

Vers la convergence des formats bibliographiques ?

ONIX, application XML du monde de l'édition

Dans un article publié dans le *BBF*¹, Dominique Lahary s'auto-citait en rappelant un de ses avertissements datant de 1994 : « *Prenons garde que nos formats d'échange ne nous permettent d'échanger qu'entre nous*². » De fait, les relations entre les bibliothèques d'une part et le monde de l'édition (éditeurs, diffuseurs, libraires) d'autre part ont toujours été largement marquées, pour ce qui est de l'échange d'informations bibliographiques, par une mutuelle ignorance, voire par une grande incompréhension.

Yves Desrichard

Médiadix
yves.desrichard@u-paris10.fr

La masse imposante des ISBD, des normes Afnor et, surtout, des formats MARC a effrayé plus d'un professionnel de l'édition, découragé par des corpus normatifs qui semblaient peu correspondre aux soucis pragmatiques et commerciaux de leur domaine. Inversement, les bibliothécaires, justement fiers de ce travail collectif de normalisation, n'ont toujours considéré qu'avec méfiance les efforts pourtant réels des éditeurs pour améliorer le travail descriptif sur leur production.

Certes, le *Cataloging in publication* est monnaie courante dans les pays anglo-saxons, s'il reste peu répandu en France. Certes aussi, la principale base de données bibliographiques du monde de l'édition française, Électre Biblio, utilise les normes

Afnor, des pratiques d'indexation similaires à celles de la majorité des bibliothèques françaises (Rameau et Dewey essentiellement), et propose ses notices dans plusieurs formats MARC.

Mais, comme le souligne Dominique Lahary dans l'article précité, les formats MARC, qui sont le volet informatique du diptyque normativo-bibliographique, n'ont jamais pénétré le monde de l'édition : si Électre Biblio propose ses notices en différents formats MARC, c'est uniquement en tant que format d'échange avec une part non négligeable de sa clientèle, et les formats MARC eux-mêmes sont absolument inconnus des autres partenaires de la « chaîne du livre », librairies et diffuseurs notamment.

Aujourd'hui, ces formats sont pour le moins en question. L'arrivée de nouveaux formats descriptifs issus du monde du web et appuyés sur les standards préconisés par le W3C³

1. Dominique Lahary, « Informatisation et bibliothèques : vers la banalisation des outils », *BBF*, 2002, n° 1, p. 60-66. Le présent article lui doit beaucoup.

2. Dominique Lahary, « Du profil de poste au métier », *Bulletin d'informations de l'ABF*, 1994, n° 164. Disponible à : <http://www.membres.lycos.fr/vacher/profess/textes/metier-fiche.htm>

* Merci à Laurent Dervieu (Cercle de la librairie) et Dominique Lahary (Bibliothèque départementale du Val-d'Oise) pour leurs précieux conseils.

3. Pour « World Wide Web Consortium », l'instance qui élabore les standards du web.

Titulaire d'un DESS option informatique documentaire, **Yves Desrichard** est responsable de formation à Médiadix. Il a auparavant travaillé à l'ABES, à la Bibliothèque interuniversitaire scientifique de Jussieu, au Centre national de la cinématographie et à la DLL. Il est l'auteur en 2001 de Julien Duvivier, cinquante ans de noirs destins et en 2003 de Henri Decoin, un artisan du cinéma populaire (*Durante/Bibliothèque du film*).

permet de penser que l'avenir est à d'autres outils d'échange, qui prennent en compte, aussi, l'avènement du document électronique et de ses propres outils descriptifs. Face au développement de la commande en ligne, dont un mastodonte de la « nouvelle économie » comme Amazon.com est le symbole, les éditeurs ont depuis longtemps réfléchi à l'informatisation la plus adéquate de leurs données, si ce n'est de leur production.

Avec ONIX, le monde de l'édition propose désormais un format d'échange d'informations bibliographiques, dont il serait dommage que les bibliothèques, au moins, n'examinent pas l'intérêt, sous peine de voir s'avérer l'avertissement cité plus haut. Le présent article n'a d'autre but que d'initier les professionnels à ce format, qui justifie quelques rappels préliminaires⁴.

De nouveaux formats

On le sait, le web est une des applications les plus utilisées de l'Internet. Pour le définir d'une façon très schématique, le web est à la fois un ensemble de protocoles, qui permettent aux machines connectées de dialoguer entre elles (HTTP est le plus répandu de ces protocoles) et de formats, qui permettent aux outils adaptés à la consultation des ressources web - comme les navigateurs - de lire les informations contenues dans les pages consultées *via* les protocoles.

4. Le standard ONIX est qualifié par EDItEUR, son promoteur, de « ONIX International ». Pour la commodité de la lecture, on se limitera à le mentionner sous le nom d'ONIX.

Quelques sigles

AAP	Association of American Publishers
AFNOR	Agence française de normalisation
BASIC	Book and Serials Industry Communication
BIC	Book Industry Communication
CIP	Cataloging in publication
DTD	Description du type de document
EAD	Encoding Archival Description
FRBR	Functional Requirements for Bibliographic Records
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
INDECS	Interoperability of Data in E-Commerce Systems
ISBD	International Standard Bibliographic Description
MARC	Machine Readable Cataloging
ONIX	Online Information eXchange
SGML	Standard Generalized Markup Language
RAMEAU	Répertoire d'autorité matière alphabétique, encyclopédique et unifié
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Jusqu'à présent, le plus répandu de ces formats est le format HTML. Comme beaucoup de formats, HTML utilise des « balises », autrement dit des « marqueurs » qui permettent aux navigateurs de savoir comment afficher et traiter les informations contenues dans les pages. Ces balises traitent essentiellement de la forme des documents traités (taille des caractères, couleurs, aspect clignotant, etc.) ; quelques-unes traitent du contenu des documents. C'est notamment le cas des balises de type « Titre » ou « Mot clé »⁵.

Ce sont ces balises, qui concernent le contenu du document et non pas sa forme, qui ont donné naissance à un concept en fait vieux comme... les catalogues de bibliothèques, la no-

5. Le standard HTML étant défini en anglais, les noms réels des balises sont bien évidemment dans cette langue et, donc, dans l'alphabet latin.

tion de « métadonnées ». Disons que, dans une application étroite, les « métadonnées » sont des éléments descriptifs d'un document électronique qui font partie de ce document.

Souple, voire fruste, facile à utiliser et à traiter, le format HTML n'est cependant guère plus évolué qu'un outil de traitement de texte. Aussi bien, la nécessité de plus en plus impérieuse de pouvoir structurer intellectuellement les documents disponibles sur le web, et qui se comptent désormais par milliards, a conduit à développer un nouveau format, XML, qui s'appuie aussi sur des balises, mais des balises cette fois signifiantes quant au contenu de l'information.

Un exemple extrait du format XML de la base Électre Biblio sera sans doute plus parlant que n'importe quelle explication (exemple 1 page ci-contre).

Et la même notice en ISBD, cette fois extraite de BN-Opale Plus :

Jacques le fataliste [Texte imprimé] / Diderot ; préf., notes et annexes par Pierre Chartier. – Paris : Librairie générale française, 2000 (72-La Flèche : Impr. Brodard et Taupin). – 414 p. : ill., couv. ill. en coul. ; 18 cm. – (Classiques de poche).

Titre de couv. : "Jacques le fataliste et son maître". – Bibliogr. p. 409-414. – Le livre de poche ; 403. – ISBN 2-253-00413-8 (br.) : 29 F

On le constate, les balises sont les équivalents des étiquettes de zones et des codes de sous-zones des formats MARC. La principale différence est que ces étiquettes sont « en clair », et en français, et qu'il existe des balises ouvrantes (début de l'information) et fermantes (fin de l'information).

Pour poursuivre la comparaison hasardeuse entre XML et MARC, un jeu de balises XML obéit à un ensemble de règles (anciennement qualifié de DTD) qui pourraient être l'équivalent de l'en-tête et du label de notice UNIMARC, mais aussi du contenu du Manuel UNIMARC qui définit les zones et leurs différentes caractéris-

tiques, mais la comparaison est grossière. La DTD définit l'ordre de présentation des éléments, et surtout les relations qui doivent être établies entre les différents éléments.

Pour autant, cette comparaison a ses limites, et notamment celle, fondamentale, du nombre de niveaux de description hiérarchisée de l'information permis. En UNIMARC, seuls deux niveaux sont possibles : la zone et la sous-zone. En XML, le nombre de niveaux est en théorie illimité, même si, dans l'exemple 1, il n'y a que deux niveaux figurés par les indentations.

Les métadonnées

Il faut reconnaître que, très tôt, les bibliothèques se sont préoccupées des évolutions bibliographiques rendues nécessaires par l'avènement de ces nouveaux formats, et des documents électroniques en général, et que les « métadonnées » ont connu, au moins en théorie, un succès foudroyant.

Mais, pour être là encore schématique, les tentatives d'évolution n'ont pas porté dans un premier temps sur l'évolution des outils informatiques bibliographiques pour tous les types de documents, mais uniquement pour les documents disponibles sous forme électronique. En d'autres termes, il ne s'est pas agi de remettre en cause le couple ISBD+MARC pourtant bâti essentiellement sur la description de ressources imprimées, pour l'adapter aux contraintes de la description de documents sous forme informatique, mais de fabriquer des outils descriptifs spécifiques aux documents sous forme électronique.

L'initiative d'OCLC, connue sous le nom de « Dublin Core », est parfaitement symptomatique de cette évolution⁶ : elle a consisté à mettre de côté les « fondamentaux » bibliothéconomiques que constituent les ISBD pour définir un ensemble très simple de 15 informations « de base », desti-

6. Tout sur le Dublin Core à : <http://dublincore.org>

Exemple 1. Notice au format XML extraite de la base Électre Biblio

```
<?xml version="1.0" encoding="UTF-8" ?>
<ENOTICE COUNT="1">
  <RECORD>
    <TITRE TYPTITRE="TP1">Jacques le fataliste</TITRE>
    <AUTEUR>Diderot, Denis</AUTEUR>
    <AUTEUR>Chartier, Pierre</AUTEUR>
    <RESPONSABLE>Denis Diderot</RESPONSABLE>
    <RESPONSABLE>Edition de Pierre Chartier</RESPONSABLE>
    <ARTICLE>
      <RELIURE>Poche</RELIURE>
      <ISBN>2-253-00413-8</ISBN>
      <EAN>9782253004134</EAN>
      <PRIX>4,40</PRIX>
      <DEVISE>EUR</DEVISE>
      <DATE>22/03/2000</DATE>
      <DISPO>Disponible</DISPO>
    </ARTICLE>
    <EDITEUR>
      <NOM>LGF</NOM>
      <LIEU>Paris</LIEU>
    </EDITEUR>
    <COLLECTION>
      <COLLTITRE>Classiques de poche</COLLTITRE>
      <COLLNUMERO>403</COLLNUMERO>
    </COLLECTION>
    <DESCRIPTION>
      <PAGE>414 p.</PAGE>
      <FORMAT>18 x 11 cm</FORMAT>
      <ILLUSTRATION>illustrations</ILLUSTRATION>
    </DESCRIPTION>
    <THEME>843 Romans et nouvelles français</THEME>
    <EDITION>Nouv. éd.</EDITION>
    <THEMGENR>
      <THEME CODE="T130080">Littérature Romans francophones</THEME>
      <GENRE CODE="G110010">Oeuvres littéraires classiques</GENRE>
    </THEMGENR>
  </RECORD>
</ENOTICE>
```

nées à la description des ressources disponibles sous forme électronique et, d'une certaine manière, uniquement de ces ressources.

Il est difficile de dire si le Dublin Core est très utilisé, mais il pêche, à notre avis, de nombreuses façons : d'abord parce qu'il « ostracise » le traitement documentaire des documents électroniques, qui obéit à des règles (d'ailleurs peu définies) beaucoup moins contraignantes que celles qui régissent le catalogage des autres types de documents.

Ensuite et surtout parce que, pour les mettre en pratique, il faut utiliser des « balises » HTML qui ne font absolument pas partie du format standard

défini par le W3C et qui semblent largement inconnues ailleurs que dans le monde des bibliothèques (les grands moteurs de recherche, par exemple, les ignorent superbement). Et l'on retombe sur l'avertissement souligné plus haut : si le Dublin Core peut être utilisé, c'est uniquement dans le cadre d'applications « fermées ». Et les modalités d'échange entre différents systèmes basés sur ce principe, comme l'interrogation simultanée de différentes bases, deviennent rapidement complexes, quand chacun ajoute aux quinze éléments de base (tous facultatifs) des éléments optionnels...

Pour ne pas accabler les créateurs du Dublin Core, disons que leur ini-

tative n'était pas adaptée à HTML, ou plutôt qu'HTML n'est pas adapté à leur initiative. Depuis, les créateurs du Dublin Core ont sans doute compris l'importance d'XML, puisque l'ensemble des métadonnées est désormais disponible sous forme de DTD⁷.

C'est que XML semble, à cet égard, beaucoup plus prometteur, comme le prouve la mise en place de l'EAD pour le monde des archives, basé sur une déclaration XML⁸. Créé en 1993 à la bibliothèque de l'Université de Californie à Berkeley, EAD est une DTD d'encodage d'instruments de recherches archivistiques et de documents d'archives : on se permet d'insister sur le « et », qui prouve qu'on peut utiliser le même format pour structurer à la fois l'information primaire et l'information secondaire.

Les Functional requirements for bibliographic records

Enfin, il faut mentionner l'élaboration, sous la houlette de l'Ifla, des Functional Requirements for Bibliographic Records (FRBR), une étude dont le but était « *d'élaborer un cadre conceptuel permettant de comprendre clairement... l'essence même de ce sur quoi la notice bibliographique est censée renseigner, et l'essence même de ce que nous attendons de la notice en termes d'adéquation aux besoins des utilisateurs*⁹ ».

Autrement dit, les FRBR se soucient de savoir à quoi sert une notice bibliographique... mais sont en réalité plus que cela : il s'agit, d'une certaine manière, d'un modèle conceptuel de données, tel qu'il en existe

dans toute application informatique construite autour de la gestion d'une base de données. En fait, c'est ce qui a manqué, dès le départ, aux formats de type MARC, qui se sont contentés d'informatiser le « pavé ISBD » et d'y ajouter des éléments supplémentaires (données codées, gestion des points d'accès) sans essayer, avant cela, d'examiner les relations entre ces différents éléments¹⁰.

Présentation d'ONIX

Le concept d'origine d'ONIX est né en juillet 1999, lors d'une réunion de The Association of American Publishers. Le but était de créer un standard permettant aux éditeurs d'offrir à leurs clients (libraires, diffuseurs) une information « à valeur ajoutée » sur les produits qu'ils diffusent. Pragmatisme américain aidant, la première version d'ONIX était disponible dès janvier 2000 (ONIX version 1).

Une initiative comparable ayant été mise en œuvre, à peu près au même moment, par le Book Industry Communication anglais, le Book and Serials Industry Communication américain, et EDItEUR, groupe international qui coordonne le développement de standards pour le commerce électronique dans le domaine du livre et des publications en série, les deux projets finirent par se rejoindre, le standard ONIX étant désormais développé et promu sous la houlette d'EDItEUR¹¹.

Il convient aussi de noter qu'antérieurement à ONIX, le projet INDECS, piloté lui aussi par le groupe EDItEUR, s'était inspiré des FRBR, même si, comme le note Patrick Le Boëuf dans un article consacré aux influences des FRBR¹², il n'en a pas réellement compris la portée. Contrairement à

ONIX, le projet INDECS (qui bénéficiait du soutien de la Commission européenne) associait, outre les industries du livre, celles de l'édition musicale, et même des compagnies cinématographiques. Le but du projet était, avant tout, de créer un modèle de métadonnées pour l'échange d'information sur les droits de propriété intellectuelle attachés aux documents concernés¹³.

Principes

ONIX est une « *norme internationale pour la diffusion de métadonnées [enrichies] concernant des livres et d'autres documents utilisés par les bibliothèques et les éditeurs. Ses principes directeurs [guidelines] comprennent des spécifications de contenu, d'éléments de données, d'étiquettes et de listes de codes et une DTD XML*¹⁴ ».

La version 2.0 en français, réalisée avec le soutien du Cercle de la librairie, est disponible sur le site d'EDItEUR¹⁵.

Les objectifs d'ONIX sont clairs : permettre de décrire tous les produits, livres ou non, proposés par « the book industry » ; prendre en compte tous les besoins de tous les secteurs de cette industrie, et pas seulement ceux des vendeurs en ligne ; prendre en compte explicitement les données de droits, de prix, de distribution et de disponibilité des documents décrits ; être utilisable dans un environnement multilingue.

Il est important de comprendre que, *stricto sensu* (voir plus haut), les données de type ONIX... ne sont pas des métadonnées, mais bien des éléments descriptifs sur un document disponible le plus souvent sous forme

7. Disponible à : <http://dublincore.org/documents/2001/04/11/dcmes-xml/index.shtml>

8. Tout sur l'EAD à : <http://lcweb.loc.gov/ead/>

9. *Functional Requirements for Bibliographic Records*, IFLA Study Group on the Functional Requirements for Bibliographic Records.

Disponible à l'adresse suivante : <http://www.ifla.org/VI/s13/frbr/frbr.pdf>

Traduction française disponible à : <http://www.bnf.fr/> ; cliquez sur Informations pour les professionnels, puis sur les Outils bibliographiques.

10. On trouvera une présentation claire et qui plus est humoristique des FRBR dans le texte de Patrick Le Boëuf : *FRBR : un p'tit coin d'paradigme*. Disponible à : http://www.ddb.de/news/pdf/papers_leboeuf_f.pdf

11. Le site d'EDItEUR se trouve à : <http://www.editeur.org>

12. *Op. cit.*

13. On pourra trouver une présentation du projet INDECS en relation avec ONIX dans :

Priscilla Caplan, *International Metadata Initiative : Lessons in Bibliographic Control*.

Disponible à : http://lcweb.loc.gov/catdir/bibcontrol/caplan_paper.html

14. Introduction à la traduction française, disponible à l'adresse suivante :

<http://www.editeur.org/> ; cliquez sur ONIX for Books, puis sur Guide ONIX 2.0 (français).

15. *Op. cit.*

imprimée. La visée commerciale de ce standard est claire, tout comme la volonté de ses promoteurs de l'imposer comme un standard international alors que le marché du commerce en ligne de produits culturels est appelé à un grand développement.

ONIX part du principe (qu'apprécieront les bibliothécaires...) que « plus on dispose d'informations sur un livre, plus on est susceptible de l'acheter » et que, à l'heure de la vente en ligne, plus d'usagers découvrent un livre sur le web qu'en librairie – et qu'il faut en quelque sorte leur offrir en ligne l'équivalent de ce qu'ils peuvent trouver en librairie.

ONIX en France

De la même manière qu'il avait été l'un des pionniers quant à la diffusion de la norme SGML¹⁶ (dont XML constitue en quelque sorte un sous-ensemble) en France, le Cercle de la librairie a compris très tôt l'intérêt de la norme ONIX, dont il assure la promotion en France. Les notices issues de la base Électre Biblio sont désormais disponibles en format XML conforme à la DTD ONIX.

De plus, le Cercle de la librairie est très actif en tant que correspondant français d'EDItEUR, tant pour la diffusion en France de la norme que pour le suivi de son évolution et de ses mises à jour.

Niveaux d'implémentation

À la manière d'une norme descriptive « allégée » ou complète, comme la norme de description des monographies, ONIX est disponible à deux niveaux : le niveau 1 est destiné aux éditeurs qui ne souhaitent pas une trop grande complexité descriptive de leur production et qui n'ont pas les moyens de mettre en place un catalogue informatisé de leur fonds ; le niveau 2, lui, est le « format complet », qui inclut bien évidemment l'en-

semble des informations disponibles au niveau 1.

Types de données

ONIX inclut principalement des données de types suivants :

- renseignements bibliographiques ;
- résumés des documents, comptes rendus critiques, biographies des auteurs, extraits du document...
- couvertures, photos des auteurs...
- droits territoriaux applicables aux documents ;
- prix et disponibilité sur différents marchés ;
- information sur les campagnes promotionnelles autour du document, les prix remportés...
- informations de type audio ou vidéo, liens vers des sites web.

On le constate, le type d'information traité par ONIX va bien au-delà de ce qu'on peut trouver dans des catalogues « traditionnels », même si, par exemple, le site d'Électre offre déjà un bon aperçu des informations « enrichies » qui peuvent être mises à disposition de tous les usagers au-delà du « noyau dur » de type ISBD : aux données de type texte peuvent s'ajouter des images, et même des documents audio ou vidéo.

Si ONIX est spécifiquement conçu comme un format d'échange de données, rien n'interdit aux éditeurs qui le souhaitent de l'utiliser à terme comme un format de production en interne. C'est bien, on s'en souvient, ce qui est arrivé au format UNIMARC.

DTD

Chaque balise présentée dans la DTD a été définie de deux façons, « longue » et « courte ». La balise « longue » est en réalité une balise qui contient du texte en anglais. La balise « courte », qui répond mieux aux objectifs de multilinguisme du projet, est composée d'une lettre et de trois chiffres, qui ne sont pas sans évoquer les étiquettes d'UNIMARC.

Ainsi, le « titre d'ensemble » d'un document est-il défini par la balise

longue « TitleofSet » et la balise courte « b023 ». L'ISBN par la balise longue « ISBN » et la balise courte « b004 ». Les balises courtes présentent les mêmes inconvénients que les structures de type MARC : elles sont incompréhensibles pour le commun des mortels. Mais elles présentent les mêmes avantages : elles facilitent les échanges entre systèmes de langue et de culture différentes (mais toujours dans l'alphabet latin).

Données

On l'a vu, la DTD d'ONIX inclut des types d'information qu'on n'a pas l'habitude de traiter en bibliothèque, par exemple les prix littéraires obtenus par un document ou, pour ce qui est des dimensions d'un ouvrage, sa largeur, élément indispensable pour la gestion des stocks et l'envoi des documents¹⁷ !

ONIX définit plus de 200 données différentes. Certaines (titre, auteur, ISBN) sont obligatoires ; d'autres (comptes rendus, images de couverture) sont optionnelles. La plupart des éléments inclus sont de nature textuelle, mais d'autres sont sous forme d'images et de sons. Les concepteurs d'ONIX insistent particulièrement sur ces derniers, comme étant ceux « qui conduisent à plus de ventes en ligne » - pragmatisme qu'on pourra trouver un peu triste, comme prenant acte de la domination des médias audiovisuels sur la chose imprimée...

Pour le reste, le standard reste relativement familier au professionnel des bibliothèques, qui divise les informations en grands blocs où l'on retrouve des notions traditionnelles au monde du catalogage : titre et variantes du titre, auteur(s), collection(s), édition(s), sujets, éditeur(s), date(s) de publication, dimensions, numérotations normalisées...

D'autres blocs sembleront, en revanche, plus originaux, ainsi celui

16. SGML, outil de normalisation de formats, a donné naissance à HTML puis à XML.

17. Merci à Dominique Lahary de m'avoir signalé cette caractéristique.

Exemple 2. Notice au format ONIX extraite de la base Électre Biblio

```

<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE ONIXMessage (View Source for full doctype...)>
<ONIXMessage refname="ONIXMessage" shortname="ONIXMessage" release="2.1">
  <Header refname="Header" shortname="header">
    <FromCompany refname="FromCompany" shortname="m174">Electre</FromCompany>
    <SentDate refname="SentDate" shortname="m182">200402241407</SentDate>
    <DefaultLanguageOfText refname="DefaultLanguageOfText" shortname="m184">fre</DefaultLanguageOfText>
  </Header>
  <Product refname="Product" shortname="product">
    <RecordReference refname="RecordReference" shortname="a001">9291000111</RecordReference>
    <NotificationType refname="NotificationType" shortname="a002">03</NotificationType>
    <RecordSourceType refname="RecordSourceType" shortname="a194">04</RecordSourceType>
    <RecordSourceName refname="RecordSourceName" shortname="a197">Electre</RecordSourceName>
    <ProductIdentifier refname="ProductIdentifier" shortname="productidentifier">
      <ProductIDType refname="ProductIDType" shortname="b221">02</ProductIDType>
      <IDValue refname="IDValue" shortname="b244">2-253-00413-8</IDValue>
    </ProductIdentifier>
    <ProductIdentifier refname="ProductIdentifier" shortname="productidentifier">
      <ProductIDType refname="ProductIDType" shortname="b221">03</ProductIDType>
      <IDValue refname="IDValue" shortname="b244">9782253004134</IDValue>
    </ProductIdentifier>
    <ProductForm refname="ProductForm" shortname="b012">BC</ProductForm>
    <ProductFormDetail refname="ProductFormDetail" shortname="b333">01</ProductFormDetail>
    <Series refname="Series" shortname="series">
      <SeriesIdentifier refname="SeriesIdentifier" shortname="seriesidentifier">
        <SeriesIDType refname="SeriesIDType" shortname="b273">02</SeriesIDType>
        <IDValue refname="IDValue" shortname="b244">1144-7877</IDValue>
      </SeriesIdentifier>
      <SeriesIdentifier refname="SeriesIdentifier" shortname="seriesidentifier">
        <SeriesIDType refname="SeriesIDType" shortname="b273">05</SeriesIDType>
        <IDValue refname="IDValue" shortname="b244">0001440400</IDValue>
      </SeriesIdentifier>
      <TitleOfSeries refname="TitleOfSeries" shortname="b018">Classiques de poche</TitleOfSeries>
      <Title textcase="01" refname="Title" shortname="title">
        <TitleType refname="TitleType" shortname="b202">01</TitleType>
        <TitleText refname="TitleText" shortname="b203">Classiques de poche</TitleText>
      </Title>
      <NumberWithinSeries refname="NumberWithinSeries" shortname="b019">403</NumberWithinSeries>
    </Series>
    <Title textcase="01" refname="Title" shortname="title">
      <TitleType refname="TitleType" shortname="b202">01</TitleType>
      <TitleText refname="TitleText" shortname="b203">Jacques le fataliste</TitleText>
    </Title>
    <Title textcase="03" refname="Title" shortname="title">
      <TitleType refname="TitleType" shortname="b202">05</TitleType>
      <AbbreviatedLength refname="AbbreviatedLength" shortname="b276">30</AbbreviatedLength>
      <TitleText refname="TitleText" shortname="b203">JACQUES LE FATALISTE ET SON MAITRE</TitleText>
    </Title>
    <Contributor refname="Contributor" shortname="contributor">
      <ContributorRole refname="ContributorRole" shortname="b035">A01</ContributorRole>
      <PersonNameInverted refname="PersonNameInverted" shortname="b037">Diderot, Denis</PersonNameInverted>
      <NamesBeforeKey refname="NamesBeforeKey" shortname="b039">Denis</NamesBeforeKey>
      <KeyNames refname="KeyNames" shortname="b040">Diderot</KeyNames>
    </Contributor>
    <Contributor refname="Contributor" shortname="contributor">
      <ContributorRole refname="ContributorRole" shortname="b035">B01</ContributorRole>
      <PersonNameInverted refname="PersonNameInverted" shortname="b037">Chartier, Pierre</PersonNameInverted>
      <NamesBeforeKey refname="NamesBeforeKey" shortname="b039">Pierre</NamesBeforeKey>
      <KeyNames refname="KeyNames" shortname="b040">Chartier</KeyNames>
    </Contributor>
  </Product>
</ONIXMessage>

```

```

</Contributor>
<NumberOfPages refname="NumberOfPages" shortname="b061">414</NumberOfPages>
<MainSubject refname="MainSubject" shortname="mainsubject">
  <MainSubjectSchemeIdentifier refname="MainSubjectSchemeIdentifier"
    shortname="b191">28</MainSubjectSchemeIdentifier>
  <SubjectCode refname="SubjectCode" shortname="b069">T130080</SubjectCode>
  <SubjectHeadingText refname="SubjectHeadingText" shortname="b070">Littérature Romans
    francophones</SubjectHeadingText>
</MainSubject>
<Subject refname="Subject" shortname="subject">
  <SubjectSchemeIdentifier refname="SubjectSchemeIdentifier"
    shortname="b067">24</SubjectSchemeIdentifier>
  <SubjectSchemeName refname="SubjectSchemeName" shortname="b171">Genre</SubjectSchemeName>
  <SubjectCode refname="SubjectCode" shortname="b069">G110010</SubjectCode>
  <SubjectHeadingText refname="SubjectHeadingText" shortname="b070">Oeuvres littéraires
    classiques</SubjectHeadingText>
</Subject>
<Publisher refname="Publisher" shortname="publisher">
  <PublishingRole refname="PublishingRole" shortname="b291">01</PublishingRole>
  <NameCodeType refname="NameCodeType" shortname="b241">02</NameCodeType>
  <NameCodeTypeName refname="NameCodeTypeName" shortname="b242">Electre</NameCodeTypeName>
  <NameCodeValue refname="NameCodeValue" shortname="b243">1780</NameCodeValue>
  <PublisherName refname="PublisherName" shortname="b081">LGF</PublisherName>
  <Website refname="Website" shortname="website">
    <WebsiteLink refname="WebsiteLink" shortname="b295">http://www.livredepoche.com</WebsiteLink>
  </Website>
</Publisher>
<CityOfPublication refname="CityOfPublication" shortname="b209">Paris</CityOfPublication>
<CountryOfPublication refname="CountryOfPublication" shortname="b083">FR</CountryOfPublication>
<PublicationDate refname="PublicationDate" shortname="b003">20000322</PublicationDate>
<Measure refname="Measure" shortname="measure">
  <MeasureTypeCode refname="MeasureTypeCode" shortname="c093">01</MeasureTypeCode>
  <Measurement refname="Measurement" shortname="c094">180.00</Measurement>
  <MeasureUnitCode refname="MeasureUnitCode" shortname="c095">mm</MeasureUnitCode>
</Measure>
<Measure refname="Measure" shortname="measure">
  <MeasureTypeCode refname="MeasureTypeCode" shortname="c093">02</MeasureTypeCode>
  <Measurement refname="Measurement" shortname="c094">110.00</Measurement>
  <MeasureUnitCode refname="MeasureUnitCode" shortname="c095">mm</MeasureUnitCode>
</Measure>
<Measure refname="Measure" shortname="measure">
  <MeasureTypeCode refname="MeasureTypeCode" shortname="c093">03</MeasureTypeCode>
  <Measurement refname="Measurement" shortname="c094">0.00</Measurement>
  <MeasureUnitCode refname="MeasureUnitCode" shortname="c095">mm</MeasureUnitCode>
</Measure>
<SupplyDetail refname="SupplyDetail" shortname="supplydetail">
  <AvailabilityCode refname="AvailabilityCode" shortname="j141">IP</AvailabilityCode>
  <Price refname="Price" shortname="price">
    <PriceTypeCode refname="PriceTypeCode" shortname="j148">04</PriceTypeCode>
    <PriceAmount refname="PriceAmount" shortname="j151">4.4</PriceAmount>
    <CurrencyCode refname="CurrencyCode" shortname="j152">EUR</CurrencyCode>
    <TaxRateCode1 refname="TaxRateCode1" shortname="j153">S</TaxRateCode1>
    <TaxRatePercent1 refname="TaxRatePercent1" shortname="j154">5.5</TaxRatePercent1>
    <PriceEffectiveFrom refname="PriceEffectiveFrom" shortname="j161">20030615</PriceEffectiveFrom>
  </Price>
</SupplyDetail>
</Product>
</ONIXMessage>

```

consacré aux droits attachés au document, aux commentaires et critiques sur le document, aux prix remportés par le document, ou un autre bloc d'une grande complexité consacré à tous les éléments de prix et de disponibilité.

Là encore, un exemple de notice ONIX sera sans doute plus explicatif que bien des commentaires. La même notice que plus haut, toujours extraite d'Électre Biblio, mais cette fois conforme à la DTD ONIX (exemple 2).

Avantages d'ONIX

ONIX est encore loin d'être un outil parfaitement adapté aux besoins des bibliothèques, même si on peut se demander, par contre, s'il n'est pas déjà adapté aux besoins des usagers. Au moins s'agit-il d'une base de travail qui pourrait être commune aux industries du livre, aux concepteurs de logiciels de gestion documentaire et aux professionnels des bibliothèques. Il a le mérite d'exister, et d'être encore en devenir.

Il s'agit essentiellement d'une structure de description, mais pas d'un outil d'élaboration de descriptifs. En la matière et quoi qu'on en dise, les éditeurs et leurs représentants, comme le Cercle de la librairie, ont déjà acquis un grand savoir-faire. Pour autant, beaucoup d'améliorations peuvent être envisagées. Les bibliothécaires et les documentalistes pourraient y avoir leur part.

Enfin, rien n'interdit de penser que, structure de saisie d'information secondaire (une notice de catalogue grandement améliorée), ONIX pourrait éventuellement être adapté pour remplacer à terme les métadonnées de type Dublin Core dans des documents structurés en XML.

Les sceptiques se demanderont si XML est vraiment l'avenir du web, alors même qu'on ne cesse de l'annoncer, sans pour autant voir les applications basées sur ce « format » remplacer les pages HTML aussi rapidement que prévu. S'agit-il bien d'une

évolution inéluctable ? Personne n'a réellement la réponse. Ce qui est sûr cependant, c'est que le web est désormais omniprésent dans la gestion des catalogues, des bases de données, de la majorité des outils bibliographiques et, bien entendu, de la documentation électronique proprement dite : ONIX est explicitement un outil adapté à ce contexte, il serait dommage de passer à côté.

Limites d'ONIX

Il serait présomptueux de dire qu'ONIX pourrait venir, à terme, remplacer les formats de type MARC, même si l'outil déjà développé par le monde de l'édition semble extrêmement prometteur.

D'une part, ONIX couvre essentiellement la description de ce que les bibliothèques ont coutume de nommer « monographies ». Son extension aux « publications en série » est en cours¹⁸, et est vivement souhaitée par tous les partenaires, celle à d'autres types de documents (sonores, audiovisuels) se heurtera sans doute à la séparation entre le monde de l'édition imprimée et... d'autres mondes, où les pratiques normatives sont moins développées et, sans doute, les contraintes commerciales (qui sont à la base de la création et de la promotion d'ONIX) tout autres.

D'autre part, ONIX présente les mêmes défauts que les formats de type MARC si on les isole du corpus normatif sur lesquels on les a bâtis, à savoir les ISBD et autres normes de catalogage : si l'information est suffisamment détaillée pour savoir à l'intérieur de quelle balise placer telle ou telle information, si la liste des informations à prendre en compte semble assez large, rien n'est dit sur l'élaboration même des données à saisir, qui sont l'objet essentiel des normes de catalogage à l'heure où la ponctua-

tion ISBD est devenue un simple programme.

ONIX et les bibliothèques

Même si l'on manque d'informations, il semble qu'ONIX ait déjà rencontré l'intérêt des bibliothèques anglo-saxonnes. La British Library, en particulier, a consacré une étude à la possible utilisation d'ONIX comme norme d'échange de données bibliographiques entre le secteur du commerce du livre et les bibliothèques¹⁹ et réfléchit à l'idée d'une sorte de « continuum bibliographique » entre éditeurs et bibliothèques.

On s'est aussi intéressé aux correspondances entre UNIMARC et MARC21 d'une part, ONIX d'autre part. Des « traductions » ont été élaborées par la Bibliothèque du Congrès et par la British Library²⁰. Alan Danskin, auteur de la version ONIX/UNIMARC, a résumé dans un article²¹ les limites de l'exercice : de nombreuses informations présentes dans ONIX n'ont pas d'équivalent dans UNIMARC. Inversement, ONIX ne prend pas en compte, même en se limitant aux documents imprimés, les caractéristiques propres aux publications en série, etc.

Cet exercice nous paraît relever du même malentendu qui a fait élaborer des DTD pour transformer MARC21 et UNIMARC en XML : croire que l'on peut faire cohabiter deux mondes irréductiblement différents, et les faire dialoguer entre eux. Car, s'il est possible de passer d'un format MARC à XML, l'opération inverse est quasiment impossible.

Les problèmes posés ne tiennent pas tant à la nature des données qu'aux caractéristiques « ontologiques » des structures : conçus pour

19. Cette étude ne semble plus disponible sur le web.

20. Disponibles sur le site d'EDITEUR à : <http://www.editeur.org/onixmarc.html>

21. Alan Danskin, *UNIMARC, ONIX and the future*. Disponible à :

<http://www.ifla.org/IV/ifla67/papers/008-188e.pdf>

18. Plus d'informations à : <http://www.editeur.org/> cliquez sur ONIX For Serials, puis sur The Serial Title Record.

l'informatique des années 1960 ou, au mieux, des années 1970, les formats MARC ne peuvent prétendre à la même souplesse et à la même pertinence, dans le monde Internet, que des formats comme XML. Leur rigidité, qui a été leur principal avantage,

**Le commerce du livre
a tout à la fois
plusieurs longueurs
d'avance et plusieurs
longueurs de retard
sur les bibliothèques
en matière de description
des documents**

permettant la création de dizaines de millions de notices bibliographiques échangeables entre systèmes différents, est désormais leur principal inconvénient : ils ne peuvent plus évoluer qu'à la marge, là où XML, qui est en fait plus un métaformat qu'un

format, a de plus grandes latitudes – même s'il n'est pas dépourvu de contraintes.

Inversement, les professionnels du livre ont sans doute beaucoup à apprendre du travail normatif portant sur l'élaboration des informations, et non plus sur la structure qui les traite, qui est le quotidien des catalogueurs. Quand on considère les problèmes auxquels se sont très rapidement heurtés les concepteurs du Dublin Core, par exemple sur la forme des noms d'auteurs, la nécessité de vocabulaires pré-coordonnés pour l'indexation, etc., on se dit que, comme le veut le proverbe, « *s'il y a une solution technique, c'est qu'il n'y a pas de vrai problème* » : autrement dit, que le problème est ailleurs, non tant dans l'utilisation d'XML que dans la normalisation des pratiques en matière de création des données bibliographiques ou autres.

Conclusion

Les FRBR, évoquées au début de cet article, ont bien montré que le carcan normatif sur lequel se sont bâtis les catalogues informatisés des

bibliothèques ces trente dernières années devrait évoluer, et commence d'ailleurs à le faire, timidement.

Une nouvelle fois, le commerce du livre a tout à la fois plusieurs longueurs d'avance et plusieurs longueurs de retard sur les bibliothèques en matière de description des documents : plus de pragmatisme à court terme, mais moins de souci de la pertinence des informations sur la durée, de leur pérennité comme de leur portabilité dans d'autres environnements.

Au moins en théorie, ONIX constitue une confluence idéale entre des intérêts qu'on aurait du mal à percevoir comme contradictoires. Qui plus est, il se trouve en prise directe avec cette « banalisation des outils » évoquée par Dominique Lahary dans son article déjà cité. On comprendrait mal que, arc-boutées sur leurs « spécificités », là où les usagers, eux, n'ont pas forcément ce souci de la différence, les bibliothèques manquent encore une fois ce rendez-vous de l'harmonisation de l'information bibliographique.

Mars 2004