

La Recherche d' Information sur INTERNET et les METADATA: Etat de la question

ALI OUALI Nadia

Chargée d'Etudes

Laboratoire de Recherche et Développement

en Information Scientifique et Technique

CERIST

e-mail : Aliouali@tassili.cerist.dz

Introduction :

Le but de cette synthèse est de retracer les caractéristiques des métadonnées et en particulier du Dublin Core afin de faire paraître les problèmes typiquement documentaires qui sont soulevés par leur utilisation. L'origine de ces métadonnées remonte jusqu'à 1995. Elle se focalise sur le développement et la promotion des éléments de METADATA qui facilitent le repérage des ressources textuelles et non textuelles dans un environnement réseaux et facilitent l'hétérogénéité l'interopérabilité des données. L'objectif est de développer une simple syntaxe qui va être utilisée par les producteurs et fournisseurs de l'information pour décrire leurs ressources électroniques et pour faciliter enfin le repérage de ces ressources.

L'expansion rapide d'INTERNET et l'introduction du World Wide Web (WWW) n'a pas été suivie par un changement fondamental de la manière de faire une recherche sur ce grand réseau. Le travail effectué par les différents outils d'indexation automatique et les moteurs de recherche est vaste et ambigu à la fois. Un chercheur qui cherche une simple information se trouve face à une vague de réponse qui parfois n'a aucune relation avec la requête formulée.

La recherche d'information sur INTERNET est dominée par deux types d'approches:

- Méthode "robots" (robot de type ALTA VISTA , ou LYCOS)
- Méthode "répertoire validé (robot de type YAHOO)

| Méthode/Référentiel | l'Internet | un sous-ensemble d'Internet |
|---------------------------------------|---|--|
| méthode "robots" | Alta-Vista, Excite, Inkotomi, Lycos, Infoseek, Open Text... | les robots francophones : (Ecila, Echo, Lokace, Francité...) les robots disciplinaires : (Hippocratis pour la médecine, Argos pour l'histoire médiévale...) |
| méthode "répertoires validés " | Yahoo, Virtual Library, Excite, Lycos A2Z... | les répertoires francophones (Urec, Nomade, Yahoo-France...) les répertoires de revues électroniques, de forums... |

Dans le premier cas un Robot comme ALTA - VISTA comportent un module de collecte automatique des ressources (ROBOTS COLLECTEURS), un module d'indexation et le module recherche d'information.

Au moment d'une recherche ces moteurs envoient systématiquement des programmes pour collecter sur tous les sites qu'ils identifient utile, pour collecter les pages HTML ensuite ils passent à l'indexation de ces pages pour constituer le fichier inversé, l'indexation se fait soit sur le titre, les premiers paragraphes et les liens sans oublier que cette indexation tient compte sur le critère de fréquence, vers la fin le modules de recherche d'information est lancé.

Dans le deuxième cas se sont des répertoires thématiques de ressources ordonnées selon un classification (CDD, LC, Mesh .Il s'agit de listes (ou annuaires, répertoires, catalogues) généralistes, classées par grandes catégories organisées hiérarchiquement. Elles sont construites à la main par des personnes filtrant les sites en fonction de leur qualité, pertinence et fiabilité. Le travail d'évaluation est très variable d'une liste à l'autre et le temps passé sur chaque site est de quelques minutes en général.). Pour ces répertoire et contrairement à l'approche robot, l'indexation se fait sur des zone limitée (URL, titre, résumé)

Donc pour le premier cas , les réponses sont hétérogènes, les doubles ne sont pas détectés et le bruit est considérable, pour ce qui est du deuxième cas , le seul intérêt et la valeur humaine ajoutée (contrôle , organisation) qui limite ainsi le bruit .Les classification utilisées permettent de répondre a des requêtes dans des thématiques bien définis par contre l'exhaustivité est limitée par rapport aux robot de recherche.

Aujourd'hui les moteurs de recherche traitent quotidiennement des millions de requêtes. Cependant ces derniers sont loin des outils efficaces pour s'y retrouver dans la masse considérable d'information qui ne cesse de s'accroître sur INTERNET. Cependant ce réseau mondial ne dispose pas encore de normes qui facilitent l'indexation automatique.

Les auteurs de pages Web actuellement se préoccupent beaucoup plus de la forme que de la structuration du contenu d'une page. Aussi il ne faut pas s'étonner si les logiciels d'indexation s'arrêtent à un niveau très pauvre d'indexation du contenu. Même des informations factuelles simples comme l'auteur, la date de publication, la taille du document ou le thème manquent, sans oublier qu'actuellement les adresses changent sans que les index soient mis à jour en conséquence.

Pour pallier ces types de problèmes classiques en informatique documentaire, plusieurs solutions ont été envisagées . Parmi ces solutions l'insetion des balises <META> (ou méta-tags). Elles contiennent la description et les mots clés que certains moteurs de recherche utilisent pour classer les pages dans leurs bases de données.

Les méta-tags que nous inséreront en tête du document html sont des éléments essentiels de la promotion d'un site. La plupart des moteurs de recherche qui indexent les pages du web pénétreront dans le document html et en extrairont des informations. Si vous avez utilisé des méta-tags, les informations qu'ils contiennent seront enregistrées par les moteurs et ensuite affichées aux utilisateurs lors d'une recherche.

Si le moteur ne trouve pas de méta-tags, il se contentera de prendre les premières lignes du document HTML et les affichera comme description, ce qui n'est peut-être pas souhaitable.

Alta Vista détecte les marques <meta name="description" content="Le texte que vous voulez voir apparaître lorsque le résultat d'une recherche pointe sur votre document"> et les utilisent dans les réponses à des recherches. <meta name="keyword" content="une suite de mots-clés séparés par des virgules">

Mais ces simples méta -Tags ne sont pas suffisant a eux seul. C'est pourquoi certains proposent de joindre des métadonnées (metadata) aux fichiers pour enrichir l'indexation automatique.

Les métadonnées où encore les metadata sont des données structurées, standardisées permettant d'identifier une ressource sur le réseaux INTERNET facilitant ainsi le travail de recherche des robots e, réduisant le têt de bruit à l'interrogation.

Plusieurs communautés ont défini des standards de métadonnées correspondant à leur besoin ; on peut citer :

- **Machine Readable Card** (MARC) pour les éléments de catalogues de bibliothèques,
- **Dublin Core** (DC) définit des métadonnées associées aux pages Web,
- **Consortium for the Interchange of Museum Information** (CIMI) définit les métadonnées associées aux informations muséographiques.

Les métadonnées améliorent la précision des recherches de documents par exemple en donnant des localisations géographiques, la langue ou des périodes historiques. Ce type de données est très difficile à extraire de documents en langage libre. Elles permettent aussi de comparer des documents issus de différentes sources.

A l'image d'Internet, il existe de nombreux formats de metadata en fonction des ressources pour lesquelles ils ont été développés. Mais pour assurer une utilisation possible par les outils de repérage comme les robots, une certaine standardisation minimum semble être adoptée le " Dublin Core Metadata standard " Premier Workshop " sur les metadata qui s'est tenu à Dublin (Ohio) en Mars 1995.

Les formats bibliothéconomiques traditionnels tels que les formats MARC sont en fait des metadata particulières, un projet mené par OCLC (InterCat) vise d'ailleurs à étendre le format MARC aux ressources électroniques en rajoutant certains champs. Dans une logique un peu similaire, les organismes de normalisation internationaux travaillent à l'extension des formats traditionnels bibliographiques pour intégrer les ressources électroniques, c'est le cas de la norme ISBD-ER (International Standard Bibliographic Description for Electronic Ressources, 1997).

Mais est-il vraiment réaliste de vouloir suivre le même modèle de repérage pour des sources papier et des sources Internet ? Le document Internet est accessible immédiatement par des clés d'accès diverses, il peut évoluer souvent et c'est un document qui doit circuler rapidement généralement.

L'approche catalogage exhaustif de type MARC est intéressante pour des documents électroniques importants (un ouvrage, une revue, une base de données...) et durables dans le temps, on ne peut envisager ce type de repérage pour tous les types de ressources circulant sur l'Internet.

Le " Dublin Core Metadata " défend une approche plus simple du catalogage avec un nombre de champs limités qui visent à être fournis directement par l'auteur ou l'éditeur des ressources.

1 /A l'origine du DUBLIN CORE

Dés 1995, l'OCLC (Online Computer Library Center) reconnue la nécessité d'utiliser des données descriptives pour l'indexation des ressources sur INTERNET.

Soucieuse du manque de données sur les pages Web, OCLC hébergea une session afin d'amener ensemble les ingénieurs, documentalistes, académicien ainsi que d'autre spécialiste pour discuter d'un standards pour les metadata. Cette session aboutie à l'identification de 13 éléments à inclure dans la tête d'un fichier HTML. Ce nombre a été étendu plus tard jusqu'à 15 éléments du DUBLIN CORE

Cela dit, les premières expériences ont montré une très grande dispersion dans l'usage de ces metadata; le Dublin Core identifie une initiative visant à améliorer la normalisation des Metadata en proposant un mode de catalogage sur INTERNET en respectant deux objectifs : d'une part, être plus accessible aux usagers que les traditionnels formats MARC. D'autre part faciliter l'interopérabilité des applications.

Le dublin core semble être le standard sur lequel la communauté internationale voudrait se mettre d'accords. Son objectif n'est pas de remplacer des outils de catalogage comme le MARC ou les outils d'indexation automatique mais d'en être le complément dans le but d'améliorer la précision des recherches effectuées sur le Web.

Les Ateliers du Dublin Core en fourni un catalyseur et un sens pour le développement de la littérature. Utilisant une approche et une méthodologie pour développer un consensus. Chaque atelier a contribué au raffinement des discussions et la redirection de la recherche sur INTERNET telle que décrite auparavant.

2 / Historique des ateliers du DUBLIN CORE

(DC-1): Le premier Atelier du DUBLIN CORE s'est tenu en Mars 1995. Sponsorisé par l'OCLC (Online Computer Library Center) et le NCSA (National Center for Supercomputing Application). L'atelier initial s'est concentré sur la sémantique de description dirigée vers le problème de la découverte des documents électronique. Ce premier atelier a donné un "CORE de 13 éléments permettant la description de ces ressources.

On désigne sous le nom de metadata des données secondaires décrivant des données primaires.

Le format proposé a été intitulé " Dublin core metadata élément set " ; il s'applique à des DLOs (Document like objects), des objets qui ressemblent à des documents. A l'instar d'un format catalographique, plusieurs champs ont été prévus pour décrire les " objets qui ressemblent à des documents ". Les différents champs du Dublin Core ont émergés à partir d'un effort international pour la construction d'un consensus manifesté à travers une série d'Ateliers

2.1 / Etendu :

Le Dublin Core (le nom retenu pour la description de documents électroniques mis sur Internet) n'ont pas la prétention de supplanter les autres descriptions de ressources mais plutôt de les compléter.

Deux façons existent pour décrire un document électronique : un index généré automatiquement utilisé par les services des engins de recherche et le fichier catalographique avec le Marc créé par des professionnels du domaine de la documentation.

Ces deux solutions sont très extrêmes : l'une contient généralement très peu d'information utilisable à cause du croisement sémantique et syntaxique des sciences d'aujourd'hui ; l'autre coûteuse (temps et argent) vu le nombre impressionnant de documents existants sur le réseau.

Pour cela, la solution proposée au cours de l'atelier est une médiation entre celles-ci.

En effet, cette description simple est mise à la disposition des auteurs et autres éditeurs de l'information pour décrire leur document. Ainsi chaque document mis sur réseau aura une description qui permettra de le localiser et de le valoriser.

D'autant plus que des principes fondamentaux ont été énoncés pour permettre le développement des Dublin Core et leur conversion de et vers les formats bibliographiques Standards (USMARC, UNIMARC, etc)

2.2/ Principes Fondamentaux :

Les différentes discussions menées durant cet atelier, ont permis de révéler plusieurs principes qui devaient guider le développement de l'ensemble des éléments du DUBLIN CORE SET.

Ainsi, les DUBLIN CORE SET sont :

- **Intrinsèques**

Les Dublin Core se concentrent sur les propriétés intrinsèques de l'objet.

Les données intrinsèques se réfèrent aux propriétés du document quand on l'a en main comme le contenu intellectuel, la forme physique etc. on distingue d'autres parts des propriétés qui sont données extrinsèques et qui décrivent le contexte dans lequel le document est utilisé comme le coût, accès à l'information etc.

- **Extensibles**

Les mécanismes d'extension des Dublin Core éléments devaient pallier à un certain nombre de problèmes en incluant par exemple des données extrinsèques pour une meilleure description de l'objet.

L'extension est très importante de par les nombreuses spécificités de chaque domaine scientifique. Ceci dit, si les Dublin Core éléments doivent changer au cours du temps; ils devront garder toujours un minimum d'éléments compatibles avec la première version.

- **Indépendance syntaxique**

Les obligations syntaxiques sont délaissées parce qu'il est très tôt de proposer des définitions formelles très rigoureuses à cause de la multitude de disciplines scientifiques que devrait couvrir les Dublin Core éléments.

- **Optionnelles**

Tous les éléments sont optionnels.

Les Dublin Core peuvent être éventuellement utilisés pour des objets mais certains éléments peuvent s'avérer sans sens (ex: l'auteur d'une image satellitaire ?).

- **Répétitives**

Tous les éléments dans les Dublin Core sont répétitifs.

Par exemple, un document peut avoir une multitude d'auteurs.

- **Modifiables**

Chaque élément dans le Dublin Core Set a une définition qui est suffisante pour une explication simple. Cependant, il est nécessaire que les définitions de chaque élément satisfassent les besoins des différentes communautés. Ce but est atteint par l'allocation à chaque élément un qualificatif optionnel. Si aucun qualificatif n'est présent, c'est que l'élément a un sens commun. Et s'il y est présent la définition de cet élément est modifiée par la valeur qu'on peut donner au qualificatif.

2.3 Fonction et objectif des METADATA

Plusieurs raisons nous poussent à utiliser les METADATA

- Ils sont faciles à créer,
- Simple à indexer
- Permettent, pour les moteurs de recherche, une meilleure indexation qu'un texte intégral,
- Permettent l'interopérabilité

Les Dublin Core soutiennent et fournissent la communauté d'information en général avec le moyen de

- Décrire une ressource (repérage, structure)
- Construire une bonne et simple description et permettre une meilleure indexation
- Découvrir les ressources
- Accessibilité aux ressources
- Améliorer la qualité de la gestion des ressources

2.4/ Description des DUBLIN CORE SET:

Les éléments obligatoires

Titre

étiquettes : *DC Title*

Définition : Titre de la ressource donné par le *créateur* ou l'*auteur*.

Elément Obligatoire

| Format | Type | Description |
|--------|-------------------|--|
| Aucun | Title | Titre donné au document |
| | Title.Alternative | Titre supplémentaire ou sous-titre donné au document |

Exemple de métadonnée

```
meta name="DC.Title" CONTENT="Les METADATA">  
<meta name="DC.Title.Alternative" CONTENT="Synthèse des travaux des  
ateliers du Dublin Core">
```

Auteur ou Créateur

étiquettes : *DC.Creator*

Définition : La personne ou l'organisation principalement responsable de la création du contenu intellectuel de la ressource. Par exemple les auteurs d'un document, ou les artistes, photographes, illustrateurs dans le cas de ressources visuelles.

Elément Obligatoire

| Format | Type | Description |
|--------|---------|--|
| Aucun | Creator | Nom ou organisation responsable de la création du document |

Les auteurs sont écrits dans le format suivant : "Nom, prénom". Dans le cas où il y aurait plusieurs auteurs, on sépare chaque entrée par un point-virgule (;). Dans le cas où il s'agit d'une organisation, on inscrit le nom complet suivi de l'acronyme, s'il y a lieu, entre parenthèses.

Exemple de métadonnée

```
<meta name="DC.Creator" CONTENT="ALIOUALI NADIA" >  
<meta name="DC.Creator" CONTENT="Centre de Recherche sur l'Information  
Scientifique et Technique (CERIST)" >
```

Sujet et mot-cle

étiquettes : *DC.Subject*

Définition : Thématique de la ressource. Typiquement, le sujet sera décrit par un ensemble de mots-clefs ou de phrases qui précisent le sujet ou le contenu de la ressource. L'utilisation de vocabulaires contrôlés et de schémas de classification formels est encouragée.

Elément Obligatoire

| Format | Type | Description |
|--------|---------|--|
| Aucun | Subject | Liste de mots-clef ou d'expressions permettant de décrire le contenu du document |

Dans le cas où il y aurait plusieurs mots-clés, on sépare chaque entrée par un point-virgule (;).

Exemple de métadonnée

```
<meta name="DC.Subject" CONTENT="METADATA ; Ressources  
électronique; Catalogage">
```

Editeur

étiquettes : *DC.Publisher*

Définition : L'entité responsable de l'édition et la diffusion de la ressource dans sa forme actuelle, telle qu'une maison d'édition, un département universitaire, une entreprise.

Élément Obligatoire

| Format | Type | Description |
|---------------|-------------|--|
| Aucun | Publisher | Éditeur du document dans sa forme actuelle |

Les éditeurs sont écrits dans le format suivant : "Nom, prénom". Dans le cas où il y aurait plusieurs éditeurs, on sépare chaque entrée par un point-virgule (;). Dans le cas où il s'agit d'une organisation, on inscrit le nom complet suivi de l'acronyme, s'il y a lieu, entre parenthèses.

Exemple de métadonnée

```
<meta name="DC.Publisher" CONTENT="Centre de Recherche sur L'Information Scientifique et Technique (CERIST)">
```

DATE

étiquettes : *DC.Date.Creation_of_intellectual_content*
DC.Date.Date.Creation/Modification_of_present_form
DC.Date.Acquisition/Accession

Définition : La date à laquelle la ressource a été publiée dans sa forme actuelle. L'usage recommandé est sous la forme d'un nombre de 8 chiffres tel que YYYY-MM-DD, comme défini dans <http://www.w3.org/TR/NOTE-datetime>, faisant partie de la norme ISO 8601. Dans ce schéma, l'élément date 1994-11-05 correspond au 5 Novembre 1994. Beaucoup d'autres schémas sont possibles, mais si un autre schéma est utilisé, il devrait être précisé de façon non ambiguë.

Elément Obligatoire

| Format | Type | Description |
|------------------|---------------------------------------|---|
| ISO 8601:1988 | Creation_of_intellectual_content | Date à laquelle le document a été créé |
| | Acquisition/Accession | Date à laquelle le document est acquis par le CRL |
| | Creation/Modification_of_present_form | Date à laquelle le document est converti en HTML ou modifié |

Dans le cas où la date exacte n'est pas connue (jours ou mois), la date indiquée devrait être le premier jour du mois, pour les documents où le jour n'est pas indiqué, ou le premier jour de l'année, pour les documents où le mois n'est pas indiqué.

Exemple de métadonnée

```
<meta name="DC.Date.Creation_of_intellectual_content" SCHEME="ISO 8601"
CONTENT="1998-12-17" >
<meta name="DC.Date.Acquisition/Accession" SCHEME="ISO 8601"
CONTENT="1998-12-17">
<meta name="DC.Date.Creation/Modification_of_present_form" SCHEME="ISO
8601" CONTENT="1998-12-20">
```

Type de ressource :

étiquettes : *DC.Type*

Définition : La catégorie de la ressource, telle qu'une page personnelle, un roman, un poème, un document de travail, un rapport technique, une dissertation, un dictionnaire. Dans un souci d'interopérabilité, un *type* devrait être choisi à partir d'une liste établie qui est en cours de définition dans la série des groupes de travail, au moment de la publication de ce document.

Elément Obligatoire

| Format | Type | Description |
|---------------|-------------|---|
| Aucun | Type | Type du document, d'abord de façon générale, puis de façon plus précise |

On inscrit d'abord le type général avec la première lettre en majuscule (Texte, Image, Animation, etc) suivi d'un point et du genre de document avec toujours la première lettre en majuscule (Article, Mémoire, Statistiques, etc.)

Exemple de métadonnée

```
<meta name="DC.Type" CONTENT="Texte.Article">  
<meta name="DC.Type" CONTENT="Image.Gravure" >
```

Format

Etiquette : *DC Format*

Définition : Le format de la ressource, utilisé pour identifier le logiciel et, éventuellement, le matériel qui peut être nécessaire pour afficher ou traiter la ressource. Dans un souci d'interopérabilité, un format devrait être sélectionné à partir d'une liste établie qui est en cours de définition dans la série des groupes de travail.

Identifiant de la ressource

Etiquette: DCidentifier

Définition : Chaîne de caractère ou nombre utilisé pour identifier de façon unique la ressource. Exemples pour des ressources réseau incluent URLs et URNs (si implémenté). D'autres identificateurs globaux et uniques, tels que ISBN (International Standard Book Numbers), ou d'autres noms formellement définis, sont des candidats potentiels pour cet élément, dans le cas de ressources privées.

Exemple de métadonnée

```
<meta name="DC.identifier" CONTENT="(SCHEME=URL)
http://www.cerist.dz/dst/metadata.html">
```

Langage

Etiquette:DC. language

Définition :Langage(s) du contenu intellectuel de la ressource. Si approprié, le contenu de ce champ devait correspondre avec la norme RFC 1766. Voir:

<http://ds.internic/rfc/rfc1766.tx>

Exemple de métadonnée

```
<META NAME="DC.Language" CONTENT="fre" >
```

Les éléments facultatifs

Autre contributeur

Etiquette: DC.Contributor

Définition : Une personne ou une organisation, non mentionnée dans un élément créateur, qui a contribué à la ressource mais dont la contribution est secondaire comparée à celle de toute personne ou organisation spécifiée dans un élément créateur (par exemple un rédacteur, un traducteur, un illustrateur).

Source

Etiquette: DC.Source

Définition : Une chaîne de caractère ou un nombre, utilisé pour identifier de façon unique le travail d'où la ressource est dérivée, si applicable. Par exemple une version PDF d'un ouvrage peut avoir un élément source contenant un numéro ISBN correspondant à la version physique du livre à partir de laquelle la version PDF a été réalisée.

Relation

Etiquette: DC.Relation

Définition : Les relations de cette ressource avec d'autres ressources. Le but de cet élément est de fournir un moyen d'exprimer les relations formelles entre des ressources qui existent aussi en temps que ressources indépendantes. Par exemple des images dans un document, les chapitres d'un livre, ou les éléments d'une collection. Une définition formelle de relation est en cours. Utilisateurs et développeurs doivent comprendre que l'utilisation de cet élément est pour l'instant expérimentale.

Exemple de métadonnée

```
<meta name="DC.Relation"  
CONTENT="(SCHEME=URL)http://www.cerist.dz/dst/logo.gif>
```

Couverture

Etiquette: DC.Coverage

Définition : Les caractéristiques spatiales et/ou temporelles de la ressource. Une définition formelle de couverture est en cours. Utilisateurs et développeurs doivent comprendre que l'utilisation de cet élément est pour l'instant expérimentale.

Exemple de métadonnée

```
<meta name="DC.Coverage" CONTENT="Algérie">
```

Les éléments du Dublin Core sont intégrés dans l'étiquette existante HTML <meta>. Ces éléments sont invisibles sur la page. Il faut consulter le code source pour la voir. Ils n'ont aucun effet sur la présentation de la page.

On peut trouver la convention établie pour intégrer les métadonnées en HTML à <http://www.oclc.org:5046/~weibl/html-meta.html>

Les métadonnées sont intégrées comme suit :

```
META NAME="schema_identifier.element_name" CONTENT="string data"
```

Par exemple, l'auteur et le titre du présent article sont :

```
<META NAME="DC.creator" CONTENT="Aliouali Nadia" >
```

<META NAME="DC.title" CONTENT="Synthèse sur les METADATA" >

(DC-2) : Une année après , en Avril 1996, un deuxième Atelier toujours sponsorisé par l'OCLC et l'UKOLN (United Kingdom Office for Library and Information Networking), s'est tenue à l'Université de WARWICK en Angleterre.

Le thème retenu pour le meeting est l'identification et la résolution des obstacles pour l'application des DC comme description de ressources sur réseaux. les résultats de l'atelier ont dépassé de loin les espérances des organisateurs puisqu'un consensus a été achevé dans différents domaines :

1-les Dublin Core :

- Une syntaxe concrète pour les DC exprimée en forme de DTD en SGML.
- Conversion de cette syntaxe en HTML pour permettre au auteurs et fournisseurs de l'information de mettre leur création directement sur Web.

2-La Structure Warwick (The Warwick Framework) :

- la structure Warwick : une structure pour l'agrégation des Eléments Metadata;

3-Guide de Création et de Maintenance de Metadata

- Guide pour les auteurs pour la génération de descriptions de ressources;
- Guide pour les administrateurs de collections.

Les DC Eléments est une collection de treize éléments proposés au cours du premier atelier comme une description pour faciliter le recouvrement d'objets-documents dans un environnement réseau. Pour faciliter le développement de ces éléments, un certain nombre de contraintes ont été imposées durant le premier atelier :

- Seuls les éléments descriptifs des données ont été retenus;
- La discussion s'est restreinte aux éléments requis pour la recherche des objets-documents (DLO);
- Les éléments sont sémantiquement compréhensibles le plus large possible par différentes communautés scientifiques;
- L'Extensibilité est jugée comme caractéristique clé. les DC non pas l'intention de remplacer les autres descriptions plus connues mais plutôt une simple description susceptible d'unifier d'autres descriptions plus complexes.

c'est ainsi que cette caractéristique est très importante dans le sens qu'elle permet de lier les DC vers d'autres modèles de description plus riches.

- Les éléments sont modifiables, optionnels et répétitifs.

Le développement des DC est motivé par différentes utilisations projetées :

- Un format simple d'échange de métadonnées descriptives;
- Une self-description d'objets sur le réseau;
- Une interopérabilité sémantique entre différents domaines scientifiques;

Il est clair que les premières implémentations ont choisies le côté sémantique des DC pour développer des formats simples de description de ressource sur réseau. les DC ont une mission c'est celle de remplir le vide qui sépare la concision des index WEB et la richesse de modèles de description tel le MARC. la Simplicité est spécialement très importante dans le contexte de Metadata générée par l'auteur.

Avec l'absence d'une syntaxe clairement définie, les DC ont retenu l'intention de bon nombre de personnes incluses dans différents projets et ont finis par adopter cette description comme description modèle de leur ressource sur réseau.

Les Projets utilisant le DUBLIN CORE SET

- **Le Nordic Metadata Project :**

Les pays scandinaves mettent à jour un système de métadonnées pour faciliter le prêt inter-bibliothèque existant déjà (The Inter-Library Loan) et le service de délivrance de documents entre ces pays. les DC est l'une des descriptions qui ont été retenues dans le but de décrire les documents mis sur réseaux.

- **TURNIP :**

The Uniform Resource Locator Interopérability Project (TURNIP) initié par le Distributed Systems Technology (DSTC) en Australie , a produit un URN Resolution Service qui utilise les DC pour les URC metadata. Les DC sont aussi utilisés pour décrire les Rapports Techniques du DSTC. le DSTC a ajouté d'autres éléments au DC (Des éléments dits Administratifs comme URC-Type, Date-

création, Owner). Trois questions ont émergés lors de l'utilisation des DC à savoir: le besoin de grouper les éléments, une syntaxe commune pour l'échange des URC et les standards pour les qualifiants des éléments.

- **OCLC :**

L'office de Recherche de l'OCLC est en train d'essayer plusieurs applications potentielles des DC Eléments, incluant :

- Une évaluation préliminaire de DC comme une Interface de recherche dans La base de données OLUC. Une expérience commune entre l'OCLC et l'université de Santa Barbara en Californie est menée dans le cadre du projet Alexandria Digital Library.
- Le projet Spectrum explore différentes interfaces pour les utilisateurs basées sur les DC.
- Le projet Scorpion est un projet de recherche de l'OCLC dans le domaine de l'indexation automatique (automatic subject assignment) basée sur le système de Classification Décimale de Dewey. Le Spectrum-Scorpion fournira une description par les utilisateurs et une génération automatique d'une classification des metadata avec les DC.

- **The National Document and Information Service :**

Le projet NDIS est un projet commun entre la National Library of Australia et la National Library of New Zealand dont le but est de fournir un service de recherche sophistiqué pour les bases de données nationales et étrangères, un service de gestion des collections et l'état de l'art dans le domaine du service de diffusion de documents.

La première phase de ce projet va implémenter un service de recherche et demande de documents a travers des bases de données bibliographiques en MARC et suite de bases de données d'index, de répertoires et de thésauri dans différents formats.

Le NDIS utiles les DC comme un outil de détermination de metadata générique pour les données bibliographiques, avec une extension de ces éléments ou adoption d'autres métadonnées pour les données non bibliographiques.

Les DC servent comme modèle d'accès et de stockage générique d'une recherche à travers les bases de données.

- **Conversion des DC vers L'USMARC :**

Après le premier atelier, la Library of Congress a mis au point deux avant-projets de conversion des DC vers l'USMARC.

Il y'a eu depuis deux conséquences dans l'USMARC :

- le Champs 856 : pour les liens aux textes intégrales (URL) actuellement intégré dans l'UNIMARC Bibliographique;
- le Champs 720 : pour Uncontrolled Author Name : Nom d'auteurs incontrôlé.

L'un des buts de l'atelier de Warwick est d'identifier les obstacles qui stopperont l'élan des DC pour une large utilisation dans différents contextes.

Les premières discussions ont identifiés quatre domaines qui requièrent une plus grande attention:

- Une spécification d'une syntaxe ;
- Elaboration de guides pour les utilisateurs ;
- Identification des mécanismes d'extension ;
- Une spécification d'une structure pour l'accommodation de différentes variétés de metadata.

Cet atelier a porté ses fruits celles de permettre une appréhension des impératifs nécessaires pour un déploiement correct et efficace des DC, de définir le cadre d'expérimentation des DC et d'achever un consensus sur les actions à mener afin de standardiser cette description auprès d'organisme concerné par Internet comme Le World Wide Web Consortium (W3C) qui se charge de la maîtrise des formats et des normes circulant sur Web.

Une démarche promotionnelle de cette description est à mettre en oeuvre afin de donner à l'élan qu'ont suscités les Dublin Core du zèle dans l'accomplissement des tâches qui restent à entreprendre et à réaliser.

Il est à noter que cet atelier a permis aux Dublin Core de prendre une dimension internationale

d'autant plus que divers projets dans divers pays et continents ont pris à charge d'implémenter les DC et de les utiliser .

Le Projet BIBLINK:

C'est un projet commun a plusieurs Bibliothèques Nationales de l'Union Européenne. Il vise à établir un lien entre les éditeurs et les agences bibliographiques Nationales, pour l'échange des enregistrements de métadatas des différents articles nouvellement publiés.

DC.3 : Le Troisième atelier sur les Metadata (Dublin Core) s'est tenu à Dublin, Ohio à

l'initiative de l'OCLC et la CNI (Coalition for Networked Information).

L'objectif de cet atelier était de définir sur la base des éléments Metadata Dublin Core une description des images sur réseau. A l'issue de cet atelier, un consensus en est sorti.

En effet, les éléments retenus par les participants sont les même éléments des Dublin core.

Etant donnée l'objectif initial des Dublin Core, définir une description sémantiquement simple et facile pour son utilisation au repérage de l'information sur réseau, le résultat de cet atelier en est une consécration.

Mais l'enthousiasme suscité par ce résultat est modulé par les fortes recommandations pour la description des textes et des images.

L'image est-elle un objet-document ?

L'un des premiers sujets abordés lors de cet atelier se rapportait au problème de savoir si l'image est elle un DLO ou non. Si elle ne l'est pas ceci nécessitera la conception d'un autre modèle de Metadata pour décrire les images.

Le consensus a permis de statuer sur le fait que les images sont des objets-documents.

La définition caractéristique d'un DLO n'est pas dans son contenu texte ou image mais plutôt dans l'apparence de celui-ci pour l'utilisateur. C'est à dire l'image est une ressource (information) qui a les même attributs et apparence qu'un texte pour un utilisateur.

Le consensus réuni au cours de cet atelier et au cours des précédents ateliers n'est du au fait qu'à la longue expérience des différents participants dans le domaine des modèles de description de ressources. Ceci n'empêche pas qu'il faut un modèle conceptuel pour guider le pragmatisme des Dublin Core.

Durant cet Atelier, quelques une des définitions des éléments du dublin core ont été mise à jour, et deux éléments supplémentaire ont été ajoutés aux treize (13) initiaux.

1-Gestion des droits

Etiquette: DC.rights

Définition :Un lien sur les droits de reproduction, les droits d'utilisation ou à un service capable de fournir l'information sur les conditions d'accès à la ressource. Une définition formelle de droits est en cours. Utilisateurs et développeurs doivent comprendre que l'utilisation de cet élément est pour l'instant expérimentale.

Exemple de métadonnée

```
<META NAME="DC.Rights" CONTENT="Domaine public" >
```

2-Description

Etiquette : DC.Description

Une description textuelle du contenu de la ressource, y compris un résumé, dans le cas d'un document ou une description du contenu dans le cas de ressources visuelles.

DC .4 : CAMBERRA en Australie hébergea le quatrième atelier du Dublin Core en Mars 1997. Cet atelier a été sponsorisé par l'OCLC, DSTC (Distributed Systems Technology Center) et le NLA (National Library of Australia) . Cet atelier à connu l'emergence de deux camps. Le camp minimaliste et le camp des qualificateurs (structuralistes).

Les minimalistes considèrent que les formats complexes style USMARC ou TEI jouent bien leur rôle et que plus le Dublin Core reste simple, meilleur il sera, donc ils souhaitent garder les 15 éléments sans qualificateurs. Les structuralistes (qualificateurs), surtout lorsqu'ils pensent XML et RDF, commencent à trouver que dans bien des cas, le Dublin Core pourrait servir à décrire plus simplement des objets plus complexes que ceux que peuvent décrire les formats MARC pour cela il souhaitent ajouter des sous élément qui, affirment-ils sont à la fois utiles et nécessaires.

Un important travail a donc consisté à proposer un nombre très limité d'extensions de la normalisation. Il a été notamment question d'autoriser l'usage de sous-éléments , "shéma" et "type", pour préciser le rôle d'un élément. Pour prendre un exemple, dans l'USMARC un titre donne lieu à un nombre impressionnant de variantes. Dans le Dublin Core, après un grand nombre de discussions, en rappelant qu'il s'agissait de produire un standard destiné à l'interrogation , seuls deux types de titre ont été introduits : un titre non qualifié repère le titre principal, toutes les autres variantes sont regroupées sous le sous-élément «alternative». Par exemple, pour ce document, on pourrait avoir les metadata suivantes:

```
<META name="DC.Title" lang="fr" content="Les Métadata: Synthèse">
```

```
<META name="DC.Title.Alternative" lang="en"  
content="The Metadata : ">
```

L'ors de cet atelier trois qualificateurs ont été définis , mis à jour l'ors du cinquième workshop qui s'est tenu à Helsinki

- Norme: définir une norme officielle du contenu des données (ex , ISO, CDD, LCSH)
- Langue: Définir la langue réel de l'élément (cas du bilinguisme)
- Type : Définir la portée d'un élément (ex: l'élément date, pourrait être la date de création ou la date de mise à jour,)

DC . 5 : En Octobre 1997 , s'est tenu à HELSINKI , Finland le cinquième Atelier (DC-5) sponsorisé par l'OCLC et la bibliothèque nationale de Finland. Les objectifs de cet Atelier était de partager les expériences d'implémentation du Dublin Core , arrêter la définition des éléments et enfin définir les sous - éléments. Les résultats de cet atelier se résume dans un modèle de données formel (RDF). Les Dublin Core ont dux caractéristiques, leur adaptation aux documents textuels et leur utilisation pour faciliter la recherche sur internet. Cependant d'autres structures de métadonnées orientées vers d'autres usages ou spécifiques à d'autres types de documents, comme les images satellitaire ou l'imagerie médicale. C'est pourquoi le consortium W3 (W3C) a souhaité mettre en place un format commun à l'ensemble des métadonnées, qui est le RDF ou (Ressource Description Framework) . Le RDF est un ensemble de conventions qui supportera l'interopérabilité entre des applications qui échangeront des métadonnées. Dans un

autres sens permettre d'introduire dans un même document des metadata propres à plusieurs communautés. Par exemple, une page Web pourrait contenir un ensemble d'éléments provenant du Dublin Core - pour un usage général, par n'importe quel moteur de recherche courant et contenir des ensembles plus spécialisés destinés à d'autres applications plus sophistiquées. Parmi les partenaires actifs dans RDF, citons Grif, IBM, Microsoft, OCLC, Netscape, SoftQuad) La syntaxe du RDF sera exprimé en langage XML mais la sémantique sera définie par les besoins des usagers. Le RDF recevra les métadonnées comme les dublin core et autres types de métadonnées et aussi les métadonnées qui seront définies dans quelques années pour répondre à d'autres besoins de recherche documentaires.

Pendant cet atelier le document traitant des qualificateurs "shéma" et "type" présenté par REBECCA GUENTER (spécialiste d'USMARC à la Bibliothèque du Congrès US) a été mis à jour à la suite de nombreuses discussions. Le qualificateur "shéma" employé pour interpréter la valeur du contenu du document basé sur des normes externe l'orsque tandis que le qualificateur "type" est utilisé pour affiner la définition des éléments de données eux même.

Voici un ensemble des qualificateurs des 15 éléments du dublin core:

1. Titre

Shéma : non nécessaire

Type:

- DC.TITRE
- DC.ALTERNATIVE (employé pour tous les type de titre sauf le titre principale.

2. Auteurs ou Créateur

Shéma:

- LCNAF (Library of Congress Name Authority File)

Type:

- DC.CREATOR
- DC.CREATOR (Nom de personne)
- DC.CREATOR (Nom de compagnie)

- DC.CREATOR (Nom de personne (inclus tout type d'adresse ou email))
- DC.CREATOR (Nom de compagnie incluse adresse)

3. Sujet et Mots- clés

Shéma :

- (Non qualifié : Mots clés pris par défaut)
- LCSH 5Library of Congress Subject Headings)
- MESH (Medical Subject Headings)
- AAT (Art and architecture thesaurus)
- LCNAF (Library of Congress Name Authority File)
- DDC (Dewey Decimal Classification)
- LCC (Library of Congress Classification)
- NLM (National Library of Medicine Classification)
- UDC (Universal Decimal Classification)

Type: non nécessaire

4. La Description

Shéma:

- Le résumé est pris par défaut
- URL

Type : non nécessaire

5. Editeur

Shéma : non nécessaire

Type

- DC.PUBLISHER (non qualifié)
- DC.PUBLISHER (Nom personne)
- DC.PUBLISHER (Nom compagnie)
- DC.PUBLISHER (Nom personne + Adresse)
- DC.PUBLISHER (Nom compagnie + Adresse)

6. Autre contributeur

Shéma : LCNAF (Library of Congress Name Authority File)

Type:

- DC. CONTRIBUTOR(non qualifié)
- DC. CONTRIBUTOR (Nom personne)
- DC. CONTRIBUTOR (Nom compagnie)
- DC. CONTRIBUTOR (Nom personne + Adresse)
- DC. CONTRIBUTOR (Nom compagnie + Adresse)

7. Date :

Shéma :

- ISO 8601 par défaut
- ANSI X3.30
- IETF RFC 822
- Autres

Type :

- DC. DATE CREATION du contenu intellectuel
- DC. DATE CREATION / Modification de la présente forme
- DC. DATE FORMAL __ PUBLI
- DC. DATE ACCESSIBLE
- DC. DATE VALIDE (inclut la vérification)
- DC. DATE CREATION.ACQUISITION /Accession
- DC. DATE ACCEPTEE
- DC.DATE DATA GATHERIN (collecte de données)

8. Type de Ressource

Shéma : Nom nécessaire

Type: La liste de type de ressources est en voie de développement

9. Format

Shéma:

- IMT (i.e.MIME)
- DCPMT (Dublin Core Physical Medium Type)

Type: Non Nécessaire

10. Identificateur de ressource

Shéma:

- URL est pris par défaut
- URN (Uniform Ressource Name)
- ISBN (International Standard Book Number)
- ISSN (International Standard Serial Number)
- SICI (Serial Item and Contribution Identifier)
- FPI (Formal Public Identifier)

Type : Non nécessaire

11. La Source

Shéma:

- Le texte libre est pris par défaut
- URI
- URN
- ISBN
- ISSN

Type : Non nécessaire

12.La Langue

Shéma :

- IETF RFC 1766

- Z39.53
- ISO 639-1
- ISO 639-2/B (après la publication finale)

Type : Non nécessaire

13. Relation

Shéma:

- Texte libre pris par défaut
- URL
- URN
- ISBN

Type:

- Créatif (e.g. traduction, annotation etc...)
- Mécanique (copier, chargement du format)
- Version (edition, draft)
- Inclusion (collection, part)
- Référence (situation)

14. Couverture

Shéma : Sera déterminé bientôt par le groupe de travail sur l'élément COVERAGE (couverture)

Type: Cette liste a été déterminé par le groupe de travail sur l'élément COVERAGE (couverture)

- DC.COVERAGE .PERIODNAME
- DC.COVERAGE.T
- DC.COVERAGE.X
- DC.COVERAGE.Y
- DC.COVERAGE.Z

- COVERAGE.POLYGON
- DC.COVERAGE LINE
- DC.COVERAGE 3D

15. Gestion des Droits

- **Texte libre pris par défaut**
- **URL**
- **URN**

Une autre série d'atelier sur les METADATA a eu lieu à l'Uxembourg le 1-2 décembre 1997, organisé par l'Unité Edition électronique et bibliothèques de la commission européenne. Ce séminaire à réuni des participants à des projets du secteur bibliothèques mis en œuvre au titre du programme application télématiques ainsi que d'autre secteurs de ce programme.

Cette série d'ateliers à pour objectifs l'établissement d'une plate-forme pour la coordination entre des projets relatifs aux métadonnées, ils visent également à sensibiliser une certaine communauté européenne aux progrès accomplis dans le domaine de la normalisation.

Le séminaire à pris en charge des travaux dirigés sur les métadonnées et sur "Dublin Core", menés par l'Office for Library and Information Networking britannique (UKOLN), des présentations de projets en cours en scandinavie et au R-U, et trois sessions en petits groupes sur la création, la récolte et l'extractions des métadonnées.

Les participants dans ce séminaire ont reconnu la nécessité des metadata pour assurer l'interopérabilité des ressources sur INTERNET, de nombreux point ont été soulevés:

- Actuellement il n'y a aucun responsable officiel pour le suivi et la maintenance du DC.
- Le développement du DC se fait au sein d'un groupe informel expert invités une ou deux fois par ans (Dublin Core Workshop série).
- L'état actuel du DC est instable, durant les différents séminaire organisé jusqu'à ce jour, des changements sont apportés sans qu'il est aucune convergence vers une version finale du standards METADATA.
- L'utilisation du format Dublin core Metadata n'est pas appuyé par la publication de principe directeurs (guidelines), la terminologie des éléments du DC n'est pas

évidentes pour l'ensemble des utilisateurs, ce qui peut les emmener à de différentes interprétations affectant le vrai sens de portabilité.

- Il a été identifié que l'adoption du DC est très longue, les éditeurs ne veulent pas investir à fournir des METADATA si les services d'indexation ne les collectent pas pour une indexation sélective en cas d'absence d'éléments pertinents, dans ce cas le DC pourrait ne pas exister en réalité.

Pour cela, les participants sont parties avec un bon nombre de recommandations à prendre en considération pour promouvoir et encourager l'utilisation du DC, on peut citer:

1. Clarifier le contrôle des différentes versions du DC et leur maintenance. Le groupe fondateur du DC,
2. Assister des projets pilotes afin de développer des expériences dans le domaine et de créer une base de critique pour le DC
3. La commission européenne et les bibliothèques nationales de chaque pays doivent jouer un rôle pour encourager la fourniture des Dublin Core dans les documents dans le cadre du dépôt légal, etc

DUBLIN CORE ET METADATA

Les METADATA sont des données sur les données, c'est pourquoi l'UNIMARC est considérée comme étant un support des metadata. Ce dernier a été développé pour permettre l'échange bibliographique entre différents systèmes de catalogage alors que les METADATA sont un jeu de 15 éléments définies pour faciliter la recherche et le repérage des ressources sur INTERNET.

Donc suite à cette définition, les données dans le format UNIMARC vont devenir quelques fois si c'est pour toujours des METADATA

TABLEAU DE CONCORDANCE ENTRE LES DUBLIN CORE ET L'UNIMARC

| Dublin Core | UNIMARC |
|--------------------|--|
| Title | 200 \$a Title Proper 200 \$e Other Title Information (for subtitle) 517 \$a Other Variant Titles (for other titles) |
| Creator | 700 \$a Personal Name - Primary Intellectual Responsibility, or if more than one: 701 \$a Personal Name - Alternative Intellectual Responsibility 710 \$a Corporate Body Name - Primary Intellectual Responsibility, or: 711 \$a Corporate Body Name - Alternative Intellectual Responsibility 200 \$f First Statement of Responsibility |
| Subject | 610 \$a Uncontrolled Subject Terms 606 Topical Name Used as Subject (for LCSH and MeSH) 675 UDC 676 DDC 680 LCC 686 Other Classification Systems |
| Description | 330 \$a Summary or Abstract |
| Publisher | 210 \$c Name of Publisher, Distributor, etc. |
| Contributors | 701 \$a Personal Name - Alternative Intellectual Responsibility 711 \$a Corporate Body Name - Alternative Intellectual Responsibility 200 \$g Subsequent Statement of Responsibility (if role known) |
| Date | 210 \$d Date of Publication, Distribution, etc. |
| Type | 608 Form, Genre or Physical Characteristics Heading |
| Format | 336 \$a Type of Computer File (provisional) |
| Identifier | 001 (mandatory for UNIMARC) 010 (ISBN) 011 (ISSN) 020 (National Bibliography Number) 300 \$a (URL) |

| | |
|----------|--|
| Source | 324 Original Version Note |
| Language | 101 Language of the Item 300 General Note |
| Relation | 300 General Note |
| Coverage | 300 General Note |
| Rights | 300 General Note |

Edition des éléments

Deux solutions peuvent être envisagées lors de la création de ces éléments:

- Ces éléments peuvent être placés au sein même des fichiers HTML soit manuellement soit en générant automatiquement les éléments et/ou leur contenu à l'aide d'un formulaire

[HTTP://www.ukoln.ac.uk/metadat/dcdot/](http://www.ukoln.ac.uk/metadat/dcdot/)

Les données du présent document ont été générées par ce programme on peut le visualiser grâce à l'option "view source" d'un browser.

Exemple

```
<META NAME="DC.Title" CONTENT="METADATA et DUBLIN CORE:
Synthèse">
<META NAME="DC.Creator" CONTENT="Aliouali Nadia">
<META NAME="DC.Subject" CONTENT="Dublin Core; DC;
generator; editor; Warwick Framework; Calogage du
Web;Ressources électronique;Recherche sur internet
RDF">
<META NAME="DC.Description" CONTENT="Synthèse sur la
genèse des dublin core et leur développement pour une
bonne description des ressources sur internet et leurs
repérage "> <META NAME="DC.Publisher" CONTENT="Centre
de Recherche sur l'Information Scientifique et
Technique">
<META NAME="DC.Date" CONTENT="1998-06-02">
<META NAME="DC.Type" CONTENT="Text">
<META NAME="DC.Format" CONTENT="text/html">
<META NAME="DC.Language" CONTENT="FRE">
```

Les éléments suivant sont générés suite au remplissage du formulaire suivant



DCDOT

DUBLIN CORE GENERATOR

If necessary, edit the values in the boxes below, and ...

Title

METADATA (DUBLIN CORE) : Synthèse



Author or Creator

ALIOUALI NADIA



Subject or keywords

Dublin core ; DC ; editeur ; Warwick Framework;
XML; RDF ; Catalogage du Web ; Description des
Ressources du Web



| | | |
|--------------------|---|---|
| Description | Synthèse sur le genèse des dublin core et leur développement pour une bonne description des ressources sur Internet et leurs repérage | ▲ |
| | | ▼ |
| ◀ | | ▶ |

| | | |
|------------------|---|---|
| Publisher | CENTRE DE RECHERCHE SUR L'INFORMATION SCIENTIFIQUE ET TECHNIQUE | ▲ |
| | | ▼ |
| ◀ | | ▶ |

| | | |
|--------------------------|-----------------------------------|---|
| Other contributor | METADATA (DUBLIN CORE) : Synthèse | ▲ |
| | | ▼ |
| ◀ | | ▶ |

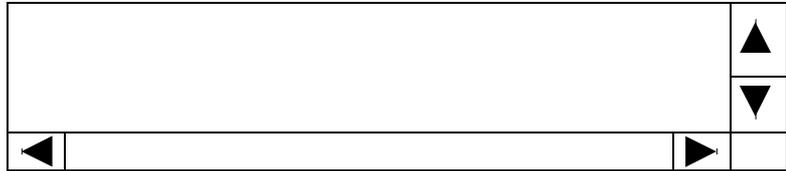
| | | |
|-------------|------------|---|
| Date | 1999-07-25 | ▲ |
| | | ▼ |
| ◀ | | ▶ |

| | | |
|-----------------------|------|---|
| Ressource Type | Text | ▲ |
| | | ▼ |
| ◀ | | ▶ |

Simple chooser : [None] ▼ Complex chooser : [None] ▼

| | | |
|-----------------------------|------------|---|
| Format | Text /Html | ▲ |
| | | ▼ |
| ◀ | | ▶ |
| Ressource identifier | | ▲ |
| | | ▼ |
| ◀ | | ▶ |
| Source | | ▲ |
| | | ▼ |
| ◀ | | ▶ |
| Language | | ▲ |
| | | ▼ |
| ◀ | | ▶ |
| Relation | | ▲ |
| | | ▼ |
| ◀ | | ▶ |
| Coverage | | ▲ |
| | | ▼ |
| ◀ | | ▶ |

Rights Management



Les éléments suivants peuvent être ajoutés à l'en-tête du document HTML par l'usage du copier/coller une fois le résultat donné par l'utilitaire

- Les données peuvent être séparées des données et être stockées dans une base de données à part. Lorsqu'une demande est adressée au serveur, les données et métadonnées correspondantes sont rassemblées en un fichier généré "on Fly"

Questions à régler et développements futurs

M. Weibel, chercheur scientifique principal à l'OCLC, lors d'un atelier sur les Metadata à la Bibliothèque Nationale du Canada, a relevé trois questions importantes encore à résoudre.

- La première question tourne autour du problème du multilinguisme.
- La deuxième question tourne autour du fait qu'il faut déterminer la responsabilité de l'inscription des éléments et des qualificatifs,
- La troisième et celle portant sur le perfectionnement à apporter aux éléments et aux qualificatifs, en particulier ceux qui sont liés à la couverture, à la relation et à la date.

Enfin, l'architecture des métadonnées, autrement dit comment elles sont intégrées, représente également un problème.

Le Dublin Core/Warwick Framework en est encore au stade du développement, mais son utilisation augmente de jour en jour. Par exemple, le Nordic Metadata Project, Distributed Systems Technical Centre, le Monticello de Solinet, ainsi que le gouvernement du Danemark se servent des éléments de métadonnées du Dublin Core dans des projets pilotes.

Adresses PURL, URL et URN

Au début de son exposé, M. Weibel a parlé de la difficulté d'être à jour sur les ressources d'Internet. Les adresses URL ne sont plus pertinentes ou échouent. Les données sont retirées ou modifiées, ou disparaissent (message d'erreur 404 : ressource introuvable)

M. Weibel a utilisé son adresse PURL comme exemple de la façon dont fonctionne ce type d'adresse. Son adresse PURL est <http://purl.org/net/weibel>. Cette adresse PURL indique <http://www.oclc.org:5046/~weibel>, l'adresse URL réelle. Si l'adresse URL de M. Weibel devait être remplacée par <http://foo.bar.com/staff/~weibel>, seule la direction de l'adresse PURL devrait être modifiée. Les gens qui se servent de son adresse PURL n'auraient pas à connaître la modification ou se soucier de la perdre et de la retrouver.

Les adresses PURL ajoutent un niveau de stabilité à Internet et sont basées sur des protocoles standard. Comme dit M. Weibel : " La persistance est une fonction des organisations, pas la technologie ". En vue d'ajouter de la persistance et de la stabilité, OCLC diffuse le code des adresses PURL gratuitement à <http://purl.org>. Jusqu'à présent, 350 téléchargements du code des adresses PURL ont été effectués du site d'OCLC.

Conclusion:

Tout au long de cette synthèse, nous avons passé en revue les METADATA du DUBLIN CORE SET et leurs évolutions. Les éléments du DUBLIN CORE de part leurs caractéristiques, représentent une valeur ajoutée à l'information sur le WWW.

Mais on peut se demander si tout ceci sera réellement respecté et utilisé?

Pour le moment, ces standards ne sont pas généralisés mais il est fort probable qu'ils puissent se généraliser assez vite. Les outils de recherche tels que les robots ne proposent pas à l'utilisateur d'exploiter ce type de données pour le moment mais ils le feront certainement dans un proche avenir.

Au bout de trois ans, le DUBLIN CORE SET est déjà opérationnel dans plus de 40 projets de part le monde dont certains commencent à utiliser cette standardisation.

Par exemple :

- La bibliothèque nationale d'Australie et de Nouvelle Zélande : le "DUBLIN CORE DATA " est utilisé pour normaliser un ensemble de formats différents existant pour les ressources (MARC et autres).
- "The nordic Metadata Project " pour partager des ressources électroniques entre différents pays nordiques, des conversions entre les formats NORMARC et DUBLIN CORE DATA ont été développées

Certaines communautés ont aussi développé des Metadata plus spécifique à leurs besoins : c'est le cas pour les données géospatiales et pour les données gouvernementales.

- Le WARWICK FRAMEWORK adopté en 1996 et qui consistait en une architecture permettant d'assembler des METADATA à structures diverses vers un schéma standardisés

Enfin pour terminer il faut souligner le fait que nous avons absolument besoin de Métadonnées. Car il est impossible de penser cataloguer une masse aussi importante qu'Internet selon les normes traditionnelles des bibliothèque (MARC-RCAA2) vue le facteur temps.

Après 25 ans d'effort conjoint de milliers de bibliothèques, le catalogue OCLC ne comporte que 40 million de notices, tandis que 5 ans après l'apparition du premier robot de recherche sur le Web, on compte plus de 320 millions de page Web. Les Métadonnées représentent la solution la plus fiable pour le traitement de cette masse d'information sans cesse mouvante ne serait-ce que pour éviter qu'une simple recherche ne génère autant d'occurrence qui ne sont des fois pas utiles. Ainsi les METADATA du DUBLIN CORE s'affirme directement un future standard de métadata international pour cela le DC a entrepris le processus de normalisation auprès de NISO (USA) et pense l'entreprendre prochainement avec l'ISO (international)

Références Bibliographiques

1- Gislaine CHARTRON

Repérage de l'information sur INTERNET : nouveaux outils, approche bibliothéconomique et micro-structure

2- Youcef AMEROUALI

Prise en compte du système descriptif de documents PROFILDOC dans le cadre des METADATA (DUBLIN CORE).__ Mémoire de DEA en science de l'information et de la communication

WEBOGRAPHIE

1- Clifford LYNCH

Searching the internet

[Http://www.sciam.com/0397_issue/0397Lynch.html](http://www.sciam.com/0397_issue/0397Lynch.html)

2- S WEIBLE, J GIDBY, E MILLER

OCLC / NCSA Metadata report

<http://www.oclc.org:5047/oclc/research/publications/weible/métadata>

3- L DEMPSEY, S WEIBLE

The WARWICK Metadata workshop : A Framework for the deployment of resource description

<http://www.ukoln.ac.uk/dlib/dlib/july96/07weible.html>

4- L BURNARD, E MILLER, L QUIN

A Syntax for Dublin Core Metadata -Recommandation

<Http://www.uic.edu/~cmsmcq/tech/metadata/dc4-notes.html>

5- DC-4 : The 4th Dublin Core Metadata Workshop

<http://www.ukoln.ac.uk/metadata/ressourcess/dc4-notes.html>

6- DC-5 the 5th Dublin Core Metadata Workshop /Helsinki, Finland, October 6-8 1997

<Http://www.linnea.helsinki.fi/meta/DC5.html>

7- Preben HANSEN

User guidelines for Dublin Core création (Nordic Metadata Project)

Http://www.sics.se/~preben/DC/DC_guide./html

8- The Dublin Core and the WARWICK FRAMEWORK :March 1995-September 1997

<Http://www.dlib/january98/01tiele.html>

8- D LERINCKX

Metadata Workshop , DGXIII/E-4, Luxembourg-1-2 Décembre 1997

Http://www.bib.ulb.ac.be/html/meta_dc.htm

9- DC-dot -Dublin Core Generator

<http://www.ukoln.ac.uk/metadata/dcdot>