

# Cours sur le Dublin Core et l'OAI

Christian Ducharme – 25 mai 2006

## Qu'est-ce que le Dublin Core

Dublin Core est un ensemble de métadonnées permettant de décrire tout type de document numérique. Le format a été créé par des bibliothécaires avec l'intention de « cataloguer » le web : il est volontairement simple voire limité, mais il est aujourd'hui le moyen le plus répandu pour décrire un fonds documentaire numérique sur le web.

## Qu'est-ce que l'OAI

L'OAI (Open Archives Initiative) est un protocole permettant de décrire et d'interroger des fonds documentaires.

« Le protocole OAI définit un standard pour transférer des collections de métadonnées et permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine ».

*<http://blogokat.canalblog.com/archives/2006/03/05/1452081.html>*

## Dublin Core : structure du format

Le format Dublin Core possède 15 champs, répétables et non obligatoires, répartis en trois catégories :

Contenu : Couverture, Description, Type, Relation, Source, Sujet, Titre

Propriétés intellectuelles : Créateur, Collaborateur, Editeur, Droits

Instanciation : Date, Format, Identifiant, Langue

<http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm#1.2>

Les métadonnées Dublin Core s'insère habituellement dans l'en-tête des documents html :

```
<html>
<head>
<meta name="DC.Identifier" content="http://www.normannia.info/document/barenton1825.html">
<meta name="DC.Creator" content="Barenton, Auguste">
<meta name="DC.Title" content="Récit d'un rêve">
<meta name="DC.Publisher" content="Avranches : Imprimerie Tribouillard">
<meta name="DC.Date" content="1825">
<meta name="DC.Description" content="58 pages">
<meta name="DC.Subject" content="littérature">
</head>
```

Il y a deux façons de « voir » le Dublin Core : le document numérique est une instance du document papier (ce qui est le cas de la description ci-dessus) ou la version du document numérique est indépendante de celle du papier et considérée comme une édition à part entière, ce qui modifierait le contenu des champs :

<meta name="DC.Publisher" content="CRL de la Basse-Normandie">  
<meta name="DC.Date" content="2006">  
<meta name="DC.Source" content="Avranches : Imprimerie Tribouillard, 1825">

## **L'architecture de l'OAI**

L'OAI a un fonctionnement basé sur la collecte de métadonnées. (Le principe est celui des archives ouvertes.) Deux acteurs sont impliqués : l'organisme qui collecte les métadonnées (moissonneur) et l'organisme qui rend disponible les métadonnées des documents qu'il possède (entrepôt). Le premier récupère les métadonnées des entrepôts afin de créer un service de recherche (à l'image d'un moteur de recherche). Le deuxième crée un serveur permettant aux moissonneurs de récupérer ses métadonnées et aux utilisateurs finaux de consulter ses documents.

L'OAI s'appuie sur deux grandes normes : http, on communique en mode web, xml, on échange des informations structurées. On utilise le format Dublin Core (pour les métadonnées) et les fichiers sont encodés en Unicode. Il y a trois niveaux d'information : les métadonnées, les ensembles de données (sets) et les documents numériques.

Le protocole prévoit 6 instructions (types de requête), appelées verbes qui peuvent être accompagnées de paramètres (facultatifs ou obligatoires).

### Identify

Demande d'information sur l'entrepôt (aucun paramètre)

### ListMetadataFormats

Demande la liste des formats disponibles (Dublin Core, EAD, etc.).

Paramètre facultatif : identifier

### ListSets

Demande la liste des ensembles de données disponibles (collections)

Paramètre facultatif : resumptionToken

### ListIdentifiers

Demande la liste des identifiants (identifiant + collection)

Paramètre obligatoire : metadataPrefix

Paramètres facultatifs : from, until, set, resumptionToken

### ListRecords

Demande la liste des enregistrements (métadonnées complètes)

Paramètre obligatoire : metadataPrefix

Paramètres facultatifs : from, until, set, resumptionToken

GetRecord : demande un enregistrement en particulier

Paramètres obligatoires : metadataPrefix, identifier

Exemple :

<http://oai.bnf.fr/repositoryOAI.php?verb=ListSets>

[http://oai.bnf.fr/repositoryOAI.php?verb=ListIdentifiers&metadataPrefix=oai\\_dc&set=gallica](http://oai.bnf.fr/repositoryOAI.php?verb=ListIdentifiers&metadataPrefix=oai_dc&set=gallica)

# Exercice pratique

## Dublin Core

Créer les métadonnées des documents suivants :

<http://www.bm-chambery.fr/scripts/opsys.asp?MODULE=3W9501&NOTICE=0017445&ETAPE=E1&NUMORDRE=1>

[http://www.normannia.info/cgi-bin/aurweb.exe/normannia/voir?tit=colza&aur\\_offset\\_rec=2&expr=+colza](http://www.normannia.info/cgi-bin/aurweb.exe/normannia/voir?tit=colza&aur_offset_rec=2&expr=+colza)

<http://opac.ge.ch/gateway?host=opac.ge.ch%2b8861%2bDEFAULT&search=KEYWORD&function=CARDSCR&SourceScreen=CARDSCR&sessionid=20060525105530334312&skin=rero&conf=.%2fchameleon.conf&lng=fr-ch&itemu1=1035&u1=1035&t1=biologie&pos=3&prevpos=2>

## Protocole OAI

L'url de l'entrepôt OAI de la BnF est : <http://oai.bnf.fr/repositoryOAI.php>

Que se passe-t-il lorsque vous l'interroger sans verbe, ni paramètre ?

Demandez des informations générales sur l'entrepôt de la BnF : Identify.

Demandez la liste des formats disponibles à la BnF.

Demandez la liste des ensembles de données (collections) : ListSets

Que se passe-t-il si vous écrivez : listsets au lieu de ListSets ?

Demandez la liste des identifiants d'une collection (un set) : ListIdentifiers  
(n'oubliez pas, le paramètre metadataPrefix=oai\_dc est obligatoire)

A l'aide d'un identifiant que vous avez obtenu, demandez l'enregistrement complet : GetRecord (n'oubliez pas le paramètre metadataPrefix et identifier sont obligatoires).

Demandez la liste des enregistrements d'une collection : ListRecords  
(n'oubliez pas, le paramètre metadataPrefix est obligatoire)

Utilisez le token pour obtenir la suite des enregistrements : resumptionToken=999999999 où 999999999 correspond au token transmis par le serveur OAI.

Essayez de limiter votre requête à une plage de date ou d'année.

## Lectures complémentaires

Oulahal, Saïd. Accès unique à des ressources numériques distribuées [en ligne].  
<http://2003.jres.org/diapo/paper.125.pdf>

Séigny, Martin. Atelier ADN X / OAI [en ligne]. <http://adnx.org/documentation/oai/>

Le site de l'Open Archives : <http://openarchives.org>