

Présentation des standards : le 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 :

Nous rappellerons le contexte historique de l'initiative « Dublin Core », puis verrons l'ensemble d'éléments de base et la notion de « Dublin Core qualifié ». Nous verrons ensuite que la description bibliographique est indépendante des formats physiques et présenterons quelques unes des représentations physiques les plus utilisées.

Mots clés : dublin core, métadonnées, indexation, web, archives ouvertes

Rappels :

Le Dublin Core n'est pas une norme. Il s'agit d'un ensemble de recommandations qu'on peut qualifier de standard de fait. Le Dublin Core est né en 1995, lors d'une conférence tenue à Dublin (Ohio). Il s'agissait de produire une recommandation pour la description bibliographique des documents ou objets se trouvant sur le web, dans une bibliothèque numérique ou même ailleurs. Les acteurs de cette première conférence, qui regroupait des personnes issues de la communauté des bibliothèques, des informaticiens et du web, ont spécifié un ensemble de quinze éléments de base, une sorte de dénominateur commun de la description bibliographique. Le Dublin Core, dans son vocabulaire propre, parle d'éléments et de termes. Eléments, parce que nous sommes ici dans un monde issu de SGML, celui de HTML (et plus tard XML). Nous parlons donc d'éléments et d'attributs là où MARC parle par exemple de zones ou de sous-zones et où d'autres encore pourraient parler de « champs ». Il est établi que les quinze éléments du Dublin Core sont « gravés dans le marbre » et que tout élément supplémentaire appartient aux extensions.

Les quinze éléments de base :

Nom de l'élément en français	Décrit	Nom de l'élément en anglais
Titre Sujet Description Source Langue Relation Couverture	contenu	Title Subject Description Source Language Relation Coverage
Créateur Éditeur Contributeur Droits	propriété intellectuelle	Creator Publisher Contributor Rights
Date Type	instance particulière	Date Type

Format Identifiant		Format Identifiant
-----------------------	--	-----------------------

Le Dublin Core qualifié et les « termes »

Les quinze éléments de base sont considérés comme un dénominateur commun et dans la plupart des cas sont insuffisamment précis. Dès le départ, les éléments de base ont été étendus (ou précisés) par un ensemble d'autres termes, parfois appelés « qualifieurs ». Deux classes de « qualifieurs » sont reconnues :

- les « raffinements d'éléments » qui rendent plus spécifique le sens d'un élément
- et les « schémas d'encodage » ou vocabulaires contrôlés.

Les éléments supplémentaires et « raffinements d'éléments »

Le DCMI¹ a défini trois éléments supplémentaires et une trentaine de raffinements d'éléments.

Les éléments supplémentaires : **Audience**, **Provenance** et **RightsHolder** étendent l'ensemble des quinze éléments de base et les « element refinements » précisent le sens d'un élément existant : voici pour information, la liste complète, que l'on retrouve à l'url <http://dublincore.org/documents/dcmi-terms/> :

Element refinement	Element refined
Abstract	<u>description</u>
Access Rights	<u>rights</u>
Alternative	<u>title</u>
Available	<u>date</u>
Bibliographic Citation	<u>identifiant</u>
Conforms To	<u>relation</u>
Created	<u>date</u>
Date Accepted	<u>date</u>
Date Copyrighted	<u>date</u>
Date Submitted	<u>date</u>
Audience Education Level	<u>audience</u>
Extent	<u>format</u>
Has Format	<u>relation</u>
Has Part	<u>relation</u>
Has Version	<u>relation</u>
Is Format Of	<u>relation</u>
Is Part Of	<u>relation</u>
Is Replaced By	<u>relation</u>
Issued	<u>date</u>
Is Version Of	<u>relation</u>
License	<u>rights</u>
Mediator	<u>audience</u>
Medium	<u>format</u>
Modified	<u>date</u>
References	<u>relation</u>
Replaces	<u>relation</u>
Requires	<u>relation</u>
Spatial	<u>coverage</u>
Table Of Contents	<u>description</u>
Temporal	<u>coverage</u>

¹ Dublin Core Metadata Initiative, organisme responsable de la définition du Dublin Core

Valid	<u>date</u>
-------	-------------

Les schémas d'encodage (vocabulaires contrôlés)

Chaque élément, ou raffinement d'élément, peut disposer d'un ou plusieurs schémas d'encodage. Le DCMI, soucieux de ne pas réinventer la roue, fait référence à des schémas existants et en a inventé un, DCMIType, pour décrire les types (logiques et non physiques) d'objets, d'autres, DCMIBox et DCMIPoint, pour délimiter géographiquement un objet ou lieu et DCMIPeriod pour situer un événement dans le temps.

Nom du schéma (vocabulaire)	« Element refined »
Box (DCMI)	<u>spatial</u>
DCMI Type	<u>type</u>
DDC (Dewey Decimal Classification)	<u>subject</u>
IMT (Internet media type)	<u>format</u>
ISO 3166 (norme iso des noms de pays)	<u>spatial</u>
ISO 639-2 (norme iso pour les codes de langues)	<u>language</u>
LCC (Library of Congress Classification)	<u>subject</u>
LCSH (Library of Congress Subject Headings)	<u>subject</u>
MeSH (Medical Subject Headings)	<u>subject</u>
Period (DCMI)	<u>date / temporal</u>
Point (DCMI)	<u>spatial</u>
RFC 1766 (codes langues-pays)	<u>language</u>
RFC 3066 (codes langues-pays, remplace RFC 1766)	<u>language</u>
TGN (Getty Thesaurus of Geographic Names)	<u>spatial</u>
UDC (Universal Decimal Classification)	<u>subject</u>
URI (Uniform Resource Identifier)	<u>identifiant/ source/ relation</u>
W3C-DTF (dates et heures)	<u>date/temporal</u>

Le vocabulaire DCMI Type

Le vocabulaire DCMI Type contient une liste de différents types de ressources :

Collection	Groupe de documents (ou ressources)
Dataset	Base de données, liste, tableau
Event	Événement ponctuel (conférence...)
Image	Image, photo, (tous formats, générique)
Interactive Resource	Objet qui demande l'interaction de l'utilisateur
Moving Image	Image animée, vidéo...
Physical Object	Un objet lui-même (et non son image)
Service	
Software	Logiciel disponible pour installation sur une machine
Still Image	
Text	

Les formats physiques

Le Dublin Core ne dépend pas d'une expression physique précise. On peut l'exprimer en html ou XML/RDF² ou encore en XML, selon l'utilisation à laquelle il est destiné. Il existe trois recommandations officielles :

- **Expressing Simple Dublin Core in RDF/XML**
<http://dublincore.org/documents/2002/07/31/dcmes-xml/>
- **Guidelines for implementing Dublin Core in XML**
<http://dublincore.org/documents/2003/04/02/dc-xml-guidelines/>
- **Expressing Qualified Dublin Core in HTML/XHTML meta and link elements**
<http://dublincore.org/documents/dcq-html/>.

Et une recommandation proposée

- **Expressing Qualified Dublin Core in RDF/XML**
<http://dublincore.org/documents/2002/05/15/dcq-rdf-xml/>

Voici un exemple simple d'utilisation du Dublin Core dans un fichier html

```
<html>
  <head>
    <title> A Dirge </title>
    <link rel = "schema.DC" href = "http://purl.org/DC/elements/1.0/">
    <meta name = "DC.Title" content = "A Dirge">
    <meta name = "DC.Creator" content = "Shelley, Percy Bysshe">
    <meta name = "DC.Type" content = "poem">
    <meta name = "DC.Date" content = "1820">
    <meta name = "DC.Format" content = "text/html">
    <meta name = "DC.Language" content = "en">
  </head>
  <body><pre>
    Rough wind, that moanest loud
    Grief too sad for song;
    Wild wind, when sullen cloud
    Knells all the night long;
    Sad storm, whose tears are vain,
    Bare woods, whose branches strain,
    Deep caves and dreary main, -
    Wail, for the world's wrong!
  </pre></body>
</html>
```

² RDF : Resource description framework : recommandation du W3C.

Voici un exemple d'encodage en html, mais en utilisant les raffinements d'éléments (seul l'entête du document HTML est montré).

```
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Expressing Dublin Core in HTML/XHTML meta and link elements</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" lang="en" content="Expressing Dublin Core in HTML/XHTML meta
and link elements" />
<meta name="DC.creator" content="Andy Powell, UKOLN, University of Bath" />
<meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2003-11-01" />
<meta name="DC.identifier" scheme="DCTERMS.URI"
content="http://dublincore.org/documents/dcq-html/" />
<link rel="DCTERMS.replaces" hreflang="en"
href="http://dublincore.org/documents/2000/08/15/dcq-html/" />
<meta name="DCTERMS.abstract" content="This document describes how qualified Dublin
Core metadata can be encoded in HTML/XHTML &lt;meta&gt; elements" /> <meta
name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
</head>
```

L'exemple le plus courant d'expression du Dublin Core en XML est celui d'un serveur OAI³ (le schéma XML « oai_dc » se trouve sur le site de l'initiative OAI) voici un exemple extrait d'un serveur OAI.

```
<metadata>
<oai_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc
http://www.openarchives.org/OAI/2.0/oai-dc.xsd" xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:creator>Moret, M.</dc:creator>
<dc:title>Notice nécrologique : Émile Cotton (1872--1950)</dc:title>
<dc:date>1949</dc:date>
<dc:identifiant>AIF_1949_1_1_0</dc:identifiant>
<dc:identifiant>http://www.numdam.org/item?id=AIF_1949_1_1_0</dc:identifiant>
<dc:identifiant>oai:numdam.org:AIF_1949_1_1_0</dc:identifiant>
<dc:identifiant>citation: Ann. Inst. Fourier 1, 1-4 (1949)</dc:identifiant>
</oai_dc:dc>
</metadata>
```

³ Voir « Etat de L'art de approche critique : du Marc au Dublin Core »

Précisons pour terminer que :

- Le Dublin Core (ses extensions) évolue assez souvent. Les documents cités en référence peuvent changer.
- Le fait que la description logique n'est pas liée à un format physique fait qu'il est possible d'inventer d'autres expressions de celle-ci sans en changer le sens sémantique.
- Une des premières idées des acteurs du Dublin Core était d'améliorer la qualité des résultats de recherche des « moissonneurs » web. Le Dublin Core est/était destiné au moissonnage par un moteur de recherche « web » tel que google, yahoo ou encore un moteur plus spécialisé. Dans un premier temps, c'est cette utilisation spécialisée qui a prévalu, car dans les faits les moteurs web généralistes ont ignoré le Dublin Core (ou plus précisément ont ignoré l'utilisation d'éléments META dans l'entête des documents html, à cause du détournement à des fins commerciaux de leur utilisation). Aujourd'hui cependant, le projet « google scholar » préconise l'utilisation du Dublin Core.
- On peut dire encore aujourd'hui que le Dublin Core est moissonné efficacement aujourd'hui avant tout par le biais d'OAI, à partir de serveurs se conformant au protocole OAIPMH.
- Actuellement, si tous les serveurs OAI utilisent (c'est obligatoire) le schéma oai_dc, bien peu se sont lancés dans l'utilisation de schémas « Dublin Core qualifié ». L'avenir nous dira si cette tendance s'inversera.
- Les métadonnées « dublin core » sont très souvent exportées d'une base de données ou application autre, dont les métadonnées sont plus précises ou plus spécialisées.

Références :

Hillman, Diane, *Guide d'utilisation du Dublin Core*

<http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>

Dublin Core Metadata Initiative Documents : <http://dublincore.org/documents/>

DCMI Metadata Terms: <http://dublincore.org/documents/dcmi-terms/>

DCMI Encoding Guidelines <http://dublincore.org/resources/expressions/>

DCMI Schemas <http://dublincore.org/schemas/>