

L'INFORMATION INDEXÉE

MÉTADONNÉES ET DUBLIN CORE

FRANCOIS.SCHMITT@epfl.ch, EPFL BIBLIOTHÈQUE CENTRALE



La facilité avec laquelle les nouvelles technologies de l'information permettent aujourd'hui à tout un chacun de publier ses travaux sur Internet ne doit pas cependant laisser croire qu'il suffit de mettre à disposition sur un serveur tous nos documents et de leur attribuer des URL pour que le tour soit joué.

Un projet de publication électronique devrait en effet être mené avec un minimum de rigueur afin que les informations diffusées puissent être non seulement disponibles sur le long terme avec des URL stables, mais également exploitables grâce à une structuration concertée des données. Ceci pour garantir le repérage, l'analyse et la récupération des documents par les moteurs d'indexation et, bien sûr, par les internautes.

Pour ce faire, il est conseillé de joindre à tout document ou fichier électronique une carte d'identité dans laquelle seront recensés les paramètres essentiels à son identification et au traçage de l'information qui y est contenue. Ces informations sur les ressources documentaires sont appelées des métadonnées (du grec *meta*: second, au-delà, de nature plus élevée et fondamentale) et forment un ensemble structuré d'informations décrivant une ressource qui soit l'accompagne, soit y est directement intégré (encapsulage). Ces informations sur l'information se retrouvent, par exemple, dans les notices bibliographiques des catalogues de bibliothèques qui nous renseignent sur la nature de l'information contenue dans le document catalogué et se traduisent par des données univoques telles que l'auteur, le titre, le sujet, etc., facilement identifiables par l'utilisateur.

Les Bibliothèques fortes de leur expérience dans le traitement de l'information ont désiré apporter à la communauté Internet leur savoir-faire dans ce domaine en proposant une indexation plus fine que celle déjà générée par les balises des documents html, mais il n'était pas question de proposer à des non-spécialistes une structuration aussi développée et complexe que, par exemple, le format MARC21 qui est utilisé pour le catalogage professionnel. Cela aurait été un frein au but visé: un emploi simple, souple et le plus largement répandu possible.

Ainsi, un standard de métadonnées consensuel a été développé depuis 1995 en collaboration entre l'OCLC (Online Computer Library Center) et le NCSA (National Center for Supercomputing Applications) à Dublin (Ohio). Ce standard est appelé le Dublin Core, *core* utilisé pour signifier ici l'aspect central, le noyau des métadonnées minimales sans lesquelles l'information ne peut pas être considérée comme exploitable.

Le Dublin Core se veut international et extensible. Il cherche à couvrir les besoins des disciplines spécialisées dans la gestion de l'information. Depuis 2003, il est devenu une norme ISO et présente un intérêt certain de par sa simplicité et son applicabilité dans tous les formats de fichiers (utilisable

avec HTML et XML). Le standard DC est en outre le format minimum de l'Open Archives Initiatives (OAI) qui permet, via l'interopérabilité des serveurs, l'exposition et la récolte de métadonnées.

STRUCTURE DU DUBLIN CORE:

Le DC est formé de 15 éléments optionnels, répétables et pouvant apparaître sans ordre. Ces éléments peuvent être répartis en 3 groupes:

1. CONTENU

- Titre / Title
- Sujet / Subject [i.e. mots-clés]
- Description [i.e. résumé]
- Source [i.e. référence originelle]
- Langue / Language
- Relation [i.e. autre(s) source(s) en rapport]
- Couverture / Coverage [i.e. spécification spatio-temporelle]

2. PROPRIÉTÉ INTELLECTUELLE

- Créateur / Creator [i.e. auteur]
- Editeur / Publisher
- Contributeur / Contributor
- Droits / Rights

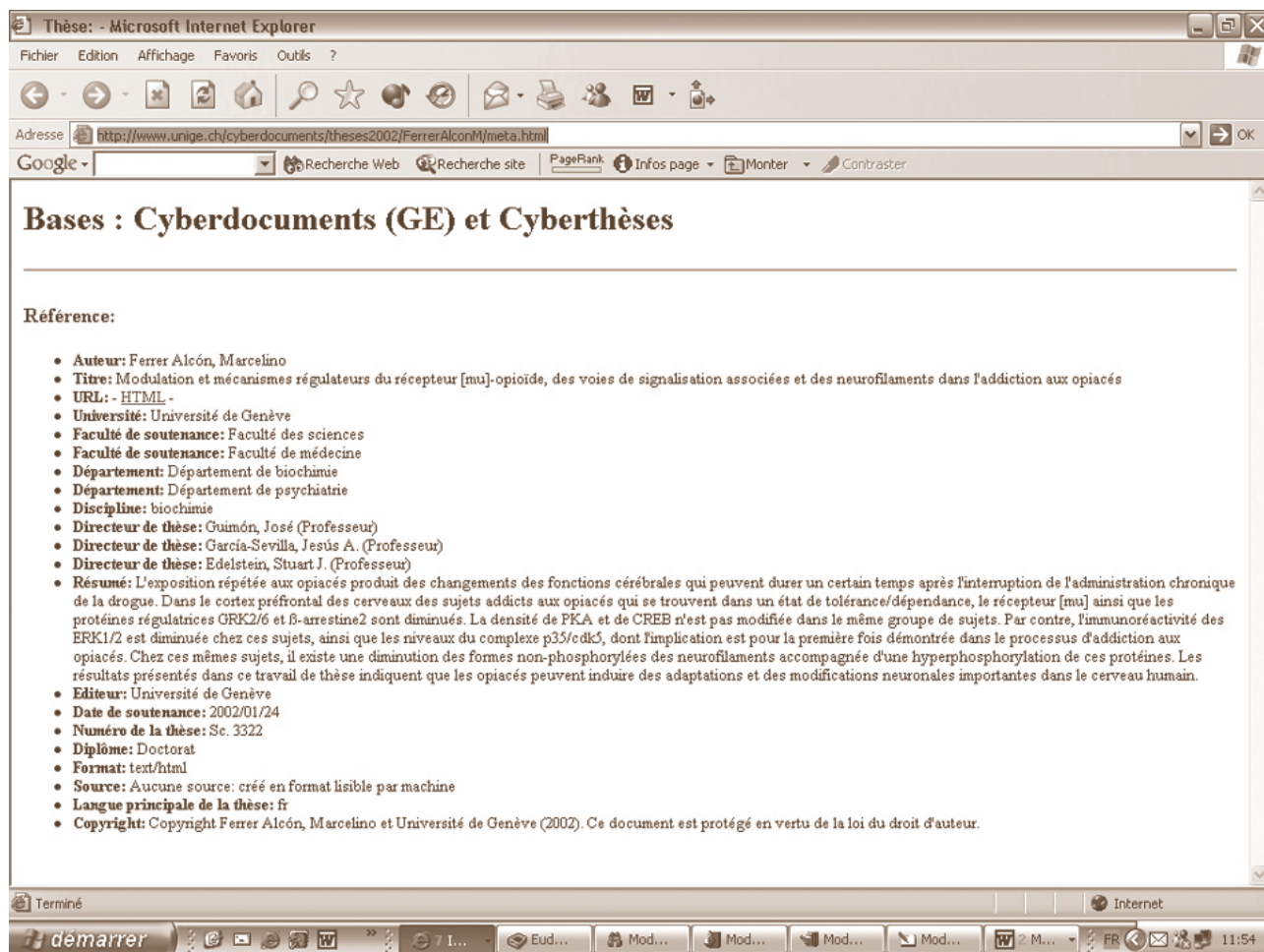
3. ADMINISTRATIF (INSTANCIATION)

- Date
- Type
- Format
- Identifiant / Identifier

APPLICATION PRATIQUE

L'élaboration d'une granularité de l'information est particulièrement efficace lorsque l'on doit traiter une collection de documents dont la forme reste plus ou moins constante comme, par exemple, une collection de thèses qui, quel que soit leur contenu, présentent des caractéristiques formelles identiques répétables. Préparer une grille de saisie avec des éléments choisis selon la profondeur de l'information souhaitée offre une gestion simple et cohérente pour l'ensemble des objets qui viendront enrichir la collection.

Voici une illustration des métadonnées retenues dans le cadre du projet des thèses électroniques de l'Université de Genève qui a choisi d'ajouter des éléments supplémentaires hors Dublin Core (directeur de thèse, etc.), mais qui présentent un ensemble homogène et structuré permettant une identification aisée et une qualité d'information optimum pour ces documents académiques.



(<http://www.unige.ch/cyberdocuments/theses2002/FerrerAlconM/META.HTML>)

ET À L'EPFL ?

La Bibliothèque centrale a inauguré en mars 2003 la mise à disposition des thèses sous forme électronique via son catalogue des thèses (<http://library.epfl.ch/theses/>) dont les notices reprendront d'ici la fin de l'année les données du Dublin Core en conformité avec les prescriptions du projet international Cyberthèses. Les docteurs et les doctresses pourront directement saisir ou compléter ces données (attribuer par exemple des mots-clés) grâce à un formulaire en ligne qui permettra aussi l'échange d'information avec d'autres serveurs, notamment dans le cadre du projet E-Helvetica de la Bibliothèque nationale suisse (<http://www.e-helvetica.admin.ch/>).

POUR EN SAVOIR PLUS:

SITES

Dublin Core <http://www.dublincore.org/>
 OCLC <http://www.oclc.org/home/>
 NCSA <http://www.ncsa.uiuc.edu/>
 OAI <http://www.openarchives.org/documents/FAQ.html>
 Norme ISO 15836:2003
<http://www.iso.org/international/SC4/n515.pdf>

Guide d'utilisation du Dublin CORE

en anglais: <http://dublincore.org/documents/usageguide/>
 en français: <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>

ÉLÉMENTS DU Dublin CORE

en anglais <http://dublincore.org/documents/dces/>
 en français <http://www-rocq.inria.fr/~vercoust/METADATA/DC-fr.1.1.html>
 Dublin Core en HTML <http://www.ietf.org/rfc/rfc2731.txt>