13-14 octobre 2009

Cours Luxembourg INAP

Histoire et principes du catalogage Evolution actuelle Avenir

5) Formats de catalogage et métadonnées

Pierre Gavin

Etat: 5 oct. 2009

www.pierregavin.ch

© Pierre Gavin

- 5.1) Formats de catalogage (codification des données)
- 5.2) Métadonnées

Pour que l'ordinateur puisse les traiter correctement, les notices de catalogage doivent être codifiées précisément; cette codification est appelée "format". Les trois grands formats utilisés dans les bibliothèques sont MARC21 (y c. sa variante IDS-MARC), UNIMARC et MAB2. Quelle évolution pour MARC21? XML joue un rôle de plus en plus important, mais ne bouleverse pas l'analyse logique des éléments à codifier. Mise à jour 5 oct. 2009

Sommaire

- 5) Formats de catalogage et métadonnées
- 5.1) Formats de catalogage (codification des données)
- 1) Que veut dire "format"?
- 2) Références
- 3) La naissance de MARC21
- 4) Pourquoi codifier
- 5) Comment codifier

Terminologie

Format fixe (de longueur fixe)

Format entièrement variable: MARC

Importance d'une bonne codification

6) MARC21

Origine

Principes de base

Evolution des formats MARC

Site Web de référence pour MARC21

- 7) ISO 2709
- 8) Quelques limites de MARC21

Article (déterminant) en début de titre

Oeuvres en plusieurs volumes

Multilinguisme

Ponctuation

- 9) IDS-MARC
- 10) UNIMARC
- 11) MAB2
- 12) Format interne / Format d'échange
- 13) L'avenir de MARC21

HTML

XML

MARC21 et XML

Les critiques les plus fréquentes contre MARC21

L'avenir est déjà présent: le CERN CDS

- 5) Formats de catalogage et métadonnées
- 5.1) Formats de catalogage (codification des données)

05.10.2009 22:34 1 sur 26

1) Que veut dire "format"?

■ Format de saisie: grille de saisie, par ex. pour le catalogage

■ Format d'affichage: présentation des notices de catalogage dans l'OPAC

■ Format du livre: 23 cm, 8°

■ Format de stockage: bases de données, par ex. Oracle

■ Format d'échange: par ex. MARC21

2) Références

Leresche, Françoise

Les formats MARC. - 2004

http://www.rnbm.org/rencontres_2004/leresche-normes.pdf

Eversberg, Bernhard. - Was sind und was sollen bibliothekarische Datenformate. - 1999

http://www.allegro-c.de/formate/formate.htm

3) La naissance de MARC21

Henriette Avram, Library of Congress, 1965. (Permalink)

Corporate Name: Library of Congress. Office of the Information Systems Specialist.

Main Title: A proposed format for a standardized machine-readable catalog record; a

preliminary draft prepared by Henriette D. Avram, Ruth S. Freitag [and] Kay D.

Guiles.

Published/Created: [Washington] 1965.

Related Names: Avram, Henriette D.

Freitag, Ruth S.

Guiles, Kay D.

Description: viii 1., 110, 4, 22 p. illus. 27 cm.

Subjects: MARC formats -- United States.

C'est le rapport qui a lancé le projet de codification MARC

Une notice MARC, simplifiée:

100 \$a Ramuz, Charles Ferdinand

245 \$a Derborence: \$b roman / \$c Charles Ferdinand Ramuz

260 \$a Paris: \$b Grasset, \$c 2003

300 \$a 182 p.: \$b ill.; \$c 23 cm

490 \$a Les cahiers rouges

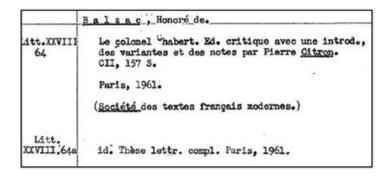
Une notice MARC contient donc:

- des **zones** (100, 245, etc.)
- dans chaque zone, des sous-zones (\$a, \$b, etc.)

Mais pourquoi faut-il codifier les notices de catalogage, et quelles sont les méthodes de codification?

4) Pourquoi codifier

L'homme sait interpréter une fiche classique de catalogage, mais pas la machine (pas encore...?)



Pour manipuler les données correctement, l'ordinateur doit pouvoir repérer précisément l'auteur, la date, la cote, etc.

5) Comment codifier

Terminologie

- Format = codification
- Notice = record
- Zone = champ = field

Format fixe (de longueur fixe)



Petit meuble à tiroirs, lkea: le nombre et la dimension des tiroirs sont prédéfinis.

Z304-ZIP	0	X(9)
Z304-EMAIL-ADDRESS	0	X(60)
Z304-TELEPHONE	0	X(30)
Z304-DATE-FROM	М	9(8)

Fichier des lecteurs Aleph, extrait:

- chaque champ est de longueur fixe
- un champ ne peut pas être répété

Format entièrement variable: MARC

```
100
       $a Bille, Stéphanie Corinna
245
       $a Les entretiens, 1964-2002 / $c S. Corinna Bille &
Maurice Chappaz ; [éd. sous la dir. de Henri Marin]
       $a [Sierre] : $b Ed. Monographic, $c 2003
      $a 3 vol. ; $c 21 cm
300
      $a Chappaz, Maurice $v [entretien] $2 rero
600
      $a Bille, Corinna $v [entretien] $2 rero
600
700
      $a Chappaz, Maurice
      $a Marin, Henri
700
```

- Chaque notice a un nombre variable de zones
- Chaque zone a un nombre variable de sous-zones
- La longueur des zones et des sous-zones est variable.

Importance d'une bonne codification

Une codification variable est absolument nécessaire pour les données bibliographiques, mais cela ne suffit pas; il faut en plus que la codification soit logique, cohérente, que tous les éléments importants soient codifiés de manière claire et univoque.

Contre-exemple: système Swissbase

Cette codification est de longueur variable, mais plusieurs éléments sont mal codés, ou pas du tout codés.

```
0 19930625
 200
        Class, Edi; Bischofberger, Sabine
 300
        Das Arbeitszeugnis und seine "Geheimcodes". Die rechtlichen Grundlagen.
Stellenwert und Bedeutung für Arbeitnehmer und Arbeitgeber. Beispiele für Formulierung
und Interpretation.
 400
        2.Auflage
 500
        Zürich
 501
        Verlag des Schweizerischen Kaufmännischen Vereins SKV
 502
        1993
 504
        94 S.
 620
        SBB-GD- 1.5163
 650
        Standort: 1=Baull-AP (195/93), 1=Baull - RC; 1=PA-A-Sekr.
 700
        331.115.3 / Arbeitszeugnis
```

Une codification logique et précise est indispensable:

- pour indexer les données de manière optimale
- pour paramétrer l'OPAC
- pour faire des statistiques fiables
- pour convertir les données dans un nouveau système (migration)
- pour faciliter l'échange des données

Un bon format est comme la colonne vertébrale des données.

6) MARC21

Origine

MARC = **Ma**chine **r**eadable **c**ataloging

Mis au point en 1965 à la Library of Congress par Henriette Avram 1919-2006



Volonté de la Library of Congress d'automatiser le catalogage, notamment pour résoudre le problème du stockage des fiches distribuées.



Vidéo de la Library of Congress

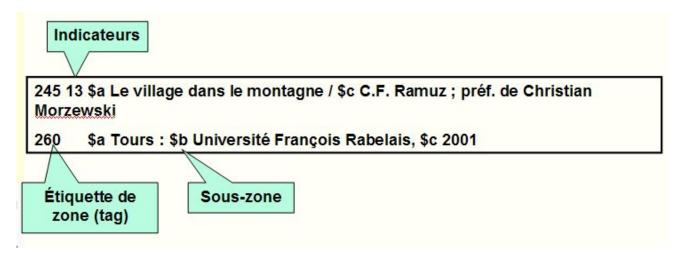
http://www.loc.gov/cds/audio/CDSshort30.ram

Principes de base

Analyse très fine, en fonction des spécificités du catalogage

Format variable (zones et sous-zones répétitives; longueur variable)

Structure "étiquette / indicateur / sous-zones"



Evolution des formats MARC

Dès 1966 aux Etats-Unis et dans le monde: LC-MARC, puis MARC II, puis USMARC, puis MARC21 (et toutes les variantes nationales de MARC, notamment IDS-MARC)

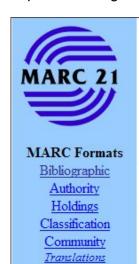
Dès 1973 en Allemagne et en Autriche: MAB1, puis MAB2, puis abandon de MAB au profit de MARC21

Dès 1977 en Europe: UNIMARC

Site Web de référence pour MARC21

Library of Congress MARC Standards

http://www.loc.gov/marc/



MARC LITE

Bibliographic

Cale List

MARC STANDARDS

Library of Congress - Network Development and MARC Standards Office

The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form.

[MARC en ESPAÑOL]

<u>Understanding MARC Bibliographic</u> - a brief description and tutorial $\underline{\textbf{Understanding MARC Authority}} - a \textit{ brief description and tutorial}$

General Information Introductory MARC Information

News & Announcements MARC FAQ MARC Forum (listserv) Unicode-MARC Forum

Remainended Reading

MARC Formats

MARC in XML Formats and Code Lists MARCXML Format Status MODS Ordering Information MADS National Level

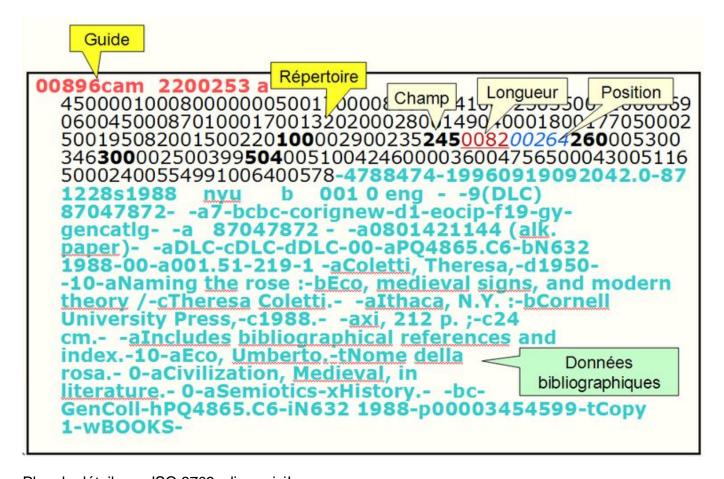
MARC and FRBR MARC Mappings MARC User Notes

Requirements

FRRR Display Tool

7) ISO 2709

La norme ISO 2709 définit l'organisation des données catalographiques MARC sur une bande magnétique. Bien que très ancienne et obsolète, cette norme est encore largement utilisée aujourd'hui, car tous les systèmes de bibliothèques savent envoyer et recevoir des notices sous forme ISO 2709.



Plus de détails sur ISO 2709, cliquer ici!

05.10.2009 22:34 11 sur 26

8) Quelques limites de MARC21

Article (déterminant) en début de titre

L'article (le déterminant) en début de titre est codé de la manière suivante: le contenu du 2e indicateur donne le nombre de caractères qu'il faut sauter pour trier le titre correctement.

245 13 \$a Le monde aujourd'hui

Cette méthode pose plusieurs problèmes:

- seule la première sous-zone peut être traitée avec cette méthode
- selon les zones, c'est le 2e ou le 1er indicateur qui est utilisé
- dans certaines zones (par exemple 246), cette méthode ne peut tout simplement pas être utilisée
- le décompte du nombre de caractères peut poser problème avec Unicode
- etc.

Une solution mieux structurée, généralisable, a déjà été définie dans le cadre de MARC21, il ne reste plus qu'à l'implanter...

Voir le rapport = http://www.loc.gov/marc/marbi/dp/dp118.html

Oeuvres en plusieurs volumes

La méthode la plus souvent utilisée est la saisie des volumes dans une zone de notes 505.

100 1 \$a Balzac, Honoré de

245 10 \$a Illusions perdues.

260 \$a Paris : \$b Librairie nouvelle, \$c 1857-1858.

300 \$a 2 t.; \$c 8°.

505 0 \$a T. 1. Les deux poètes ; Un grand homme de province à Paris. - 1857 --

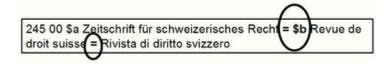
T. 2. Un grand homme de province à Paris suite ; Eve et David. - 1858.

Mais il n'y a pas vraiment de méthode pour les cas complexes avec notices chaînées.

Multilinguisme

Il manque dans MARC21 une méthode pour coder les langues des éléments.

Exemple du titre parallèle: il n'est même pas codé spécifiquement comme titre parallèle, donc pas non plus de code de langue.



Ponctuation

Dans MARC21, on doit saisir simultanément les codes de sous-zones et la ponctuation ISBD; très souvent, il y a redondance. En bonne logique, on ne doit pas saisir dans la base de données des éléments qui concernent la présentation dans l'OPAC.

245 10 \$a Condition de l'homme : \$b essai / \$c Henri-L. Miéville

":" veut dire: début du sous-titre, \$b également

"/" veut dire: début de la mention de responsabilité, \$c également.

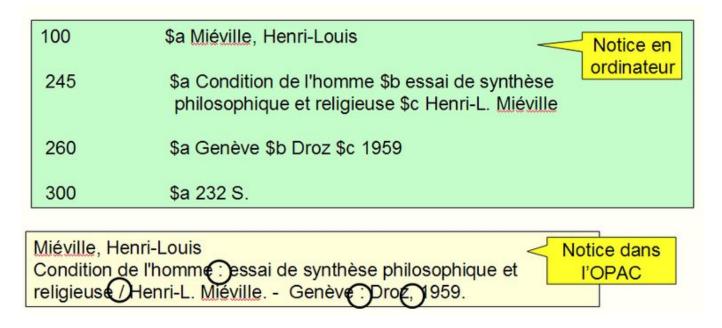
9) IDS-MARC

IDS (Informationsverbund Deutschschweiz) a choisi MARC21 pour garantir la meilleure compatibilité sur le plan national (RERO et Bibliothèque nationale) comme sur le plan international.

Mais IDS a aussi cherché à résoudre certains problèmes de MARC21; IDS a donc **adapté** MARC21 sur certains points, et le résultat de cette adaptation est appelé "IDS-MARC".

Par rapport à MARC21, IDS-MARC présente les caractéristiques suivantes:

a) la ponctuation ISBD n'est pas saisie, elle est ajoutée par programme lors de l'affichage



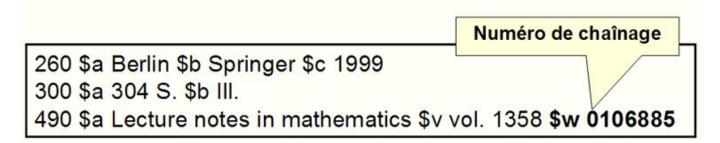
b) L'article (le déterminant) en début de titre est codé par <<...>>

Article en début de titre:	
MARC21	
245 1(3) \$a Le monde aujourd'hui	
<u> </u>	
Article en début de titre:	
IDS MARC	
245 \$a < <le>> monde des livres</le>	

c) La codification des non-nooks est faite en clair



d) Une technique simple permet de chaîner les notices entre elles



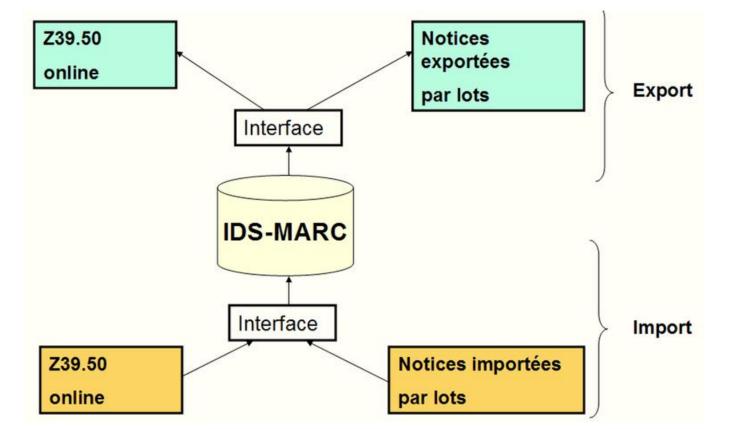
Cette technique est utilisée pour chaîner les volumes d'une collection à la notice de collection, ou les volumes d'une oeuvre en plusieurs volumes à la notice de niveau supérieur.

Résultat dans l'OPAC:

Lecture notes in mathematics	1 : Seminar über Funktionen-Algebren
Berlin : Springer	2 : Cohomologie des espaces localement comp
Früher mit Untertitel: an informal s	3 : Stable homotopy theory
Enthält 1971(1973)- : Ecole d'été	3 : Stable homotopy theory
Séminaire de probabilités ,	4 : Groups of homotopy classes
Non-commutative harmonic analy	4 : Groups of homotopy classes
Representation theory ,	5 : Cohomologie Galoisienne
representation theory,	

e) Compatibilité avec MARC21

La compatibilité entre IDS-MARC et MARC21 est garantie grâce à un programme-interface



Grâce à cette interface, les notices peuvent être échangées entre MARC21 et IDS-MARC sans perte ni d'information ni de codification; par exemple notice dérivée de RERO vers IDS:

Notice RERO en format MARC21

100 \$a Rebetez, Martine 245 13 \$a La Suisse se réchauffe : \$b effet de serre et changement climatique / \$c Martine Rebetez \$a Lausanne: \$b Presses polytechniques et universitaires romandes, \$c cop. 2002 300 \$a 140 p.; \$c 18 cm 4901 \$a Le savoir suisse \$v 2. \$a Nature et environnement 504 \$a Bibliographie (p. 137-140)

Conversion automatique de format de MARC21 à IDS-Marc

100 \$a Rebetez, Martine

245 \$a <<La>> Suisse se réchauffe \$b effet de serre et changement climatique \$c Martine Rebetez

260 \$a Lausanne \$b Presses polytechniques et universitaires romandes \$c cop. 2002

300 \$a 140 p.

490 \$a Le savoir suisse \$v 2

504 \$a Bibliographie (p. 137-140)

05.10.2009 22:34 17 sur 26

10) UNIMARC

A l'origine d'UNIMARC, il y a la volonté (européenne!) de définir un format qui soit meilleur que MARC et qui soit vraiment international. UNIMARC bénéficie d'un soutien de la FIAB.

Site Web de référence pour UNIMARC

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

UNIMARC est utilisé surtout en Europe (France bibliothèques universitaires, Italie, Portugal, etc.)

Exemple de notice UNIMARC, simplifiée:

```
001 83.041470.3
010 ## $a 3515023550$bkart.$d DM46.0 0
020 ## $a DE $b 88,A22,0260
100 ## $a 19801020d1976####|||y0gery01#####ba
101 0# $a eng
102 ## $a DE
200 1# $a NSB The NSE phonology of Old High
German $f by Joseph B. Voyles
210 ## $a Wiesbaden $c Steiner $d 1976
215 ## $a XII, 323S. $c 1 Kt. $d 24cm
225 0# $a Zeitschrift fur Dialektologie und
Linquistik / Beihefte
410 #1 $1 54011 $a Zeitschrift fur Dialektologie
und Linguistik / Beihefte
606 ## $a Phonologie
606 ## $a Althochdeutsche $x o. a. Sprache
700 #1 $a Voyles, $b Joseph B.
```

UNIMARC est un excellent format!

- codification logique et cohérente
- pas de ponctuation à saisir
- mécanisme pour le chaînage des notices
- etc.

Cependant, développé 10 plus tard que MARC, UNIMARC n'a pas réussi, malgré ses qualités, à s'imposer sur le plan international.

11) MAB2

MAB = Maschinelles Austauschformat; MAB2 = 2e version

A l'origine de MAB, il y a la volonté (allemande!) de définir un format qui soit meilleur que MARC et qui soit moins lourd en temps machine.

Site Web de référence pour MAB

http://www.d-nb.de/standardisierung/formate/mab.htm

MAB est utilisé quasi exclusivement en Allemagne et en Autriche.

```
001 $a FU01515567
025a $a 95.003401.0
100 $a Gates, Bill $9 FU00600447
304 $a <The> road ahead
331 $a < Der > Weg nach vorn
335 $a die Zukunft der Informationsgesellschaft
359 $a Bill Gates. In Zusammenarbeit mit Nathan Myhrvold und Peter
  Rinearson. Aus dem Amerikan, von Friedrich Griese und Hainer Kober
403 $a Völlig überarb, und aktualisierte Taschenbuchausg.
410 $a München
412 $a Heyne
425 $a 1997
433 $a 459 S. : Ill., graph. Darst.
451 $a Heyne-Bücher: 19: Heyne-Sachbuch; 521
454c $a Heyne-Bücher / 19
455 $a 521
540a $a 3-453-12296-8
574 $a 97,A18,0762
```

L'Allemagne et l'Autriche ont décidé récemment d'abandonner le format MAB comme format d'échange (tant sur le plan national qu'international), et d'adopter MARC21.

Umstieg auf MARC21 (Passage à MARC21)

http://www.d-nb.de/standardisierung/formate/marc21.htm

12) Format interne / Format d'échange

Le format interne (codification des données dans la base de données d'une bibliothèque) n'est pas forcément le même que le format d'échange. Exemple de la Deutsche Nationalbibliothek:

- format interne: un format spécifique, le format PICA
- formats en exportation: à choix MAB2, MARC21 ou UNIMARC

```
0100 027252086
0500 Aau
1100 1980
1500 /1de
1700 /1de
2000 3-489-60534-9
2065 801020409
2100 80A080107
2105 79N390243
2190 1116/23P
3000 Rupert@Riedl
3010 Robert@Kaspar [Mitarb.]
4000 @Biologie der Erkenntnis : die stammesgeschichtlichen
Grundlagen der Vernunft / Rupert Riedl. Unter Mitarb. von Robert
Kaspar
4030 Berlin [u.a.] : Parey
4060 230 S
4061 Ill
4201 Literaturverz. S. 215 - 223
5050 03a ; 18a
5301 !106405799!42.88 Physische Anthropologie
5302 !106420259! 42.65 Tiergeographie
5550 !104591862!Evolutionäre Erkenntnistheorie
5550 !104755482! Humanbiologie
5550 !106098691! Evolution
```

05.10.2009 22:34 20 sur 26

13) L'avenir de MARC21

HTML

HTML est un langage de balises; c'est la méthode universelle pour codifier une **page Web**. Tous les navigateurs savent comment interpréter une page HTML. Toutes les balises HTML sont prédéfinies sur le plan international, on n'a aucune liberté.

```
<!-- Titelbalken -->
<table border=0 cellpadding=5 cellspacing=0 width="100%"
style="margin-top:5pt">
<font face="Arial,Helvetica">IDS Pace!/Ports (font) (font)
                                = tableau
size=+2>Exemplarstatus: Liste der v
</font>
                                = nouvelle ligne dans le
tableau
 = cellule
 
                               etc. etc.
<!-- End Titelbalken -->
```

Exemples de balises HTML: ...

XML

XML est aussi un langage de balises, pour codifier non pas des pages Web, mais des **données**. Dans XML, aucune balise n'est prédéfinie, on dispose d'une liberté totale. Mais on doit suivre des règles strictes de syntaxe.

```
<?xml version="1.0" encoding="UTF-8"?>
<MyPhotoArchive>
<photo title="Walking on the Beach at Sunset"</pre>
filename="http://home.earthlink.net/~banerjek/i
    <name>Keiko</name>
  </dogs>
  <people>
   <name>Banerjee, Kyle</name>
  </people>
  <place>
   <country -state>OR</country-state>
   <city>Manzanita</city>
  </place>
  <date>1999-08-13
  </date>
</photo>
```

Exemples de balises XML: <name> ... </name>

Vous n'avez jamais entendu parler de XML? Alors cliquer ici.

MARC21 et XML

Le format MARC21 peut être présenté sous forme XML.

Soit sous la forme MODS, avec des balises alphabétiques, correspondant aux étiquettes de MARC21

```
<titleInfo>
<title>Arithmetic /</title>
</titleInfo>
<name type="personal">
<namePart>Sandburg, Carl,</namePart>
<namePart type="date">1878-1967.</namePart>
<role>creator</role>
</name>
<name type="personal">
```

Soit sous la forme MARCXML, avec des balises "MARC21 original"

```
/datafield>
-<datafield tag="260" ind1=" " ind2=" ">
         <subfield code="a">San Diego :</subfield>
         <subfield code="b">Harcourt Brace Jovanovich.</subfield>
         <subfield code="c">c1993.</subfield>
</datafield>
-<datafield tag="300" ind1=" " ind2=" ">
         <subfield code="a">1 v. (unpaged) :</subfield>
         <subfield code="b">ill. (some col.);</subfield>
         <subfield code="c">26 cm.</subfield>
</datafield>
```

Il n'y a donc pas incompatibilité entre MARC21 et XML. MARC21 peut être présenté sous forme XML, que ce soit avec la solution MODS ou avec la solution MARCXML.

Pour un exemple concret d'utilisation de MARCXML, cliquer ici.

05.10.2009 22:34 23 sur 26

En fait il ne faut pas confondre

- l'analyse logique des éléments à codifier (titre, sous-titre, etc.): sur ce point, MARC21 est excellent (mais peut être amélioré sur certains points)
- la méthode de codification MARC (étiquette 245, sous-zone \$a, etc.): cette méthode est tout à fait spécifique au monde des bibliothèques; il est normal d'étudier des alternatives
- le formatage ISO 2709: c'est une technique effectivement obsolète, qui fera sans doute place à XML

Les critiques les plus fréquentes contre MARC21

- Beaucoup de points de détail (article, ponctuation ISBD, etc.): il existe des solutions à ces problèmes, dans le cadre de MARC21
- Format inadapté aux ressources numériques: en partie vrai
- Technique de codification exotique: vrai
- Trop compliqué, trop détaillé: en partie vrai
- Marc conçu seulement pour imprimer des fiches: faux

Quelle évolution? Il faut étudier divers scénarios.

Voir par exemple MARC scenarios

http://futurelib.pbwiki.com/MARC-Scenarios:

- Etendre MARC21 en utilisant MARCXML
- 2. Ou: Etendre MODS pour en faire un nouveau format
- 3. Ou: Créer un nouveau format

Probablement: stabilité et continuité

Evolution lente de MARC21 dans le sens XML

Abandon progressif de ISO 2709

Pourquoi une évolution lente? Il y a des centaines de millions de notices MARC21 dans le monde, et il y a des dizaines de milliers de système automatisés installés. Tout changement radical est quasiment impossible. Mais la notice de catalogage classique, en MARC21, sera certainement de plus en plus complétée par des informations non-MARC.

05.10.2009 22:34 24 sur 26

L'avenir est déjà présent: le CERN CDS

La même notice est convertie dynamiquement en divers formats!

CERN CDS = http://cdsweb.cern.ch/



Autre exemple: Library of Congress

http://lccn.loc.gov/93182445

LIBRARY OF CONGRESS CATALOG RECORD



View LC holdings for this title in the: LC Online Catalog View this record in: MARCXML | MODS | Dublin Core

Le bonheur d'être Suisse

LC Control No.: 93182445

Type of Material: Book (Print, Microform, Electronic, etc.)

Personal Name: Ziegler, Jean, 1934- » More like this

Main Title: Le bonheur d'être Suisse / Jean Ziegler.

Published/Created: Paris: Seuil: Fayard, c1993.

Description: 312 p.; 24 cm.

ISBN: 2020201461:120F

5) Formats de catalogage et métadonnées

5.2) Métadonnées

Situation classique:

On a des notices de catalogage

- faites par des catalogueurs professionnels
- dans des bibliothèques
- selon des normes de catalogage

Ces notices sont codifiées avec les formats classiques

- MARC21 (et variantes, par ex. IDS-MARC)
- UNIMARC

Nouvelle situation:

On a de manière plus générale des métadonnées

- faites par des non bibliothécaires
- en dehors des bibliothèques
- souvent sans normes de catalogage

Ces métadonnées décrivent **toutes sortes de documents et d'objets** (vidéos, sites Web, travaux de recherche, full text, objets, tableaux, etc.).

Ces métadonnées ne sont pas codifiées avec les formats "classiques", mais avec de **nouveaux formats**, par exemple Dublin Core, RDF, MODS, METS, etc. etc.

Ces nouveaux formats, très nombreux, sont souvent plus simples que les formats classiques; ils sont adaptés au Web et aux documents électroniques.