Bibliothèque numérique de l'enssib

Il y a une vie après MARC, 19 novembre 2007 : journée d'étude en hommage à Pierre-Yves Duchemin

---

De MARC à XML : Les nouveaux formats bibliographiques

CLAVEL, Thierry

CLAVEL, Thierry. De MARC à XML : Les nouveaux formats bibliographiques. In *Il y a une vie après MARC : journée d'étude organisée en hommage à Pierre-Yves Duchemin, l'enssib à Villeurbanne, 19 novembre 2007* [en ligne]. Format PDF.

Disponible sur : <http://www.enssib.fr/bibliotheque-numerique/notice-1333>

Ce document est diffusé sous licence « **Creative Commons by-nc-nd** ». Cette licence signifie que le document est mis à disposition selon le contrat **Paternité-Pas d'Utilisation Commerciale-Pas de Modification**, disponible en ligne à l'adresse http://creativecommons.org/licenses/by-nc-nd/2.0/fr/ Il est ainsi possible de reproduire, distribuer et communiquer cette création au public, à condition de le faire à titre gratuit, mais ni de le proposer à titre onéreux ni le modifier sans le consentement explicite de l'auteur.

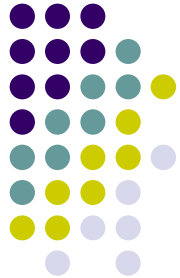
L'ensemble des documents mis en ligne par l'enssib sont accessibles à partir du site : http://www.enssib.fr/bibliotheque-numerique/

De MARC à XML : les nouveaux formats bibliographiques

En hommage à Pierre-Yves Duchemin

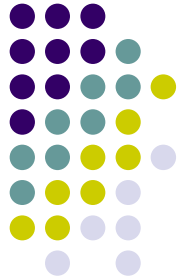
Thierry Clavel (enssib)

Origine des formats MARC



- MARC I : 1965 pour la production de fiches papier
- MARC II : 1968 pour la production de notices informatisées
- Contexte :
  - Explosion documentaire des années 1960-70
  - Contrôle bibliographique universel (CBU)
  - Création de l' ISBD
- Objectif de MARC :
  - Diffuser et échanger de l'information bibliographique

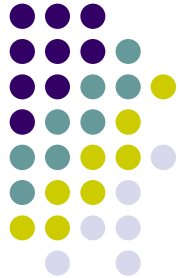
Caractéristiques des formats MARC



- Un contenu très spécialisé :
  - Information bibliographique calquée sur l'ISBD
  - Pas (ou peu) de données de gestion

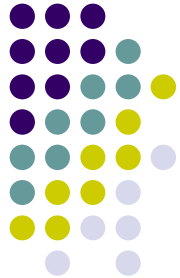
- Des utilisateurs peu nombreux :
  - Bibliothèques
  - Métiers de la chaîne éditoriale
  - Fournisseurs de SIGB

Structure des formats MARC



- structuration sur deux niveaux
  - Zones de données de type numérique (étiquette)
  - sous zones
- Correspondance stricte avec les normes de catalogage
  - ISBD
  - AACR puis RDA pour les formats anglo-américains (US MARC, MARC 21)
- Structuration normalisée
  - Norme ISO 2709 : échange
  - Z 3950 (ISO 23950) : protocole de requête associé

Evolution des formats MARC



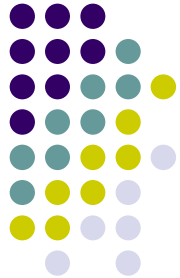
- En 40 ans, les formats MARC ont peu évolué :
  - même type de contenu
  - même structuration à deux niveaux
  - mêmes utilisateurs : les bibliothèques

- Qu'est ce qui a changé : Rien fondamentalement
  - De nouvelles zones et sous-zones pour s'adapter aux besoins des bibliothèques
  - Une tendance à structurer plus finement que par le passé
  - Davantage de données codées

Quels sont les besoins de nos catalogues aujourd'hui ?

- Le Web, les moteurs de recherche et l'arrivée massive de documentation électronique ont radicalement transformé les usages en termes de recherche documentaire
- Aujourd'hui nos catalogues
  - Manquent de visibilité sur le Web
  - Souffrent de la comparaison avec l'offre commerciale (Google, Amazon, etc.) : les notices sont peu ou pas enrichies
  - Ne sont pas assez tournés vers l'usager
    - Absence de recherche fédérée
    - Absence de Web services (Web 2.0)

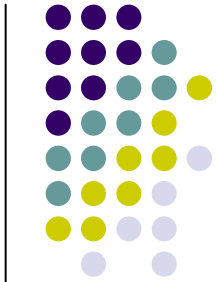
Que manque t-il aux formats MARC ?



- Structuration des données
  - Il faudrait davantage de niveaux hiérarchiques
- Visibilité sur le Web et interopérabilité des données
  - MARC et les normes qui lui sont associées ne sont pas des standards du Web :
    - ISO 2709 : outils spécifiques
    - Z3950 : non exploitable avec HTTP (nécessité d'une passerelle Web)

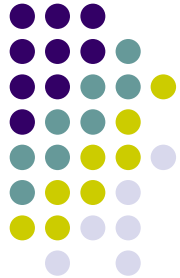
    MARC contribue à isoler les catalogues
- Données de gestion
  - Non prévu avec MARC mais pourtant indispensable avec la documentation numérique :
    - Droits
    - Archivage
    - Conservation

Que manque t-il aux formats MARC ? (2)



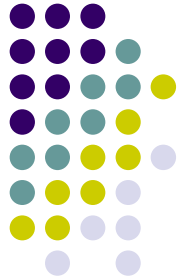
- Services aux usagers
  - Recherche fédérée :
    - Possible avec Z3950 mais lourde et coûteuse à mettre en place
  - Enrichissement de notices :
    - MARC en permet certains (Lien hypertexte en zone 856, etc.
  - Web services : nécessite des adaptations importantes

MARC peut il faire évoluer les catalogues ?



NON, car les formats MARC eux mêmes peuvent difficilement évoluer

Quels outils pour succéder à MARC ?



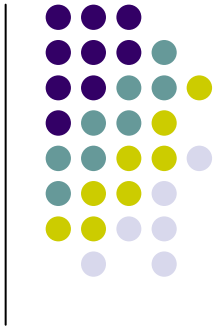
XML (eXtensible Mark up Language)

- Un langage à balises structuré issu de SGML :
- Un métalangage : langage générique pour établir des modèles de documents (DTD et schémas)
- avantages d'XML :
  - Un standard du Web :
    - Ouverture :
    - interopérabilité et visibilité
  - Structuration :
    - hiérarchisation théorique illimitée,
    - Extensibilité : adaptation aux besoins
  - fonctionnalités : potentiel important :
    - RSS, OAI PMH, etc.

Formats MARC / Formats XML

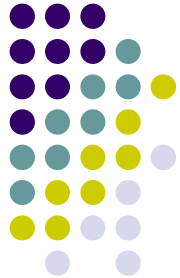
Structuration logique : Étiquette et sous-zone	Balises ouvrantes et fermantes.imbrication d'éléments dans une arborescence logique
200 $aLe nom de la rose	<titre>Le nom de la Rose</titre>
MARC précise les éléments par des indicateurs	XML précise les éléments par des attributs
200 1# $a Le nom de la Rose	<titre langue="it">Il nome della Rosa</titre>

Exemple très simple d'arborescence XML



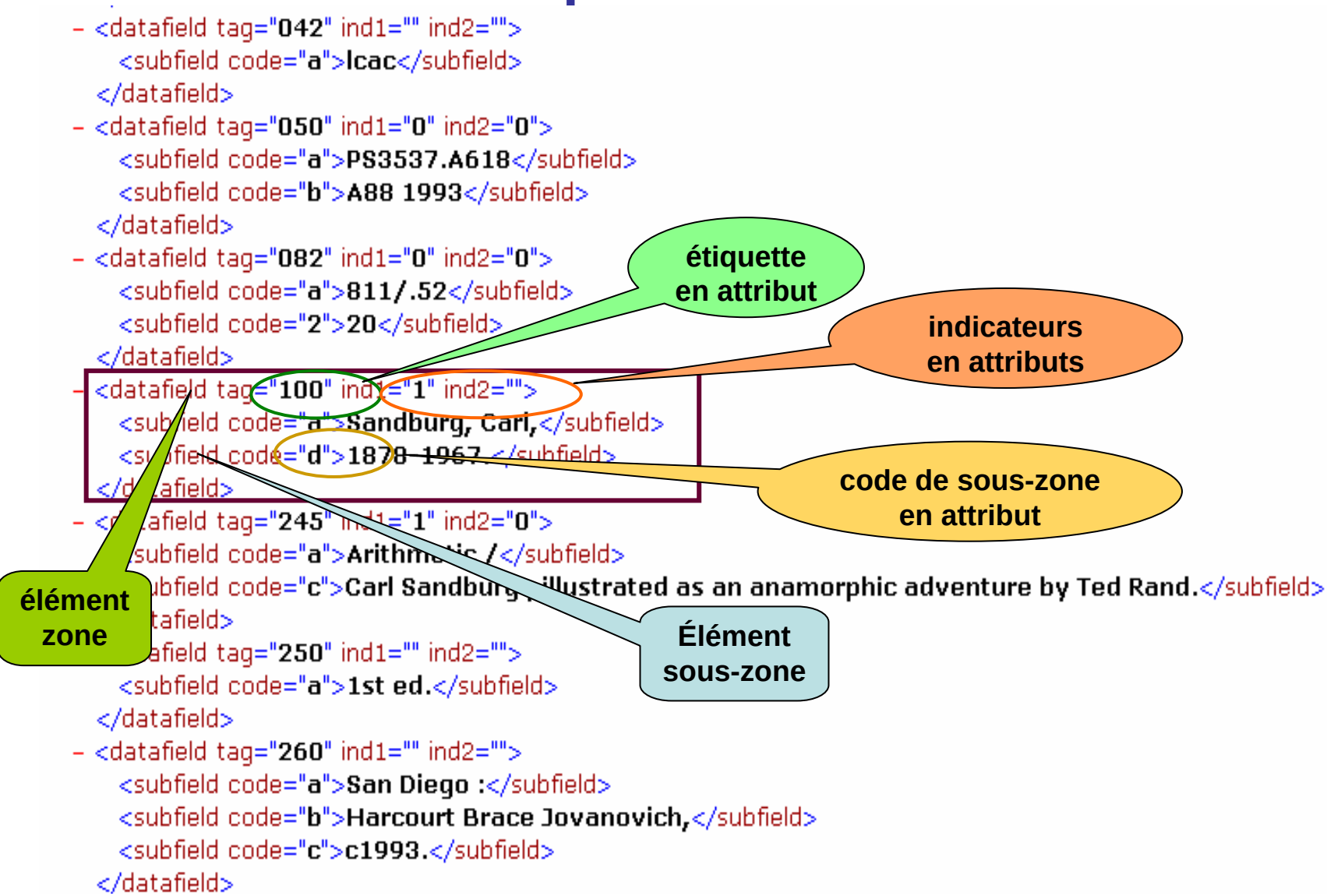
```
<biblio>
        <auteur>
                <nom>…</nom>
                <prénom>…</prénom>
        </auteur>
        <titre>
                <titre propre>…</titre propre>
                <complément de titre>…</complément de titre>
        </titre>
</biblio>
```

formats XML basés sur MARC
MARC XML



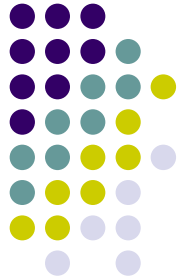
- Schéma XML créé en 2000 et maintenu par la Bibliothèque du Congrès.
- habillage en XML de MARC 21 avec ses étiquettes, ses indicateurs, ses codes de sous-zones.
  - Zones et sous-zones : éléments XML
  - indicateurs et codes de sous-zones : attributs XML
- à quoi sert MARC XML?
  - convertir des notices MARC 21 en XML

Exemple MARCXML



Source: F. Leresche, 2007

Les formats XML basés sur MARC : MARCXchange

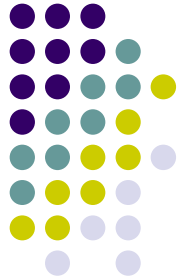


- Schéma XML devenu une norme ISO 25577 en 2006
- habillage MARC en XML sur le même principe que MARC XML
- A quoi sert MARCXchange ? :
  - à convertir des notices de différents formats MARC en XML (UNIMARC, etc.)
- Exemple : Echanges de notices du Sudoc vers Google Scholar

extraits d'une notice UNIMARC en MARCXchange

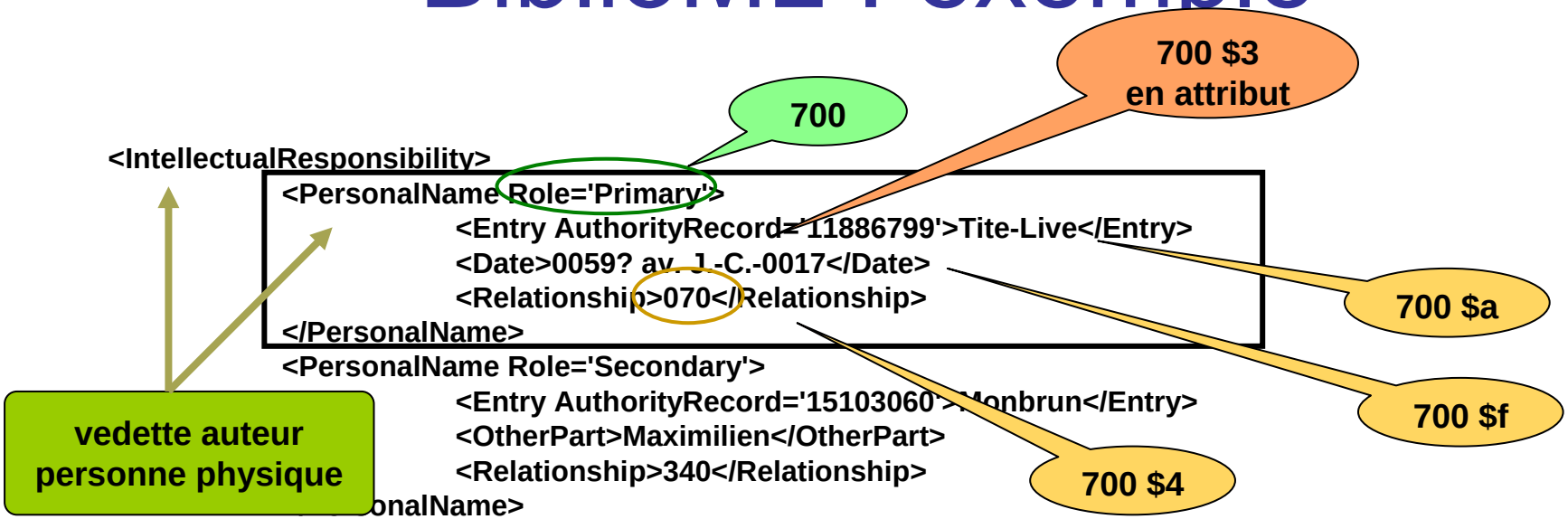
```
<?xml version="1.0" encoding="windows-1251"?>
<collection xmlns="http://www.bs.dk/standards/MarcXchange" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.bs.dk/standards/MarcXchange http://www.bs.dk/standards/MarcXchange.xsd">
    <record format="UNIMARC" type="Bibliographic">
        <leader>01307nam0 2200349 i 450 </leader>

    <datafield tag="105" ind1=" " ind2=" ">
        <subfield code="a">y   ||||||||</subfield>
    </datafield>
    <datafield tag="200" ind1="1" ind2=" ">
        <subfield code="a">The Liberal Party in rural England, 1885-1910</subfield>
        <subfield code="e">radicalism and community</subfield>
        <subfield code="f">Patricia Lynch</subfield>
    </datafield>
```

Les formats XML basés sur MARC Biblioml (bibliographic mark up language)



- DTD XML basée sur l'UNIMARC créée en 1999 par le Ministère de la culture français
  - plus de maintenance
  - Utilisation de mots à la place des étiquettes et des codes de sous zones
- à quoi sert Biblioml ?
  - - convertir en XML des données bibliographiques et d'autorités initialement saisies en Unimarc.
  - - créer directement en XML des notices complètes conformes à l'ISBD
- Exemple d'utilisation :
  - Bibliographie nationale française

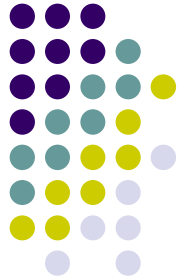
BiblioML : exemple




```
<IntellectualResponsibility>
        <PersonalName Role='Primary'>
                <Entry AuthorityRecord='11886799'>Tite-Live</Entry>
                <Date>0059? av. J.-C.-0017</Date>
                <Relationship>070</Relationship>
        </PersonalName>
        <PersonalName Role='Secondary'>
                <Entry AuthorityRecord='15103060'>Monbrun</Entry>
                <OtherPart>Maximilien</OtherPart>
                <Relationship>340</Relationship>
        onalName>
</IntellectualResponsibility>
<PublicationGroup>
        <CountryOfPublication><CountryCode Country='FR'/></CountryOfPublication>
        <Publication>
        <Publisher><Place>Nantes</Place><Name>Éd. du Temps</Name></Publisher>
        <Manufacture>
        <Manufacturer><Place>52-Langres-Saints-Geosmes</Place><Name>Impr. IDG</Name></Manufacturer>
        </Manufacture>
        <Date>impr. 2003</Date>
        </Publication>
</PublicationGroup>
<PublicationDate>
        <MonographDate Status='Normal'><Year>2003</Year></MonographDate>
</PublicationDate>
<PhysicalDescription>
        <PhysicalItem>
        <ItemDescription><Extent>1 vol. (95 p.)</Extent><OtherPhysicalDetails>couv. ill. en coul.</Other
        PhysicalDetails><Dimensions>25 cm</Dimensions></ItemDescription>
```

Source : F Leresche, BnF

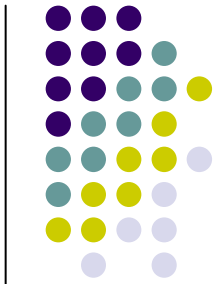
formats XML basés sur MARC MODS (Metada Object description schema)



- Schéma XML de métadonnées descriptives créé par la LC en 2002, basé sur **MARC 21 simplifié**

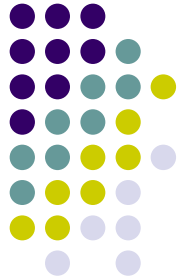
- MODS est un sous-ensemble de MARC 21
  - 19 éléments principaux, qui se subdivisent eux-mêmes en plusieurs sous-éléments.
  - Comme Biblioml, les noms de balises sont signifiants et non plus numériques
- A quoi sert MODS ?
  - à réutiliser une partie d'une notice déjà codée en format MARC21
  - à créer intégralement des notices en XML
- Un schéma XML similaire développé par la LC pour les autorités existe depuis 2005 : **MADS**

Exemple MODS pour une carte : extraits




```
<mods version="3.0"
    xsi:schemaLocation="http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
− <titleInfo>
<title>Campbell County, Wyoming /</title>
</titleInfo>
−<name type="corporate">
<namePart>Campbell County Chamber of Commerce</namePart>
</name>
<typeOfResource>cartographic</typeOfResource>
<genre authority="marc">map</genre>
−<originInfo>
<place>
<placeTerm type="text">[Gillette, Wyo.]</placeTerm>
</place>
<publisher>Campbell County Chamber of Commerce</publisher>
<dateIssued>[1982?]</dateIssued>
```

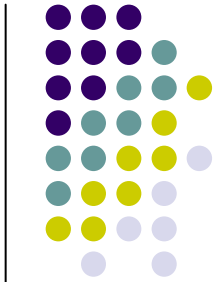
Jeux de Métadonnées indépendants de MARC : Dublin core (DC)



- Jeu de métadonnées créé en 1995
  - norme ISO 15836 depuis 2003, maintenu par le DCMI
  - Objectif de départ : Avoir un ensemble de métadonnées simple utilisables par tous pour décrire des ressources en ligne et améliorer leur recherche sur le Web
- DC simple : 15 éléments de base tous facultatifs et répétables (Title, Description, Creator, Rights, identifier, etc.)
- DC qualifié : une cinquantaine d' attributs validés par le DCMI qui peuvent qualifier les éléments de base

- DC est devenu un noyau de métadonnées de référence.
  - DC est de plus en plus utilisé par les bibliothèques numériques pour exposer leurs données en OAI
    - seul format de métadonnées obligatoire dans le protocole OAI PMH.

Exemple de notice XML en Dublin Core destinée à un entrepôt OAI



```
- <record>
- <header>
  <identifier>oai:bnf.fr:catalogue/ark:/12148/cb300007082/description</identifier>
  <datestamp>2007/10/22</datestamp>
  <setSpec>catalogue:collections:d1</setSpec>
  <setSpec>catalogue:imprimes:monographies</setSpec>
  </header>
- <metadata>
- <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
  <dc:identifier>http://catalogue.bnf.fr/ark:/12148/cb300007082/description</dc:identifier>
  <dc:title>Un Procès de cloches à Metz. (Signé : Ch. Abel.)</dc:title>
  <dc:creator>Abel, Charles (1824-1894)</dc:creator>
  <dc:publisher>impr. de Rousseau-Pallez (Metz)</dc:publisher>
  <dc:date>1858</dc:date>
  <dc:language>fre</dc:language>
  <dc:type xml:lang="fre">texte imprimé</dc:type>
  <dc:type xml:lang="eng">printed text</dc:type>
  <dc:type xml:lang="eng">text</dc:type>
  <dc:rights xml:lang="fre">Catalogue en ligne de la Bibliothèque nationale de France</dc:rights>
  <dc:rights xml:lang="eng">French National Library online Catalog</dc:rights>
  </oai_dc:dc>
  </metadata>
  </record>
```

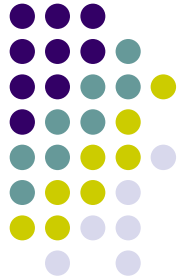
Balises en Dublin core simple

Source : BnF, novembre 2007

Formats XML de métadonnées indépendants de MARC

- **ONIX** = ONline Information eXchange. 1re version de la DTD publiée en 2000
  - Gérée par le groupe EDItEUR
  - 2 versions : (ONIXfor books) et (ONIX for serials)
  - Plus de 200 éléments dans ONIX for books dont certains obligatoires(ISBN, titre…)
  - A quoi sert ONIX ?
    - Proposer une description bibliographique complète et enrichie aux clients des éditeurs,
    - Fournir des métadonnées de gestion pour les libraires
    - Exemple d'utilisation : Electre
- **EAD** = Encoded Archival Description Publiée en 1998
  - Maintenue par la Bibliothèque du Congress et Society of American Archivists S'appuie sur ISAD(G)
  - Utilisée par la BnF et l'ABES (Calames)pour les manuscrits

Formats XML de métadonnées indépendants de MARC (2)



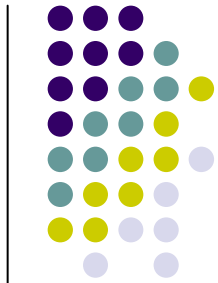
- **METS (Metadata Encoding and Transmission Standard)**
  - Schéma XML maintenu par la Library of Congress
  - une enveloppe XML qui sert à relier ensemble différents blocs de métadonnées pour décrire des objets numériques complexes.
    - un fichier METS est constitué de sept blocs (sections) parmi lesquels :
      - métadonnées descriptives,
      - métadonnées de gestion (techniques, source, droits, provenance)
      - carte de structure
      - Etc.
  - A quoi sert METS ?
    - A gérer complètement un document électronique et sa structure
      - Feuilletage de documents électroniques, consultation d'articles de presse numérisés
    - A conserver des documents numériques
    - Exemples d'utilisation :
      - TEF (Thèses électroniques françaises),
      - Numérisation des manuscrits de la BnF

Les protocoles de requête

- **SRU** =Search retrieval via URL
  - Protocole de requête qui s'appuie sur XML
  - Développé pour l'environnement Internet par des experts en Z3950.
  - A quoi sert SRU ?
    - évolution de Z39 50 pour le Web

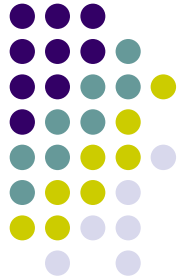
- **OAI PMH** = Open Archive Protocol for Metadata Harvesting
  - protocole pour le moissonnage des métadonnées qui s'appuie sur HTTP et XML ; Dernière version : juin 2002
  - A quoi sert OAI ?
    - A exposer ses métadonnées : les rendre visibles

Conclusion



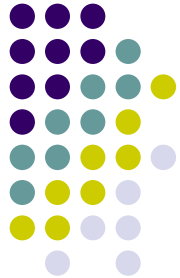
- En 2007
  - Description, échange, dérivation : les formats MARC sont toujours très utilisés.
  - Exposition des données sur le Web fait de plus en plus appel à XML (MARxchange, etc.)
  - Bibliothèques numériques : fonctionnent déjà avec XML
  - Si les SIGB le permettaient on pourrait dés maintenant abandonner MARC, recevoir et créer directement des notices en XML
    - Electre en ONIX
    - BnF en EAD

Conclusion
Avenir des catalogues



- **Changer de formats** pour s'ouvrir au Web et offrir davantage de services aux usagers
- **Garder les fonctions premières du catalogue** fortement liées aux missions des bibliothèques.
  - identification,
  - recensement et localisation des documents
- **Maintenir la qualité du catalogue** : gage de pérennité :
  - les normes de catalogage évoluent, : elles s'appliquerons demain à XML aussi bien quelles se sont appliquées hier aux formats MARC.

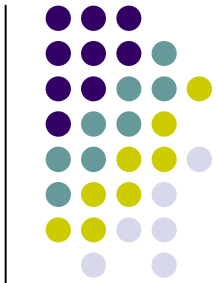
Conclusion : Avenir d'XML



- *L'évolution qui se profile : Le **Web sémantique** basé non plus sur des documents mais sur des données*

- Dans cette perspective, c'est RDF (Resource description framework) cadre intellectuel de description des ressources qui est le mieux placé pour prendre le relai des formats XML

Liens utiles



- Normes et standards maintenus par la Bibliothèque du Congrès :
  - MARC 21, MARCxml, MODS, METS, Z3950, SRU http://www.loc.gov/standards/
  - MarcXchange : http://www.loc.gov/standards/iso25577/
- ONIX (DTD en français) : http://www.editeur.org/ONIX2.0TradFr.Guide.pdf

- Dublin core
  - Site officiel : http://dublincore.org/
  - Guide d'utilisation en français : http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm
- Protocole OAI PMH : http://www.openarchives.org/OAI/openarchivesprotocol.html
- RDF : http://www.w3.org/RDF/