou de ferment dans la pâte.

\$g Uniform Resource Name (Repeatable)	\$g - Uniform Resource Name (R)
\$h Processor of request (Not repeatable)	\$h - Processor of request (NR)
\$i Instruction (Repeatable)	\$i - Instruction (R)
\$j Bits per second (Not repeatable)	\$j - Bits per second (NR)
\$k Password (Not repeatable)	\$k - Password (NR)
\$l Login/login (Not repeatable)	\$l - Logon (NR)
\$m Contact for access assistance (Repeatable)	\$m - Contact for access assistance (R)
\$n Name of location of host in subfield \$a (Not repeatable)	\$n - Name of location of host in subfield \$a (NR)
\$o Operating system (Not repeatable)	\$o - Operating system (NR)
\$p Port (Not repeatable)	\$p - Port (NR)
\$q Electronic Format Type (Not repeatable)	\$q - Electronic format type (NR)
\$r Settings (Not repeatable)	\$r - Settings (NR)
\$s File size (Repeatable)	\$s - File size (R)
\$t Terminal emulation (Repeatable)	\$t - Terminal emulation (R)
\$u Uniform Resource Locator (Repeatable)	\$u - Uniform Resource Locator (R)
\$v Hours access method available (Repeatable)	\$v - Hours access method available (R)
\$w Record control number (Repeatable)	\$w - Record control number (R)
\$x Nonpublic note (Repeatable)	\$x - Nonpublic note (R)
\$y Access method (Repeatable)	
\$z Public note (Repeatable)	\$z - Public note (R)
	\$2- Access method (NR)
	\$3- Materials specified (NR)
	\$6- Linkage (NR)

USMARC 856 \$2 est remplacé par **UNIMARC 856 \$y**

USMARC 856 \$3 est destiné à spécifier textuellement la nature de la relation codée dans le 2^e indicateur.

USMARC 856 \$6 existe dans tous les champs USMARC pour traiter d'éventuelles graphies différentes.

UNIMARC 856 ne concerne que le lien vers la ressource faisant l'objet de la notice alors que **USMARC 856** peut concerner une version de la ressource ou une ressource liée. Ces deux derniers liens sont traités en UNIMARC dans la nouvelle sous-zone **\$u** (URL) des zones du bloc **4XX**.

1 C'était notamment le cas de la BnF dans son cédérom DSAM qui, jusqu'à la fin de 1999, comportait des notices à niveaux issues de plusieurs notices liées.

2 Les métadonnées (ou *metadata*) sont des données qui renseignent sur la nature de certaines autres données, ce qui permet leur utilisation pertinente. Une métadonnée est une donnée qui en décrit une autre. Une notice bibliographique peut être considérée comme un ensemble de métadonnées. Les métadonnées sont normalement constituées à partir du document électronique lui-même : par exemple, la page de titre d'un document électronique fournit

l'auteur, le titre et l'éditeur de l'ouvrage.

3 Syndrome affectant les systèmes répondant « oui » à la question « Avez-vous *Carmen* de Mozart ? » s'il s'agit d'un document sonore comportant à la fois un extrait de *Carmen* de Bizet et d'un morceau quelconque de Mozart.

4 Un intergiciel, terme québécois qui désigne l'appellation française « logiciel intermédiaire » (JO du 16 mars 1999), est un logiciel qui, dans une application client/serveur, sert d'intermédiaire entre, par exemple, un SIGB et un système de GED (gestion électronique de documents).