

Projet Metadonnées – Dublin Core

- Partie 1 -

Proposez une présentation synthétique du Dublin Core. Présentez les notions de raffinements et de vocabulaires pour Dublin Core.

- Définition -

Le Dublin Core est une norme de description des ressources d'information. Pour se faire, on utilise des métadonnées, littéralement des données décrivant d'autres données. Le plus souvent, ces métadonnées sont appliquées à des documents numériques (comme le contenu d'une page web), mais elles peuvent aussi décrire des documents papiers (comme une facture ou une étiquette de produit ; une notice bibliographique est un ensemble de métadonnées).

Il existe 2 types fondamentaux de métadonnées, qui se différencient par leur localisation :

- Tout d'abord, une métadonnée peut être interne, elle est alors encapsulée dans la ressource qu'elle décrit. C'est le cas des informations EXIF dans une photographie, ce sont des métadonnées qui se créent automatiquement au moment de la prise de vue.
- D'autre part, une métadonnée peut être externe. Les informations sont alors rédigées à part, en dehors de la ressource initiale. Par exemple, une notice numérique d'une monographie, dans une base de données.

- Utilisation -

Le Dublin Core a été conçu afin d'être utilisé par un nombre maximum de personnes, même non initiées, afin de rendre le web plus transparent. Ainsi, il est simple de création et de gestion et d'utilisation. Mais ce n'est pas la seule caractéristique ; il utilise un vocabulaire usuel, aisément compréhensible. Même si les éléments du Dublin Core étaient dans un premier temps seulement anglophones, aujourd'hui, de nombreuses versions sont disponibles. Ainsi, le DC a pour but d'être utilisé à l'échelle internationale. Enfin, les concepteurs du DC sont prévoyants, le schéma peut alors, en cas de besoin étendre ses éléments. C'est un « carcan ouvert » !

- Historique -

Le Dublin Core a été établi par un groupement international de professionnels provenant de disciplines liées telles que la bibliothéconomie, l'informatique, le balisage de textes ou la muséologie, et ce, depuis 1995. Le nom de "Dublin Core" a été adopté en référence au premier atelier de réflexion, tenu à Dublin, dans l'Ohio aux États-Unis.

- Comment ? -

La norme Dublin Core décrit les objets grâce à une structuration particulière. Les champs de métadonnées du Dublin Core sont dénommés « éléments ». Il comprend ainsi un ensemble de 15 éléments généraux. Pour l'exemple de la page web, le contenu peut ainsi être décrit de manière précise à travers de ces éléments. Techniquement, cela suppose l'inclusion de balises de métadonnées Dublin Core (DC) dans l'en-tête (head) de cette page.

Concrètement, chacun de ces éléments est facultatif et peut être répété. L'application du Dublin Core nécessite le recours à des vocabulaires spécialisés et surtout normalisés, pour renseigner les champs.

Chaque élément peut disposer d'un ou plusieurs schémas d'encodage, plus couramment appelés vocabulaire spécifique. Les 15 éléments sont les suivants : Titre, Sujet, Description, Source, Langue, Relation, Couverture, Créateur, Éditeur, Contributeur, Droits, Date, Type, Format, Identifiant.

Le Dublin Core possède également un ensemble de qualificatifs : les raffinements. Ils sont utilisés afin de préciser le sens d'un élément, sans en changer la signification, simplement la restreindre. Ils sont facultatifs.

- Partie 2 -

Fournissez pour chaque élément du Dublin Core : une courte définition en quelques mots, puis un exemple de valeur et lorsque cela est possible de décliner pour chaque raffinement (sauf pour les éléments DATE et RELATION pour lesquels il suffira de faire cette déclinaison sur le 1er raffinement et simplement citer les autres raffinements possibles).

Pour ma part, j'ai choisi de décrire le nouveau roman d'Anna Gavalda, comme exemple.

Élément (Identifiant)	Description, liste des raffinements
	Exemple
Titre (Title)	Il s'agit du titre principal du document, c'est le nom par lequel la ressource est connue. Il est possible de mettre une abréviation ou une traduction, c'est le raffinement « alternative ».
	<META NAME= « DC.Title » CONTENT= « La consolante »/> Mon exemple n'a pas d'alternative pertinente, mais pour un autre document, il ressemblerait à cela <META NAME= « DC.Title.Alternative » CONTENT= « ... »
Créateur (Creator)	Nom de la personne, de l'organisation, ou du service à l'origine de la création du contenu de la ressource.
	<META NAME= « DC.Creator » CONTENT= "Anna GAVALDA"/>
Sujet et mots-clé (Subject)	Le sujet est décrit par un ensemble mots-clé, de phrases de résumé ou par un code de classification. Afin de tendre vers une uniformisation, il est intéressant d'utiliser des vocabulaires contrôlés et de schémas formels tels que le vocabulaire MESH, la classification DEWEY, le codage LCSH..
	<META NAME= « DC.Subject » CONTENT= "Littérature, Généralités, Français"/> <META NAME= « DC.Subject.Classification » Scheme=«DDC» CONTENT "804"/>
Description (Description)	La description peut contenir un texte libre un résumé ou la table des matières. D'où les 2 raffinements : la table de matières (TableOfContent) et le résumé (Abstract).
	<META NAME= « DC.Description.Abstract » CONTENT= « Le quotidien de Charles, 47 ans. Sa vie, sa famille, son boulot, et soudain, le décès de la mère d'un de ses amis d'enfance. Il est désorienté mais épaulé et bien entouré par « ses deux femmes », il reprend goût à la vie. » Mon exemple ne contient pas de table des matières, mais il ressemblerait à cela : <META NAME= « DC..Description.TableOfContent » CONTENT= « I.xxxxx, II. Xxxxx... » />

Éditeur (Publisher)	<p>Nom de la personne, de l'organisation, ou du service à l'origine de la publication de la ressource dans sa forme actuelle. Le plus souvent, ce sont les maisons d'éditions qui sont ici citées, pour les monographies.</p> <p><META NAME= « DC.Publisher » CONTENT= « Le Dilettante »/></p>
Contributeur (Contributor)	<p>Personne physique ou morale qui contribue ou a contribué à l'élaboration du document</p> <p>Pas de contributeur, seulement l'auteur en tant que créateur, dans mon exemple, le champ resterait donc vide.</p>
Date (Date)	<p>Date d'un événement dans le cycle de vie du document. Il peut s'agir de la date de création ou de mise à disposition de la ressource. Il est conseillé d'utiliser le format d'encodage W3CDTF, selon la norme ISO 8601, soit AAAA-MM-JJ.</p> <p>Il existe 8 raffinements pour préciser la nature de la date :</p> <ul style="list-style-type: none"> • la date de création (created) • la date ou période de validité (valid) • la date ou période de mise à disposition (available) • la date de publication (issued) • la date de modification (modified) • la date d'acceptation, pour les thèses par une université par exemple (dateAccepted) • la date du copyright (dateCopyright) • la date à laquelle le document a été soumis, à un comité de lecture par exemple (dateSubmitted) <p><META NAME= « DC.Date.Issued » Scheme= « W3CDTF » CONTENT = « 2008/03/11 »/></p>
Type de la ressource (Type)	<p>Nature ou genre du contenu de la ressource, classé par grandes catégories. Le DCMI Types propose une typologie basique.</p> <p><META NAME= « DC.Type » CONTENT= « Monographie »/> <META NAME= « DC.Type » SCHEME= « DCMI Types » CONTENT=« Text »/></p>
Format (Format)	<p>Format physique ou électronique du document, sa matérialisation. Pour un média, par exemple, on précise sa durée. Il est recommandé d'utiliser une liste de vocabulaire contrôlé. Les types les plus couramment utilisés sont MIME (internet media type) et IMT.</p> <p>Deux raffinements sont possibles :</p> <ul style="list-style-type: none"> • extent : taille ou durée • medium : support physique <p><META NAME="DC.Format.Medium" CONTENT= « 14x21 cm »/> <META NAME="DC.Format.Medium" SCHEME="IMT" CONTENT="Text/html"></p>

Identifiant de la ressource (Identifiant)	<p>Identificateur non ambigu de la ressource. Il est recommandé d'utiliser un système de référencement contrôlé, tel que "Uniform Resource Identifier" (URI), qui inclut le "Uniform Resource Locator" (URL), pour les ressources sur le web, le "Digital Object Identifier"(DOI) et le "International Standard Book Number"(ISBN).</p> <p><META NAME= « DC.Identifiant » CONTENT= «2842631528 »/></p>
Source (Source)	<p>Ressource à partir de laquelle la ressource actuelle a pu être dérivée. La ressource en question peut dériver d'une autre ressource « source », en totalité ou en partie. Il est recommandé de référencer cette source par un vocabulaire contrôlé comme une chaîne de caractère ou un nombre conforme à un système formel identification. Comme pour l'identifiant, la dénomination utilisée peut être celle de la typologie URI.</p> <p>Dans mon exemple, le document choisi est le document primaire, à part le document brut de l'auteur, puisque c'est un roman. Le champ resterait donc vide.</p>
Langue (Language)	<p>Langue du contenu intellectuel de la ressource. Il est recommandé d'utiliser comme valeur de l'élément Langue celles définies par la RFC 1766 qui comprend un code de langage à deux caractères, suivi de façon facultative, d'un code à deux lettres pour le pays.</p> <p><META NAME=DC.Language CONTENT= « fr »/></p>
Relation (Relation)	<p>Cet élément propose le(s) lien(s) vers une ressource liée, qui a un rapport avec le document initial. Pour le transcrire, il est recommandé d'utiliser un système formel d'identification des ressources, par exemple, URI. Il existe 13 raffinements disponibles dans cet élément :</p> <ul style="list-style-type: none"> • isVersionOf : pour une nouvelle version, au-delà de la forme, les changements concernent aussi le contenu. • hasVersion : le document lié est une version modifiée du document initial. (<i>réciproque d'isVersionOf</i>) • isReplacedBy : pour signifier que ce document a été remplacé par le document lié. • Replaces : le document remplace le document lié. (<i>réciproque d'isReplacedBy</i>) 5 / 9 • isRequiredBy : le document permet d'interpréter le document lié. • Requires : le document a besoin du document lié pour être

	<ul style="list-style-type: none"> • <code>isReferencedBy</code> : le document est référencé ou cité par le document lié. • <code>references</code> : le document cite ou pointe vers le document lié. (<i>réciproque d'<code>isReferencedBy</code></i>) • <code>isFormatOf</code> : le document a le même fond que le document lié, mais sous une forme différente. • <code>hasFormat</code> : le document possède une variante sous une forme différente. (<i>réciproque d'<code>isFormatOf</code></i>) • <code>conformsTo</code> : référence à un standard utilisé, auquel se conforme le document.
	<pre><META NAME= « DC.Relation.isPartOf » CONTENT « Collection des ouvrages d'Anna Gavalda, déjà parues aux éditions Le Dilettante »./></pre>
Couverture (Coverage)	<p>Portée du document, tant géographique que temporelle. Elle inclut de même la juridiction, telle qu'une entité administrative. Il est recommandé d'utiliser une typologie des représentations normalisées. La couverture peut être précisée par 2 raffinements :</p> <ul style="list-style-type: none"> • <code>spatial</code> : couverture spatiale. On peut utiliser le codage par point géographique, la norme ISO 3166 qui les pays grâce à 2 lettres, le codage Box qui classe en régions ou encore en fonction du TGN (thesaurus Getty des noms de lieux). • <code>temporal</code> : couverture dans le temps. On peut utiliser les codages <code>Period</code> (intervalle de temps) ou <code>W3CDTF</code> (format de dates, déjà utilisé dans l'élément <code>Date</code>.)
	<pre><META NAME= « DC.Coverage.spatial » CONTENT= "fr"/></pre>
Droits (Rights)	<p>Gestion des droits relatifs à la ressource. Cet élément permet de donner des informations sur le statut du document. Ces droits couvrent les droits de propriété intellectuelle (IPR), Copyright, et divers droits de propriété. L'absence de renseignements concernant cet éléments ne présume pas qu'il est libre de droits.</p>
	<pre><META NAME= « DC.Rights » CONTENT= « © Le Dilettante »</pre>

- Partie 3 -

En quelques lignes présentez, selon vous, 3 avantages essentiels du recours à Dublin Core et 3 inconvénients (problèmes possibles).

Dans un premier temps, on comprend bien que le schéma du Dublin Core est simple, il peut ainsi décrire une ressource de manière rapide et efficace. Il a alors une vocation universelle et doit être utilisé par le plus grand nombre. De cette manière, et notamment sur le web, les métadonnées pourraient permettre aux moteurs de recherche une réponse pertinente des pages web sélectionnées parmi toutes celles disponibles sur le web. Cela éviterait le bruit et pourrait permettre une traçabilité et une transparence des informations disponibles aujourd'hui sur la toile.

En effet, faut-il rappeler que tous les 6 mois, les données présentes sur le web doublent de volume ?! Dans notre branche de métier, nous nous devons d'être efficace et de nous frayer un chemin parmi les méandres de l'information. Le schéma du Dublin Core est une méthode en réponse notamment à cette surabondance.

De même comme le Dublin Core recommande l'utilisation de vocabulaires contrôlés tout au long de ses éléments, il est possible d'harmoniser à une envergure internationale les renseignements. On peut ainsi facilement rêver à la puissance de cet outil de compréhension structurelle des documents.

D'autre part, une des qualités fondamentale du Dublin Core est son interopérabilité entre tous les systèmes d'information, c'est à dire qu'il peut être utilisé dans tous les cas de figure, quel que soit l'entreprise, les logiciels utilisés ou encore le concepteur du matériel (Mac/PC).

Enfin, du fait de sa construction, on voit bien que le Dublin Core est un outil ouvert. Il est certes relativement complet aujourd'hui, mais suivant les évolutions des technologies (web 3.0., ontologies !), il s'adaptera à de nouvelles contraintes, peut-être de nouveaux éléments à faire apparaître, par exemple.

D'autre part, les créateurs du Dublin Core ont autorisé que chaque champ soit facultatif. Ceci vaut mieux que de remplir un élément de manière totalement libre, néanmoins, cette possibilité peut engendrer une non valorisation de l'information contenue. Ainsi, on peut observer que l'utilisation de vocabulaires contrôlés est tant un avantage qu'un inconvénient. En effet, leur mise en pratique n'est pas toujours aisée et peut en devenir elliptique donc non pertinente.

Cette affirmation met en valeur que, certes, le schéma de base peut paraître simple, mais tout aussi complexe s'il est correctement et sur le long terme appliqué, avec tous ses raffinements et pour tout le fonds documentaire. Ainsi, même si, à la base, tout un chacun peut et doit mettre en place et renseigner le Dublin Core, je pense que la place du spécialiste de l'information - documentation conserve un regard privilégié, pointu et vigilant à l'égard de l'utilisation du DC.

Enfin, je trouve que, autant pour la description d'informations provenant du web (page, site...) le schéma du DC est compréhensible, pertinent et concret, autant, paradoxalement, pour les objets physiques, il paraît peu adapté. De même, s'il est simple dans sa construction, il se complexifie du fait de la densité de ses raffinements des possibilités qui s'y rapportent. Mais on peut observer que pour l'utilisation en série, à part dans les raffinements de relation, peu d'informations permettent un travail en lot.

Ainsi, pour la saisie, a-t-on réfléchi à des logiciels (toujours pensés en interopérabilité) capable de cadrer le renseignement d'informations des éléments du Dublin Core, grâce à des thésaurus intégrés et à des listes déroulantes disponibles ? Il serait probablement relativement simple de développer un tel outil et de le rendre « logiciel libre » sur le web pour que leur

utilisation se répande. Je pense qu'aujourd'hui, il existe un fossé entre les capacités et les fonctionnalités avancées (mais aussi le fait de sa simplicité d'utilisation basique) de cet outil et le manque d'informations diffusées au grand public. Pourquoi ne pas organiser des ateliers de formation, afin de sensibiliser les gens à ces nouvelles pratiques, en leur faisant notamment entrevoir la puissance de ce schéma s'il est universellement ancré dans les pratiques (voir avantages !).

- Partie 4 -**Références bibliographiques**

Dico du Net. DC-Dublin Core. In Dico du Net, [En ligne] <<http://www.dicodunet.com/definitions/referencement/dublin-core.htm>>, (consulté le 22.03.2008)

HILLMANN, Diane. Guide d'utilisation du Dublin Core. In Bibliothèque de l'Université de Laval. [Disponible en ligne] <<http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>>, (page consultée le 08.04.2008).

DCMI. In Dublin Core Metadata Initiative [En ligne] <<http://dublincore.org/>>, (page consulté le 22.03.2008).

JACQUET, Christophe. Métadonnées et Dublin Core. In Open Web. [Disponible en ligne] <http://www.openweb.eu.org/articles/dublin_core/>, (page consultée le 08.04.2008)

WIKIPEDIA. Dublin Core. In : fr.wikipedia.org. [Disponible en ligne] <http://fr.wikipedia.org/wiki/Dublin_Core>, (Consulté le 22.03.2008).

Groupe AFNOR, Glossaire. In TEF 2.0 [Disponible en ligne], <<http://www.abes.fr/abes/documents/tef/recommandation/glossaire.html>>, (Page consultée le 22.03.2008)