

L'enrichissement des catalogues ? Et après ?

Fin 1998, bibliothécaires et fournisseurs de données bibliographiques ont à plusieurs reprises évoqué la question de la fourniture d'images de couverture, de quatrièmes de couverture et de tables des matières, voire de ressources électroniques accessibles en ligne ; quelques bibliothèques ont lancé le mouvement en donnant accès à des notices bibliographiques « enrichies » dans leur catalogue en ligne. Mais le problème crucial de l'échange des données n'était pas résolu pour autant. Quelle est la situation aujourd'hui ?

Pierre-Yves Duchemin

École nationale supérieure
des sciences de l'information
et des bibliothèques
duchemin@enssib.fr

L'idée continuant son chemin, le salon professionnel du congrès ABF de La Rochelle en 1999 fut l'occasion de réunir bibliothécaires et fournisseurs de systèmes intégrés de gestion de bibliothèque (SIGB), d'intergiciels¹, de notices, ce qui n'a pas manqué de faire dire à certain(e)s collègues que cette démarche n'était pas la bonne et qu'il fallait discuter entre bibliothécaires... Le concept était de réunir de façon informelle bibliothécaires, fournisseurs de données, éditeurs de logiciels et administrations. Au cours d'une réunion de plus de deux heures qui a connu une affluence inattendue, la discussion s'avéra si abondante et positive que la décision fut prise de poursuivre la réflexion lors d'une journée de travail en octobre 1999.

19 octobre 1999... Séance décisive, riche, passionnante, foisonnante qui a lancé des pistes qui, ultérieurement, n'ont cessé de s'approfondir et

poser des questions qui vont bien au-delà de la proposition initiale de l'enrichissement des catalogues et, insensiblement, remettent en cause toute l'architecture normative sur laquelle est construite la bibliothéconomie informatisée.

Puisque le besoin existait, il fallait battre le fer et c'est ainsi qu'est née la « Concertation pour une information bibliographique enrichie », rapidement abrégée par ses membres en « Enrichi ». La tâche était immense et les objectifs ambitieux : faire de la pédagogie dans la profession, notamment pour convaincre les bibliothèques de sortir de leur isolement normatif et technique, faire une évaluation de la normalisation en vigueur, étudier différents cas de figure, suivre de nombreuses pistes, promouvoir des solutions, tenter d'influer sur les applications, convaincre les administrations centrales...

Une réflexion s'imposait également autour d'une idée qui apparaissait de plus en plus évidente : à l'échelle du web, se posait la question de l'avenir de formats compris et pratiqués par une seule profession. Il fallait donc inciter les collègues à découvrir et comprendre les normes et standards extérieurs à la profession, les adopter, voire les enrichir en y

1. Intergiciel : terme créé par les Québécois pour désigner le « middleware », c'est-à-dire un « logiciel fonctionnant sur le modèle client/serveur autorisant des applications à communiquer les unes avec les autres, alors qu'elles n'étaient pas conçues jusque-là pour dialoguer ensemble. Il facilite ainsi l'accès à des données stockées dans des systèmes qui ne sont pas toujours compatibles ». Depuis le début des années 1990, l'intergiciel représente un marché nouveau dans le monde des bibliothèques.

* Tous mes remerciements à Dominique Lahary (Bibliothèque départementale du Val-d'Oise), Patrick Le Bœuf (Bibliothèque nationale de France) et Laurent Dervieu (Électre) pour leurs précieuses indications.

** Toutes les URL citées fonctionnaient le 15 avril 2005.

Analyste en informatique documentaire et titulaire d'une maîtrise d'anglais,

Pierre-Yves Duchemin est directeur des ressources documentaires à l'Enssib, après avoir exercé à la Bibliothèque nationale de France et à la Bibliothèque publique d'information. Auteur de *L'art d'informatiser une bibliothèque* (Éditions du Cercle de la librairie, 2^e éd., 2000), il a collaboré au *Catalogage : méthodes et pratiques* (id., 2^e éd., 2003) et à *Développer un fonds de référence en bibliothèque* (id., 2001).

transférant tout l'acquis de la normalisation bibliothéconomique et persuader les bibliothécaires que toute la culture de normalisation qui était la leur depuis des années devait se fondre dans une normalisation plus large dont ils n'étaient plus les éléments moteurs, en préservant les éléments fondamentaux, en sauvant les concepts et en sacrifiant le point espace tiret, voire le \$ et ses indicateurs.

Marc

Les limites imposées par les formats Marc et la norme ISO 2709 font qu'« Enrichi » s'est rapidement tourné vers XML. Était-ce l'attirance du chant des sirènes ? Ou bien celle des yeux de Chimène ? Pour quelles raisons le regard s'est-il ainsi porté hors du cadre conceptuel traditionnel des bibliothèques conçu au milieu des années 1960 ?

Les formats Marc présentent une description textuelle et des accès limités ; leur gestion informatique permet le traitement de l'information secondaire (description et indexation) et la gestion des exemplaires physiques, mais montre ses limites dès qu'il est question de document numérique. On peut certes introduire des notions nouvelles, ainsi que le font la zone 856, qui gère le lien à une ressource électronique, et les zones 327 et 359, conçues pour traiter des tables des matières, mais alourdir indéfiniment les formats Marc de notions pour lesquelles ils n'ont pas été conçus ne semble pas une solution d'avenir : qui parlait de numérisation, de document numérique ou d'échange

Les formats Marc et l'ISBD

Quelques jalons

- Rapport du groupe de travail présidé par Henriette Davidson Avram dont le titre est *Machine Readable Cataloguing*, 1964.
- Marc I, LC-Marc, 1965.
- Marc II, USMarc, 1968.
- ISO 2709¹, 1973, norme internationale d'échange de données bibliographiques adoptée par l'Afnor en 1987 sous l'indice NF ISO 47-300, version actuelle 1996.
- InterMarc, 1975.
- Unimarc, 1977, version actuelle 2002.
- Marc21, 1997, version actuelle 2004.
- InterMarc intégré, 1998, version actuelle 2002.

Les formats Marc sont construits sur les concepts bibliographiques traditionnels formalisés dans les ISBD :

– *Instruction pour procéder à la confection du catalogue de chacune des bibliothèques sur lesquelles les directeurs ont dû ou doivent incessamment apposer des scellés*, 1791. Ce texte, peut-être l'aïeul des normes de catalogage, préconise le classement par auteur et la rédaction des notices par des bibliographes.

– *Principes de Paris sur les accès aux notices bibliographiques*, sous l'égide de l'Ifla, 1961 (les *Principes de Paris* n'ont jamais été révisés) ;

– ISBD², 1971, publiés par l'Ifla, régulièrement révisés et déclinés selon les types et supports de documents, jusqu'à l'ISBD (CR), 2002, qui est une révision de l'ISBD (S) pour les publications en série et dont le champ d'action s'étend désormais aux « ressources continues », c'est-à-dire aux périodiques électroniques accessibles en ligne et aux sites web. Un groupe de travail réfléchit actuellement à un ISBD « intégré ».

– Les normes et fascicules de documentation publiés par l'Afnor, adaptations françaises des ISBD.

– AACR, 1967, AACR2r³, 1978, version actuelle 2002, règles de catalogage anglo-américaines, en cours de révision (AACR3), Rica en Italie, Rak en Allemagne, Gost en ex-URSS, etc., toutes adaptations nationales des ISBD.

1. ISO 2709 : <http://www.iso.org/cate:d7675.html>

2. Liste en français des ISBD : <http://www.ifla.org/VI/3/nd1/isbdfren.htm>

3. Anglo-American Cataloguing Rules (AACR) : <http://www.aacr2.org/>

de données primaires au milieu des années 1960 quand la carte perforée à 12 lignes et 80 colonnes était encore en utilisation ?

En outre, les formats Marc sont gérés par des bibliothécaires pour des bibliothécaires et c'est peut-être là leur plus grave défaut : en quarante ans d'existence, ils n'ont jamais réussi à convaincre archivistes, documentalistes, conservateurs de musée, éditeurs et libraires, ni qui que ce soit qui traite de l'information. Dans un monde de la documentation ouvert sur le web, le monde Marc est un royaume oublié régi par les ISBD, les formats Marc, les normes ISO 2709 et ISO 239-50, alias ANSI Z 39.50, ensemble qui constitue une législation certes cohérente, fiable, indispensable, puissante, mais refermée sur elle-même.

Enfin, la norme ISO 2709, à laquelle tous les formats Marc doivent se conformer, offre une arborescence à deux niveaux : zone, sous-zone. Cette limitation structurelle est aujourd'hui une contrainte majeure. Une autre contrainte est la difficulté de la norme ISO 2709 à gérer les sous-notices imbriquées dans une notice principale, fonctionnalité pourtant offerte par Unimarc. Pas directement utilisable sur le web, ISO 2709, comme les formats Marc, est « réservée » aux bibliothécaires et son existence ignorée des autres acteurs du monde documentaire.

Les langages à balises

Ils sont nés hors du monde des bibliothèques, eux aussi au milieu des années 1960 :

- ML (Markup Language), 1967 ;
- GML (Generalized Markup Language), 1969 ;
- SGML² (Standardized Generalized Markup Language), 1980, devenue norme ISO 8879 en 1986 et Afnor NF Z 71-010 en 1990, référence internationale en matière d'édition électronique ;

2. W3C SGML : <http://www.w3.org/MarkUp/SGML/>

- HTML (HyperText Markup Language), 1992, devenu XHTML³ en 2000 ;
- XML⁴ (eXtensible Markup Language), 1998, version 1.1 en 2004.

C'est à SGML que l'on doit la création des DTD (Définition de Type de Document = Document Type Definition). Qu'est-ce qu'une DTD ? Une DTD est un « format », c'est-à-dire un modèle de document permettant d'exprimer la description plus ou moins complexe de la structure d'un type de document en précisant l'intitulé des balises, leur contenu, leur caractère obligatoire ou non, leur ordre de succession, leur imbrication, les relations existant entre différents éléments, etc., ce qui permet d'exprimer la structure arborescente du document.

C'est également aux langages à balises que l'on doit la naissance du concept de « métadonnées ». Que sont les métadonnées ? Tout simplement, des données sur les données... c'est-à-dire de l'information secondaire, donc du catalogage, ce que faisait déjà Callimaque de Styrière dans la Bibliothèque d'Alexandrie. Depuis, des générations de bibliothécaires ont, comme Monsieur Jourdain, produit des métadonnées sans le savoir. La subtile distinction qui existe entre catalogage traditionnel et métadonnées est que ces dernières, à la différence d'une notice ISBD ou d'un enregistrement en format Marc, sont souvent incluses dans le document primaire et permettent aux ressources électroniques accessibles en ligne d'être repérées, voire « moissonnées », par les moteurs de recherche. Les métadonnées permettent en outre la gestion des droits de propriété intellectuelle du document électronique ou des droits de disponibilité et d'accès aux ressources en ligne sur le web et facilitent grandement l'archivage des ressources électroniques.

Un format répandu de métadonnées est le Dublin Core⁵, apparu en

1995 et devenu norme ISO 15836 en 2003, dont les quinze éléments permettent la description « bibliographique » de ressources électroniques, mais ne peuvent prétendre rivaliser en précision avec les 999 zones possibles des formats Marc. Des tables de concordance existent entre Dublin Core et les principaux formats Marc, notamment Unimarc⁶. Conçu pour être utilisé par les auteurs eux-mêmes des documents électroniques, c'est-à-dire des non-bibliothécaires qui ignorent les arcanes du catalogage et les subtilités de Z 44-050, Dublin Core est aujourd'hui au cœur de l'OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting⁷).

XML

XML, résultat des réflexions d'un groupe de travail du Web Consortium (W3C) de 1996 à 1998, a été conçu pour être un « métalangage » moins lourd que la norme SGML, aussi simple d'utilisation que HTML, et offrant des fonctionnalités propres à baliser des documents à la structure complexe et multimédia. XML, comme SGML, permet le balisage de la structure du document électronique, conserve la séparation établie par SGML entre contenu et présentation et emprunte à HTML sa facilité d'utilisation, les liens hypertextuels et son interopérabilité sur le web. C'est aussi un « métaformat », c'est-à-dire une structure générique conçue pour être transmise, lue et traitée sur le web et qui permet de générer des formats, c'est-à-dire des DTD.

5. Dublin Core Metadata Initiative (DCMI) : <http://dublincore.org/>

6. Tables de concordance Dublin Core-Unimarc : http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html
<http://www.iccu.sbn.it/dubluni.html>

7. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) : <http://www.openarchives.org/OAI/openarchivesprotocol.htm>

Pourquoi les bibliothèques s'intéressent-elles à XML ?

L'apport essentiel de XML est de permettre au document électronique d'échapper aux contraintes d'un support quelconque et à celles des logiciels et des systèmes d'exploitation qui permettent de le créer, de le rechercher, de le visualiser, de l'imprimer, de l'éditer.

De ce point de vue, XML est vraisemblablement appelé à un développement important dans le monde des bibliothèques, notamment dans le domaine de l'enrichissement des catalogues. Toutefois, XML ne se situe pas sur le même plan que les formats Marc, mais plutôt sur celui de la norme ISO 2709 : en effet, XML ne propose pas de définition des éléments bibliographiques pour lesquels les ISBD demeurent toujours aujourd'hui la référence, mais permet une structuration de données respectant la syntaxe d'une DTD. Sans DTD, c'est-à-dire sans contenu, XML n'est qu'une coquille vide et, pour remplir cette coquille, les bibliothécaires disposent d'une solide et réelle expérience et ont des arguments à faire valoir qui font d'eux des interlocuteurs privilégiés, voire incontournables.

Offrant une arborescence *a priori* illimitée, XML permet le traitement « à niveaux » dans une perspective plus simple et plus structurée qu'auparavant, ce qui permet d'envisager la gestion de tables des matières structurées.

La codification en XML des caractères (non latins compris) est basée sur Unicode⁸, ce qui représente un sérieux espoir pour toutes les bibliothèques.

N'autorisant aucun caractère non affichable, XML offre une pérennité à long terme des documents structurés et son formalisme simplifie le traitement automatique en permettant l'échange de données entre systèmes différents : un fichier XML n'est qu'un fichier texte, sans aucun attribut de mise en forme.

8. Unicode : <http://www.unicode.org>

3. W3C XHTML : <http://www.w3.org/TR/xhtml1/>

4. W3C XML : <http://www.w3.org/XML/>

Comme format de stockage, XML apporte une solution intéressante à l'archivage des données numériques puisque c'est un format générique qui peut s'appliquer à toutes sortes de documents, y compris des données non bibliographiques.

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 celle de la norme ISO 2709, en importation et en exportation, l'échange, dans ce dernier cas, pouvant aller jusqu'à la mise en ligne de documents électroniques et d'éléments d'enrichissement du catalogue sur le web. Il y a vingt-cinq ans, c'est par leur aptitude à l'échange des données bibliographiques que les formats Marc se sont imposés en France. C'est par l'échange que XML s'introduit aujourd'hui dans le paysage bibliothéconomique...

XML en bibliothèque

La plupart des DTD aujourd'hui reconnues au niveau international n'ont pas été conçues dans le monde des bibliothèques, même si nombre d'entre elles rendent hommage à la normalisation bibliothéconomique en faisant référence « aux normes internationales de catalogage en vigueur pour les catalogues de bibliothèque » pour structurer leurs métadonnées. Ces DTD peuvent toutefois souvent trouver leur utilité en bibliothèque dans la mesure où elles permettent de structurer le contenu d'un document et non plus sa seule description bibliographique. Parmi les DTD couramment utilisées dans le monde documentaire, on peut citer :

9. TEI : <http://www.tei-c.org/>

(Sparte, Cyberdocs), la base Frantex ou le Trésor de la langue française, ainsi que pour de nombreux corpus littéraires (Voltaire, Proust, Rabelais, etc.) ;

- DocBook¹⁰ (1992, version actuelle 2005) qui permet le balisage de documentation à caractère nettement plus technique (PHP, Linux, etc.) ;

- EAD¹¹ (Encoded Archival Description, 1998, version actuelle 2002) qui permet de baliser des fonds de type archivistique ou, plus généralement, toute collection de documents, y compris les documents dits « spécialisés », organisés en fonds et qui permet aux bibliothécaires et aux archivistes de pouvoir enfin parler le même langage et d'envisager la « reconstitution virtuelle » de fonds séparés pour des raisons historiques ; l'EAD est utilisée pour la conversion rétrospective des 170 000 notices contenues dans les 108 volumes du *Catalogue général des manuscrits des bibliothèques publiques de France* (CGMBPF), initiative de la Direction du livre et de la lecture.

Plus proches des préoccupations quotidiennes des bibliothécaires :

- ISO 12083¹² (1994), DTD bibliographique qui se décline pour les monographies, les périodiques, les articles de périodiques ou les formules mathématiques ;

- MarcXML¹³ (1995 en SGML, 2001 en XML, version actuelle 2003), projet de la Bibliothèque du Congrès, qui permet de convertir des enregistrements Marc21 en XML, etc. ;

- BiblioML¹⁴ (1999), initiée par la Mission recherche et technologie du ministère de la Culture, qui permet la conversion d'Unimarc bibliographique et autorités en XML ;

10. DocBook : <http://www.docbook.org/>

11. EAD : <http://www.loc.gov/ead/>

Voir aussi l'article de Ghislaine Chartron, « La DTD EAD dans les archives et les bibliothèques », *BBF*, n° 2, 2003.

http://bbf.enssib.fr/bbf/html/2003_48_2/2003-2-p112-chartron.xml.asp

12. ISO 12083 :

<http://www.xmlxperts.com/12083xml.htm>

13. MarcXML :

<http://www.loc.gov/marc/marcxml.html>

14. BiblioML : <http://www.biblioml.org/fr/>

- Onix¹⁵ (2000, version actuelle 2005) au cœur du programme EDItEUR et utilisée par le Cercle de la librairie pour sa base Électre ; Onix contient des tables de concordance vers les principaux formats Marc¹⁶ ;

- BookMarc¹⁷ (en cours), prototype conçu pour exprimer les règles sémantiques des formats Marc en XML, élaboré dans le cadre du programme fondamental Unimarc de l'Ifla.

Et maintenant ?

Les catalogues de bibliothèque évoluent, mais une nouvelle conception des catalogues reste à inventer tout en conservant, voire en transfigurant, toutes les innombrables notices Marc existantes et leurs éléments d'enrichissement. Par ailleurs, force est de constater qu'il n'existe aujourd'hui aucun outil ergonomique de production d'information bibliographique en XML, même si plusieurs éditeurs de logiciels de bibliothèque réfléchissent activement à la question.

Marc distingue deux aspects : l'aspect sémantique (données, zones et sous-zones, indicateurs) et l'aspect logique (étiquettes, zones obligatoires et/ou répétables, codage des zones et sous-zones, valeur des indicateurs, caractères de contrôle). Il est évidemment possible de conserver l'aspect sémantique de Marc et d'utiliser la structure logique de XML : du point de vue du catalogueur, encoder un auteur 700\$a ou <auteur> ne change rien sur le fond puisque l'auteur est identifié comme tel pour l'utilisateur. Marc peut ainsi devenir une DTD... C'est ce que font BiblioML, MarcXML ou BookMarc. Ces DTD, utilisées avec un programme de transformation XSL¹⁸ (zones affichables, ponctuation

15. Onix : <http://www.editeur.org/onix.html>

16. Voir l'article de Yves Desrichard, « Vers la convergence des formats bibliographiques ? », *BBF*, n° 3, 2004.

http://bbf.enssib.fr/bbf/html/2004_49_3/2004-3-p55-desrichard.xml.asp

17. BookMarc : <http://www.bookmarc.pt/unimarc/>

18. W3C XSL : <http://www.w3.org/Style/XSL/>

restituée, libellé en clair des zones, etc.) peuvent permettre l'interprétation par des systèmes différents au lieu d'imposer, comme aujourd'hui, un paramétrage différent pour chacun.

Les convertisseurs Marc-XML semblent toutefois constituer les outils d'une phase transitoire car, si l'on bénéficie en sortie d'un fichier qui permet d'utiliser toutes les potentialités de XML, ce fichier n'est finalement rien de plus que la transposition en XML des ISBD avec les deux niveaux d'arborescence des formats Marc. C'est pourquoi les efforts de définition de DTD adaptées aux données que manipulent les bibliothèques et les autres professions du livre se poursuivent.

Quels formats pour l'avenir des catalogues de bibliothèque ?

Depuis plusieurs années, l'Ifla a entamé, autour de la révision des règles internationales, une réflexion de fond. En 1998, ont été publiées les FRBR¹⁹ (Functional Requirements for Bibliographic Records = Spécifications fonctionnelles des notices bibliographiques²⁰), qui explorent un nouveau modèle de description bibliographique centré non plus sur l'auteur mais sur la notion d'œuvre. Ce modèle se prête à la création d'une DTD XML.

À la lumière des commentaires d'utilisateurs de provenances diverses qui proposent des ajouts au modèle, l'Ifla entame la révision du modèle FRBR dont le périmètre va s'étendre au modèle Franar²¹ (Func-

tional Requirements And Numbering of Authority Records = Spécifications fonctionnelles et numérotation des notices d'autorité) et à la modélisation de l'indexation-matière dans le but de mettre en œuvre un modèle unique pour l'ensemble des données bibliographiques. La question de la liaison du modèle en gestation pour les Ressources continues - ISBD (CR) - avec le modèle FRBR est également à l'étude. Autre piste intéressante : le rapprochement entre le modèle FRBR et le modèle CRM²², conçu pour les musées, semble prometteur dans le cadre du web sémantique ou pour l'interrogation simultanée de réservoirs de données hétérogènes.

Depuis 2003, la révision des ISBD est une autre grande préoccupation de l'Ifla. Il est vraisemblable que l'« ISBD nouveau » ne verra pas le jour avant trois ou quatre ans. Ce n'est qu'alors que de nouveaux formats « bibliographiques » fondés sur XML pourront apparaître et que, sans doute dans la prochaine décennie, s'ouvrira l'ère de l'« après-Marc » et que naîtra une nouvelle forme de bibliothéconomie, fruit de toute la réflexion aujourd'hui engagée. D'ici là, la révolution prématurément annoncée se poursuivra lentement : les catalogues actuels s'enrichiront peu à peu de ressources électroniques, mais il est vraisemblable que leur nature ne devrait pas changer en profondeur.

Déjà les limites des DTD apparaissent : elles possèdent une syntaxe éventuellement spécifique, leur modèle doit être complet et unique, elles ne permettent pas de typer les contenus et n'offrent que de faibles possibilités pour typer les attributs. Les DTD ne seraient donc pas la panacée attendue ?

Il semble en fait que la possibilité riche d'avenir soit le concept de Schéma. Les Schémas reprennent les

acquis des DTD, permettent le typage des contenus, utilisent les espaces de noms et permettent l'utilisation de définitions incomplètes et de modèles partiels, ce qui, pour les bibliothèques, offre l'immense avantage de pouvoir utiliser des balises provenant de différentes DTD : en restant dans le contexte aujourd'hui disponible, un document XML contenant des balises BiblioML convertibles en Dublin Core pour les métadonnées descriptives et d'indexation dans une perspective OAI, une instance EAD décrivant l'arborescence d'un fonds et des balises TEI codant la structure d'un document textuel numérisé appartenant à ce fonds, l'ensemble structuré dans un Schéma Mets²³, n'est plus une utopie.

Aujourd'hui, plusieurs modèles de Schémas existent : SchemaXML²⁴, recommandation orientée objet du W3C en date du 2 mai 2001 et RelaxNG²⁵, spécification du Comité Oasis et « draft » ISO depuis 2002, ainsi que Trex²⁶, 2001 et Schematron²⁷, 2004. Les recherches continuent, les expériences se multiplient et ce sont sans doute là les ferments sur lesquels se développeront les formats bibliographiques de la bibliothéconomie de demain.

Que sont les images de couverture et les tables des matières devenues ?

C'était là, avec les quatrième de couverture, la raison de la naissance de la Concertation pour une information bibliographique enrichie. Où en est-on cinq ans après alors que, on l'a vu, le contexte normatif de la bibliothéconomie est en mutation et que, à ce jour, seules des pistes de réflexion sont disponibles ?

19. Groupe de révision des FRBR :

<http://www.ifla.org/VII/s13/wgfrbr/frbr.htm>

20. Traduction française des FRBR :

<http://www.bnf.fr/pages/infopro/normes/rft/frbr.rtf>

21. Françoise Bourdon, *Franar : to what extent authority control can be supported by technical means ?* :

<http://www.ifla.org/IV/ifla67/papers/096-152ae.pdf>

Gerhard J.A. Riesthuis ; Zumer, Maja, *FRBR and Franar : subject access* :

<http://www.ucl.ac.uk/isko2004/sysweb/4bRiesthuisZumer.ppt>

22. Cidoc Conceptual Reference Model (CRM) :

<http://cidoc.ics.forth.gr/>

Traduction française du CRM :

http://cidoc.ics.forth.gr/docs_crm_french_version_3.4.zip

23. Metadata Encoding and Transmission Standard (Mets) :

<http://www.loc.gov/standards/mets>

24. W3C XML Schema :

<http://www.w3.org/XML/Schema>

25. Relax NG : <http://www.relaxng.org>

26. Trex : <http://www.thaiopensource.com/trex/>

27. Schematron : <http://www.schematron.com/>

Force est de reconnaître que la situation n'a pas évolué comme on aurait pu l'espérer...

Le Comité français Unimarc a adopté une structure logique permettant d'intégrer une table des matières jusqu'à huit niveaux dans une notice bibliographique Unimarc. Les titres des différents niveaux renseignent les différentes sous-zones de la zone 359²⁸. Le SIGB peut ainsi indexer le texte pour enrichir les accès et afficher une table structurée. Cette structure a été acceptée par le Puc (Permanent Unimarc Committee de l'Ifla) qui l'a intégrée dans l'édition 2002 du format bibliographique Unimarc en choisissant la zone 327²⁹ déjà utilisée pour les notes de contenu et les notes de dépouillement.

Les autres acteurs ont plutôt bien joué le jeu : l'offre existait avant « Enrichi », mais les fournisseurs de données bibliographiques (Électre, Librairie Decitre, etc.) ont notablement augmenté le nombre de tous ces éléments d'enrichissement (image de couverture, table des matières, quatrième de couverture, résumé, biographie de l'auteur, voire extraits de l'ouvrage) dans leurs propres bases. De leur côté, certains éditeurs de systèmes intégrés de gestion de bibliothèque ont incorporé les zones permettant à Unimarc de gérer ces enrichissements. Enfin, les bibliothécaires sont toujours demandeurs. Où se situe alors le problème ?

On constate par ailleurs que tous les sites de librairies en ligne (amazon.fr, alapage.com, chapitre.com, la-librairie.com, fnac.com, auchan.fr, libriszone.com, etc.) proposent les images de couverture, voire la quatrième de couverture en mode image pour rendre leurs catalogues plus attractifs aux internautes, de plus en plus nombreux, qui ont ainsi la possibilité de flâner dans les rayonnages

28. Contenu des sous-zones de la zone 359 : <http://abf.asso.fr/enrichi/unimarc/cfu/359.htm>

29. Mise à jour 2002 du format Unimarc : <http://www.ifla.org/IV/ifla67/papers/031-188e.pdf>
Contenu des sous-zones de la zone 327 : <http://www.ifla.org/VI/3/p1996-1/concise2.pdf>

« Enrichi »

Née à l'automne 1999, la Concertation pour une information bibliographique enrichie, coordonnée par Dominique Lahary et Pierre-Yves Duchemin, s'est structurée en trois groupes de travail :

- Groupe « Unimarc », piloté par Françoise Bourdon, dont l'objectif était de vérifier l'adaptation d'Unimarc aux besoins définis, c'est-à-dire rechercher et proposer des solutions pour la gestion de l'information bibliographique enrichie dans le cadre normatif en vigueur, notamment pour l'échange de données et faire des propositions au comité permanent Unimarc. Cet objectif a ensuite été confié au comité français Unimarc, créé le 31 mars 2000. « Enrichi » a proposé la création d'une zone pour gérer les tables des matières et le comité français Unimarc a créé la zone 359 qui permet l'introduction de tables des matières, mais il faut reconnaître que les applications pratiques n'ont pas suivi en grand nombre. Cette proposition a été reconnue au niveau international par l'Ifla qui a inclus les tables des matières dans la zone 327 (mise à jour 2002 du format bibliographique Unimarc).

- Groupe « XML », piloté par Pierre-Yves Duchemin, dont l'objectif était d'explorer les possibilités de cette structure de métaformat, notamment les potentialités de XML pour le traitement combiné de l'information secondaire et primaire. Il s'agissait d'une démarche à moyen terme passant par une phase de familiarisation et d'apprentissage, notamment des présentations de DTD par des experts (TEI, EAD, BiblioML, Onix, etc.), et une réflexion sur l'usage possible de XML pour enrichir les catalogues. Le groupe a également mis en place une page de signets sur l'univers XML.

- Groupe « Signalement partagé des ressources en ligne », piloté par Suzanne Santiago puis par Thierry Samain, dont

l'objectif était de promouvoir la coopération entre bibliothèques dans la description et l'organisation de l'accès aux ressources du web et d'explorer les techniques documentaires à mettre en œuvre pour le signalement de ces ressources (pages de liens, intégration dans un catalogue de bibliothèque, bases de données spécialisées).

Ces trois groupes de travail se sont réunis régulièrement pour réaliser leur programme. Des réunions plénières périodiques ont permis de faire le point sur l'ensemble des travaux, et quelques séances thématiques transversales ont été organisées.

« Enrichi » a connu une quarantaine de participants réguliers (bibliothécaires, documentalistes, fournisseurs de SIGB, d'intergiciels, de données et d'administrations centrales) et un paradoxe intéressant : les bibliothécaires n'étaient pas les plus nombreux.

Grâce au site aimablement hébergé par l'ABF*, le site « Enrichi » a pu constituer la plate-forme de travail en ligne des participants : convocations aux réunions et comptes rendus, ainsi que documents de travail et liens vers des sites et ressources accessibles en ligne, comme les liens proposés par le groupe XML. Complété par l'usage de la messagerie, ce site a permis à « Enrichi » de fonctionner sans moyens financiers, sans secrétariat, sans papier à en-tête ni timbres-poste. Il a permis à quiconque y trouvait intérêt de suivre ses travaux.

« Enrichi » était une étape, sans doute nécessaire. Actuellement en sommeil, il n'attend que de nouvelles forces vives pour vivre sa renaissance... ou sa transfiguration.

* <http://www.abf.asso.fr/enrichi>

d'une librairie virtuelle. Les bibliothèques ne sont certes pas des sites marchands, mais est-ce la raison qui les empêche de rendre leur catalogue plus attractif ? Peut-être peut-on le déplorer, mais ce serait jouer la politique de l'autruche que de ne pas le constater, on ne peut que remarquer que les utilisateurs d'aujourd'hui sont devenus plus friands d'illustrations et de documents primaires que de simples descriptions bibliographiques, même si ces dernières sont plus

riches et plus fiables dans les catalogues de bibliothèque que les descriptions figurant sur les sites des librairies en ligne.

L'enrichissement graphique

Toutefois, quelques problèmes subsistent : tout d'abord, puisqu'il s'agit d'images numériques, en ce qui concerne les couvertures et les quatrièmes de couverture en mode

image, se pose l'inévitable question des droits de cession, d'accès et d'utilisation. Sous réserve qu'il soit lisible aussi bien en intranet que sur le web, une solution pourrait être d'établir un lien vers l'image numérique stockée sur un serveur chez le fournisseur de données à partir d'une zone 856 présente dans le SIGB. Dans ce cas, il n'y aurait pas cession d'image mais un simple lien et le problème juridique ne devrait ainsi pas être insurmontable. C'est ce que, moyennant l'ouverture d'un compte, propose aujourd'hui notamment Électre à ses abonnés qui pourraient ainsi bénéficier du lien 856 vers l'image de la couverture dans la notice Unimarc qu'ils reçoivent, parmi une offre de plus de 250 000 images de couverture et plus de 70 000 tables des matières disponibles.

Ce type de dispositif existe et fonctionne aux États-Unis³⁰, au Canada, en Grande-Bretagne, en Suisse, etc. Pourquoi pas en France ? Il est vrai que, notamment dans les pays anglo-saxons, la question juridique est posée de manière différente et que les bibliothèques sont plus souvent considérées par les éditeurs comme des partenaires et des vecteurs de promotion de leurs ouvrages...

L'enrichissement textuel

Le problème se pose en d'autres termes en ce qui concerne les informations textuelles (table des matières, quatrième de couverture en

mode texte, résumé, biographie de l'auteur, extraits). En effet, de plus en plus, ces informations sont stockées chez le fournisseur sous un format XML, Onix ou autre. Dès lors, se pose la question de la récupération de ces informations dans un SIGB.

S'il est techniquement possible d'inclure un fragment XML dans un enregistrement Marc, cette solution ne semble pas porteuse d'avenir et n'est pas la voie à suivre. Il serait plus intéressant d'envisager la récupération de ces données en BiblioML ou en Onix, ces deux propositions en XML offrant l'avantage d'une interrogation possible sur le texte balisé ou le texte intégral.

Une récupération sous forme d'une simple page HTML interrogeable en texte intégral pourrait aussi être une solution. Le fournisseur pourrait également indexer ces zones textuelles et les inclure dans l'extraction Unimarc qu'il fournit à ses abonnés. Différentes solutions apparaissent possibles, mais, sur ce point, le problème se situe du côté des éditeurs de SIGB qui devront développer des solutions adaptées, notamment au niveau de l'Opac ou en développant des moteurs de recherche après avoir implémenté les zones nécessaires d'Unimarc. La récupération des données XML est déjà effective chez les éditeurs et un nombre croissant de libraires. Pourquoi pas dans les bibliothèques ?

En fin de compte, la balle semble bien être dans le camp des bibliothécaires... S'ils sont intéressés par l'enrichissement de leur catalogue, dans la mesure où ils savent où se trouvent les informations désirées, et dans la mesure où, *a priori*, un Opac web permet la gestion de ces documents d'enrichissement de leur catalogue, il semble que la stratégie à suivre soit de réunir et leur éditeur de SIGB et leur fournisseur de données et de

leur poser les bonnes questions sur les solutions proposées par l'un et par l'autre, le rôle respectif de chacun, le coût, la spécificité ou non de la solution proposée. En ce qui concerne les tables des matières, deux questions particulières se posent : le fournisseur de données propose-t-il une table des matières structurée ? La place-t-il dans la zone 327 ou dans la zone 359 ?

La recherche de solutions aussi génériques que possible doit impérativement être au cœur de la réflexion afin de ne pas compromettre la compatibilité future avec les pistes suivies au niveau international pour concevoir des modèles bibliothéconomiques adaptés à la gestion des documents numériques.

L'enrichissement des catalogues n'est pas un gadget inutile, un effet de mode passager, un artifice superflu. Ce peut être au contraire l'occasion de ré-attirer les internautes habitués à trouver des images et du texte sur tous les sites web qu'ils visitent... tous, sauf les sites de bibliothèques. Les bibliothécaires doivent se persuader qu'un Opac enrichi n'est pas un exercice vain, mais le moyen d'apporter une importante valeur ajoutée à la richesse et à la pertinence de son contenu, ainsi qu'un attrait et une apparence plus conformes aux attentes de l'utilisateur.

En les rendant plus attirants, l'enrichissement pourrait être une façon de rendre aux catalogues de bibliothèque un rôle prépondérant dans la diffusion de l'information.

Avril 2005

30. Le Beat (Bibliographic Enrichment Advisory Team) de la Bibliothèque du Congrès utilisait la zone 856 pour insérer des liens vers les tables des matières. Depuis le 1^{er} février 2005, c'est la zone 505 qui est utilisée pour « enrichir les notices bibliographiques de tables des matières numérisées. Les données figurant dans la zone 505 seront générées à partir de la table des matières et fournies par programme, précédées de la légende : Note de contenu générée par programme ».