

Etat de l'art en matière de normes et approche critique :

De Marc à Dublin Core

Journée « Indexation des ressources pédagogiques », Lyon, 16 Novembre 2004

Elizabeth CHERHAL, Cellule MathDoc, UMS5638, CNRS/Université Joseph Fourier, Grenoble.

Elizabeth.Cherhal@ujf-grenoble.fr

Résumé :

La question des normes de description bibliographique « classiques » sera abordée d'abord d'un point de vue historique, puis d'un point de vue « utilitaire » : l'adaptation aux besoins et aux types de documents décrits.

Nous verrons le contexte historique de création des normes MARC¹ et de son corollaire, la norme de catalogage « sur fiches » ISBD², mais aussi l'utilité de cette normalisation « universelle » pour l'échange de métadonnées entre bibliothèques et pour l'accès réparti à des bases de données distantes et différentes.

Nous verrons ensuite le contexte de l'initiative « dublin core », ses protagonistes et citerons des applications importantes, telles qu'OAI³.

En conclusion, et au-delà des normes, nous aborderons l'interconnexion souhaitable de deux mondes encore un peu distincts : celui des bibliothèques et celui de la documentation en ligne et des archives ouvertes.

Mots Clés : indexation, catalogage, marc, dublin core, métadonnées, archives ouvertes

Des normes pour la description bibliographique « généraliste »

Marc et Dublin Core ont pour caractéristique commune de se proposer de décrire toute ressource documentaire. En ce sens, on peut les qualifier de normes (ou standards) de description bibliographique généralistes. Comme toutes les normes (ou standards) de description bibliographique, Marc et Dublin Core ont pour but de décrire l'objet documentaire afin de le retrouver, puis éventuellement de le partager. Ces deux standards ont néanmoins des utilisations et histoires différentes :

La norme MARC est née en 1965 dans le but de procéder à « l'informatisation » des catalogues des bibliothèques. De manière concomitante, les normes ISBD, publiées à partir de 1971, constituent un ensemble de spécifications pour la description bibliographique, ainsi que sa présentation sur fiches. MARC apparaît aujourd'hui comme très empreint des contraintes imposées à la fois par le catalogage sur fiches et par l'informatique naissante des années 60-70. Cependant, il s'agit d'une norme aujourd'hui incontournable et utilisée dans le monde entier par toutes les bibliothèques importantes.

¹ MARC MACHine-Readable Cataloging

² International Standard Bibliographic Description

³ Open Archives Initiative

Le standard Dublin Core, qui date de 1995, a pour premier objectif de décrire les ressources (documents, mais aussi d'autres ressources) en ligne. Le but était de permettre une certaine interopérabilité dans un monde sans description bibliographique, le monde du web. Les acteurs du Dublin Core ont voulu établir un « plus petit dénominateur commun » de la description, quitte à permettre des extensions connues sous le nom de « Dublin Core qualifié ».

MARC aujourd'hui

Dans le passé, il existait pratiquement une version de MARC par pays, ce qui était assez dommageable pour l'interopérabilité. Aujourd'hui il y a deux formats reconnus par l'IFLA⁴, les formats MARC21 et UNIMARC. UNIMARC est le résultat d'une normalisation des différentes variétés de MARC et est surtout utilisé en Europe. MARC21 est le successeur de LC-MARC, utilisé par la bibliothèque du congrès et plus généralement en Amérique du nord.

MARC permet une description bibliographique très détaillée, groupée (logiquement) en blocs, zones et sous-zones, comprenant des informations y compris sur les zones qui doivent être indexées.

Il comporte un format physique d'échange, le format ISO2709, dont voici un exemple :

```
01235cam 2200337 450 (label)
0010021000000100031000210200017000520210053000069100004100122101
000800163102000700171105001800178106000600196200009400202205002      répertoire
60029621000630032215002700385320001000412324006200422345001800
484606002900502676001200531686007000543700004700613801003900660
930004200699960004100741960003800782960003900820960003800859-
FRBNF3900058000000008- $a2-87647-210-4$bbr$d56
EUR- $aFR$b00317406- $5759999999:39000580001001$aFR$bDLE-20030521-
19871- $a20030521e1979
    k y0frey0103 ba-0 $afre- $aFR- $ay z 000yy- $ar-1 $aPanorama des      Contenu et
mathématiques pures$b Texte imprimé$e le choix bourbachique$f Jean      indicateurs
Dieudonné- $a[Reprod. en fac-sim.]- $aParis$c J. Gabay$d 2003$e 14-Condé-sur-      de traitement
Noireau
$g Impr. Corlet- $a XV-306 p.$c ill.$d 24 cm- $al Index- $a Fac-sim. du second tirage de
l'éd. de Paris: Gauthier-Villars, 1979- $b 9782876472105- $3 11932434
$a Mathématiques- $a 510 $v 21- $a 510 $2 Cadre de classement de la Bibliographie
nationale française- 1$3 11900186$a Dieudonné$b Jean$f 1906-
1992$4 070- 0$a FR$b BNF$c 20030521$g AFNOR$2 intermrc-$5 759999999:39000
580001001$a 2003 115383- $3 11932434$a Mathématiques
appliquées- $3 11932434$a Mathématiques modernes- $3 11932434$a Méthodes
mathématiques- $3 11932434$a Sciences mathématiques- 0
```

On devine, au vu de cet exemple, que le format hérite de contraintes techniques qui, aujourd'hui, n'ont pas lieu d'être.

Pour résumer, on peut citer à l'actif de MARC :

- ses spécifications ouvertes,
- une reconnaissance universelle,
- l'existence du format d'échange qui a permis d'une part le partage de métadonnées entre bibliothèques et la mise en place de catalogues tels que le SUDOC⁵ et d'autre part le déploiement de l'outil Z39.50, permettant une interrogation normalisée et à distance de différents catalogues (intégration dans de nombreux outils de catalogue et mise en place d'interfaces d'interrogation de plusieurs catalogues simultanément (CCF⁶)).

⁴ International Federation of Librarians Associations

⁵ système Universitaire de Documentation (catalogue collectif des bibliothèques universitaires)

⁶ Catalogue Collectif de France

Et au passif :

- sa complexité, qui le rend difficilement utilisable en dehors du contexte des bibliothèques classiques : on verrait difficilement son utilisation dans le cadre d'une grande collection de documents numériques,
- sa relative inadaptation aux conditions technologiques actuelles,
- le fait que, vu la complexité, la plupart des logiciels qui s'y conforment sont assez coûteux, (mais il faut souligner l'exception : le logiciel libre Koha).

Dublin Core de sa naissance à nos jours

Le Dublin Core (« core » signifie « noyau »), n'est pas une norme. On peut le qualifier de standard de fait. Il est né en 1995 à Dublin (Ohio), lors d'une réunion entre différents acteurs, dont en particulier l'OCLC⁷, lui-même acteur de la communauté des bibliothèques. Aujourd'hui il est maintenu par le DCMI⁸ et donne lieu à une conférence annuelle au contenu riche et varié. Le DCMI comporte de nombreux groupes de travail, par exemple le groupe « éducation » qui tient une conférence commune avec le groupe LOM⁹.

L'objectif de l'initiative, déclaré sur son site web, est :

« The Dublin Core Metadata Initiative (DCMI) is an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems ».

L'initiative Dublin Core est partie du constat que d'une part beaucoup de documents (et autres ressources) étaient disponibles en ligne sans aucune description, donc difficilement trouvables, et d'autre part qu'il n'était pas possible de « cataloguer » ces ressources en utilisant des moyens traditionnels. Il fallait donc un moyen de décrire simplement des ressources en ligne ou plus généralement les ressources numériques.

Les participants sont arrivés à un consensus sur les points suivants :

- la description devait être assez simple pour être comprise et utilisée par un grand nombre de personnes/ projets. Il fallait donc établir un dénominateur commun de base,
- la description devait être indépendante des formats physiques,
- la description devait être suffisamment généraliste pour permettre de décrire des ressources autres que les documents classiques (gouvernement, environnement, enseignement, etc...).

Ces principes de base sont toujours respectés, comme on le verra plus tard.¹⁰

Une des premières idées des acteurs du Dublin Core était d'améliorer la qualité des résultats de recherche des « moissonneurs » web. Le Dublin Core peut/pouvait être destiné au moissonnage par un moteur de recherche « web » tel que google, yahoo ou encore un moteur plus spécialisé. Dans les faits, dans un premier temps, c'est cette utilisation spécialisée qui a prévalu et, dès son lancement, le Dublin Core a servi de base à de nombreux projets et applications dans le domaine des bibliothèques numériques, tels que le projet CISMef (Catalogue et Index des Sites Médicaux Francophones), le « Victorian Education Channel » ou le projet EULER (European Libraries and Electronic Resources in Mathematical Sciences). De fait, ces projets ont souvent aménagé le Dublin Core à leur manière et sa véritable utilité comme dénominateur commun a été démontrée par l'initiative OAI (Open Archives Initiative). OAI a en quelque sorte « établi » le Dublin Core comme standard. Tout serveur OAI a pour

⁷ OCLC : Online Computer Library Center

⁸ DCMI : Dublin Core Metadata Initiative

⁹ LOM : Learning Object Metadata, voir d'autres contributions dans ces actes

¹⁰ Voir «Présentation des Standards: le Dublin Core», dans ces actes

obligation de fournir des métadonnées en « dublin core » de base. La préexistence du Dublin Core a grandement facilité le lancement fulgurant de l'initiative OAI et de son protocole OAIPMH.¹¹

On peut résumer à l'actif du Dublin Core :

- sa simplicité qui le rend facilement compréhensible et utilisable,
- son aspect généraliste,
- la grande variété de ressources à décrire,
- son indépendance par rapport aux formats physiques,
- la vivacité de ses différentes composantes (réunions, groupes de travail)...

Au passif, et corollaire de l'actif :

- son côté généraliste et incomplet, nécessitant souvent des extensions,
- sa relative jeunesse et le fait qu'il évolue encore (bien que les éléments de base semblent être « gravés dans le marbre »).

Cohabitation et avenir communs ?

Schématiquement on pourrait dire que MARC représente le monde des bibliothèques classiques et Dublin Core celui des bibliothèques numériques et du mouvement pour les archives ouvertes et qu'ils sont apparemment à des lieues l'un de l'autre. En réalité, ces deux mondes se croisent et se parlent relativement souvent. Les bibliothèques classiques se préoccupent de plus en plus de ressources numériques diverses et variées et les responsables de bibliothèques numériques se rendent compte que les bibliothèques classiques ont beaucoup à leur apprendre. En témoignent certainement les nombreux « emprunts » à MARC que l'on peut trouver dans le Dublin Core « qualifié »¹².

A l'avenir, sur le plan purement technique, les initiatives en cours, telles que MARCML ou BIBLIOML, pour exprimer MARC en XML (et s'affranchir des contraintes de l'ISO2709) peuvent également faciliter ce dialogue entre bibliothèques classiques et numériques. Dans le même ordre d'idée, des outils pour la conversion des formats, telles que le « Dublin Core/MARC/GILS Crosswalk » ou le « Marc to Dublin Core Crosswalk » ont pu montrer leur utilité.

Sur un plan un peu plus politique, OAI a indéniablement favorisé la cohabitation. On peut prendre comme exemple, la BnF qui a lancé un serveur OAI devant sa bibliothèque numérique Gallica. ou le fait que l'ABES¹³ ouvre un portail où l'on trouve de nombreuses notices moissonnées par OAI.

¹¹ OAIPMH : Open Archives Initiative Protocol for Metadata Harvesting

¹² Voir « Présentation des standards : le Dublin Core »

¹³ Agence Bibliographique de l'Enseignement Supérieur

Références :

ISBD : <http://www.ifla.org/VI/3/nd1/isbdlist.htm>

MARC : <http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>

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

:: F Leresche « les formats MARC » http://www.rnbm.org/rencontres_2004/leresche-marc.pdf

Dublin Core : <http://dublincore.org/>

Outils de conversion : <http://www.loc.gov/marc/dccross.html>

<http://www.loc.gov/marc/marc2dc.html>

OAI : <http://www.openarchives.org/>

Z39.50 : <http://www.loc.gov/z3950/agency/>

: Dominique Lahary « La norme Z39.50 »

<http://membres.lycos.fr/vacher/profess/cours/mediadix/z3950/>