


enssibEcole Nationale Supérieure des Sciences
de l'Information et des BibliothèquesMémoire de DEA
Sciences de l'Information et de
la CommunicationOption :
Système d'information
documentaire

**PRISE EN COMPTE DU SYSTEME DESCRIPTIF
DE DOCUMENTS PROFILDOC DANS
LE CADRE DES METADATA (DUBLIN CORE)**

Réalisé par : Mr. **YOUCEF AMEROUALI**Sous la direction de Madame **Sylvie Lainé-Cruzel**

Laboratoire Recodoc

Université Lyon1

1997 / 1998


Université Lumière
Lyon 2Ecole Nationale supérieure
des Sciences de l'information
Et des BibliothèquesUniversité Jean Moulin
Lyon 31998
DEA
1

enssib

Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques

Mémoire de DEA
Sciences de l'Information et de
la Communication

Option :
Système d'information
documentaire



**PRISE EN COMPTE DU SYSTEME DESCRIPTIF
DE DOCUMENTS PROFILDOC DANS
LE CADRE DES METADATA (DUBLIN CORE)**

Réalisé par : Mr. **YOUCEF AMEROUALI**

Sous la direction de Madame **Sylvie Lainé-Cruzel**

Laboratoire Recodoc

Université Lyon1

1997 / 1998

Université Lumière
Lyon 2

Ecole Nationale supérieure
des Sciences de l'information
Et des Bibliothèques

Université Jean Moulin
Lyon 3

**PRISE EN COMPTE DU SYSTEME DESCRIPTIF
DE DOCUMENTS PROFILDOC DANS LE
CADRE DES METADATA (DUBLIN CORE)**

Youcef Amerouali

Sous la direction de

Madame Sylvie Lainé-Cruzel

Laboratoire Recodoc

Université Lyon1

Résumé :

Les standards de metadata ont été développés dans le but d'organiser, de représenter et de décrire l'information numérique. Les premiers standards importants qui furent mis en place sont des standards fédéraux américains, pour les domaines géospatial et administratif. D'autres standards furent développés pour l'information en réseau, dont l'un émergea du lot pour sa simplicité et ses qualités. Nous nous sommes basés sur son architecture pour structurer les éléments de description de Profildoc en éléments de metadata, avec les mêmes propriétés que le Dublin Core.

Descripteurs :

Metadata, information, description, structure, syntaxe, TEI, Dublin Core, Profildoc, unité documentaire, profil utilisateur, représentation, propriété, affiliation, communauté.

Abstract :

The standards of metadata have been developed for the purpose to organize, to represent and to describe the numerical information. The first important standards that were put in place are American federal standards, for areas géospatial and administrative. Others standards were developed for the information in system, whose emerged a the share for its simplicity and its qualities. We have been based on its architecture to structure elements of description of Profildoc in elements of metadata, with the same properties that the Dublin Core.

Descriptors:

Metadata, information, description, structure, syntax, TEI, Dublin Core, Profildoc, documentary unit, user profile, representation, property, affiliation, community.

TABLE DES MATIERES

INTRODUCTION

Première partie :

INTRODUCTION AUX METADATA

| | |
|---|----|
| 1) DÉFINITION : | 7 |
| 2) HISTORIQUE : | 8 |
| 3) LES DIFFÉRENTS STANDARDS DE METADATA : | 13 |
| 3.1) <i>CSDGM</i> | 13 |
| 3.2) <i>GILS</i> | 17 |
| 3.3) TEXT ENCODING INITIATIVE : | 22 |
| 3.4) IAFA TEMPLATES..... | 24 |
| 3.5) SOIF | 29 |
| 3.6) IMS | 30 |
| 3.7) <i>MCF</i> | 32 |
| 3.8) <i>FSMS</i> | 34 |
| 3.9) <i>CERA</i> | 35 |
| 3.10) <i>CERES</i> | 36 |
| 3.11) <i>MRLC</i> | 37 |
| 3.12) <i>CSDNGM</i> : | 39 |
| 3.13) <i>ICPSR</i> | 40 |
| 3.14) <i>NEDRES</i> | 41 |
| 3.15) <i>DUBLIN CORE</i> : | 43 |
| 4) METADATA ET OUTILS DE RECHERCHE : | 68 |
| 5) CONCLUSION : | 69 |

DEUXIEME PARTIE:

| | |
|--|------------|
| 1- INTRODUCTION AU PROJET PROFILDOC: | 72 |
| 1.1-INTRODUCTION : | 72 |
| 1.2- PRINCIPE DE PROFILDOC : | 76 |
| 1.3- LA PHASE D'INTERROGATION : | 76 |
| 2- ÉLÉMENTS DESCRIPTIFS DE PROFILDOC:..... | 80 |
| 2.1-PROPRIÉTÉS PROPRES AU DOCUMENT:..... | 80 |
| 2.2-PROPRIÉTÉS PROPRES À L'UNITÉ DOCUMENTAIRE: | 81 |
| 3-STRUCTURATION DE PROFILDOC SELON LE DUBLIN CORE: | 82 |
| 4- DÉFINITION DES ÉLÉMENTS DE METADATA DE PROFILDOC:..... | 89 |
| 5- DÉFINITION DE QUALIFICATEURS PROPRES À PROFILDOC: | 90 |
| 5.1-DÉFINITIONS: | 90 |
| 5.2-QUALIFICATEURS PROPOSÉS POUR LES ÉLÉMENTS DE PROFILDOC:..... | 91 |
| 6- SYNTAXE DE TRANSCRIPTION EN HTML :..... | 96 |
| 6.1- DÉFINITIONS: | 96 |
| 6.3- SYNTAXE HTML 4.0 : | 104 |
| 7- METADATA ET PROFIL UTILISATEUR : | 105 |
| 7.1. LE PROFIL UTILISATEUR « CLASSIQUE » : | 105 |
| 7.2. PROFIL UTILISATEUR ET DUBLIN CORE : LE PROJET NEWSAGENT | 107 |
| 9-CONCLUSION :..... | 112 |

BIBLIOGRAPHIE

| | |
|--|-----|
| 1- OUVRAGES : | 114 |
| 2- ARTICLES DE PERIODIQUES: | 114 |
| 3- DOCUMENTS RETROUVES SUR INTERNET :..... | 117 |
| 3.1- DOCUMENTS GÉNÉRAUX: | 117 |
| 3.2- DUBLIN CORE :..... | 118 |

| | |
|---|-----|
| 3.4-ENCODED ARCHIVAL DESCRIPTION:..... | 124 |
| 3.5-STANDARDS DE METADATA GEOSPATIALS :..... | 124 |
| 3.6-GOVERNMENT INFORMATION LOCATOR SERVICE :..... | 125 |
| 3.8- IAFA/WHOIS++ TEMPLATES : | 126 |
| 3.9- MARC :..... | 127 |
| 3.10-META CONTENT FORMAT :..... | 127 |
| 3.11-PICS :..... | 127 |
| 3.12-SOIF :..... | 128 |
| 3.13-TEXT ENCODING INITIATIVE :..... | 128 |
| 3.14-AUTRES DOCUMENTS :..... | 129 |

INTRODUCTION

Face à l'évolution rapide des moyens d'information et de communication, une masse importante d'informations est produite chaque jour à travers des milliers de livres, des dizaines de milliers d'articles de périodiques et d'innombrables pages web. Par conséquent, la nécessité de structurer l'information et de la rendre utilisable et accessible de la façon la plus optimale possible ne s'est jamais autant fait sentir. Pour cela, il faut pouvoir décrire cette information d'une façon synthétique. Dans le milieu des bibliothèques, la notice bibliographique constitue un moyen de description et de représentation de l'information, très utilisé. Pour l'information numérique stockée dans des bases de données, ou en réseau, ce besoin s'est fait sentir aussi très tôt. Ce fut le cas pour les bases de données géographiques : leurs gestionnaires ont ainsi été amenés à développer des standards de metadata, pour décrire leur information.

Devant le foisonnement d'informations sur Internet, des initiatives pareilles se mirent en branle ; et d'autres standards de metadata furent développés. L'un d'eux émergea du lot, de par sa simplicité d'abord et ses qualités ensuite ; et fut très vite adopté par une large communauté, pour des besoins aussi divers que spécifiques.

Dans le cadre de notre mémoire de DEA, nous allons passer en revue les principaux standards de metadata ; en voir la genèse et identifier la nature, la sémantique et le codage des informations qui sont prises en compte. Après cela, nous allons comparer la nature des informations recueillies dans ce système de caractérisation avec la nature des informations exploitées dans Profildoc. Nous devons, à la faveur de cela, proposer un mode de prise en compte des données utiles à Profildoc, dans le contexte des metadata.

Première partie

1) Définition :

Les metadata (ou métadonnées) peuvent être définies comme étant des données relatives à d'autres données (*data about data* : données sur des données). Par conséquent, une notice catalographique classique peut-être considérée comme une métadonnée. Le terme « metadata » est surtout utilisé pour désigner l'information « lisible par machine » concernant des fichiers de données « lisibles par machine » : donc ce terme désigne en quelques sortes une information référentielle sur des données électroniques.

Utilisée dans le contexte de l'information numérique géospatiale, metadata sera perçue comme l'information de fond qui décrit le contenu, la qualité, les conditions et autres propriétés et caractéristiques des données.

Tim BARNERS L. du **W3C** définit les metadata comme suit :

« Metadata est une information compréhensible (interprétable) par machine (ordinateur) sur des ressources d'information du Web ou d'autres choses (autres sources de données). L'expression « compréhensible par machine » est la clef. Nous parlons ici de l'information que des logiciels traitent pour nous faciliter la vie, nous assurer que nous obéissons à nos principes, à la loi, vérifier que nous pouvons croire en ce que nous faisons, et exécuter tous les travaux régulièrement et rapidement. Metadata définit bien la sémantique et la structure. Metadata a été appelée ainsi à sa création; et est encore actuellement une information sur des ressources du Web, ou des données sur des données. Dans l'avenir lorsque les metadata, les langages, les moteurs de recherches seront plus développés ; ils formeront une base plus forte pour l'information « compréhensible par machine » du Web, sur n'importe quoi : les gens, les choses, les concepts et les idées. Gardons dans notre esprit ce modèle (préconisé), bien que la première étape soit de réaliser un système d'information sur l'information. »

2) Historique :

2.1) Première époque :

Le premier standard de metadata, le format MARC (Machine Readable Cataloguing) fût développé en 1965 par la bibliothèque du Congrès américain à des fins d'échanges de notices bibliographiques sur bandes magnétiques. A côté du format USMARK vers lequel évolua le format MARC, apparurent rapidement d'autres formats dérivés de ce dernier. Ce fut le cas d'abord en 1968 de UKMARC qui fut établi par la British National Bibliography en Grande-Bretagne. La bibliothèque Universitaire de GRENOBLE (France) définit sur la même lancée, un format analogue en 1970 : Monocle. A partir de là, ce fut le déferlement : AUSMARC en AUSTRALIE CANMARC au Canada, DANMARC au DANEMARK, et finalement INTERMARC pour l'EUROPE. Ce foisonnement de formats nationaux risquait de devenir un sérieux obstacle aux échanges d'information bibliographique entre pays. Ce qui amena l'IFLA a développer un format d'échange international : UNIMARC. La première version de ce format apparut en 1977. Il fut révisé et complété en 1987 et en 1994. Actuellement une instance de l'IFLA, le *Permanent Unimarc Committee* est charge de son évolution et de la promotion de son utilisation dans le monde...

Quant au terme *metadata*, il fut employé pour la première fois par un informaticien américain **Jack E. Meyers** en définissant des architectures informatiques appelées plus tard metamodèles, au début de l'été 1969. Il fit d'ailleurs des recherches dans des publications et des bases de données pour s'assurer que ce mot n'a pas été utilisé avant lui. Après confirmation, il décida en 1986 d'enregistrer le mot « metadata » aux USA comme une marque d'une compagnie de services informatiques. Cela n'a évidemment aucun lien avec le sens que nous connaissons au mot metadata, aujourd'hui.

En novembre 1987, un projet de norme d'échange de documents électroniques (TEI pour Text Encoding Initiative) fût lancé, lors d'une conférence préparatoire qui s'est tenue au VASSAR Collège de New York. Les objectifs généraux de la TEI ont été définis par la résolution finale,

d'ailleurs connu sous le nom de 'Principes de Poughkeepsie' (du nom de l'endroit où s'est tenue la conférence). La TEI connue deux cycles de développement :

- ✓ Le premier cycle (1988\1990) déboucha sur la publication d'une ébauche des recommandations de la TEI (document TEI P1).
- ✓ Le deuxième cycle (1990\1994) permit la publication de la version officielle des recommandations de la TEI, en mai 1994.

De par leur structure, les entêtes TEI peuvent être perçues comme des metadata.

2.2)Deuxième époque :

Nous situons le début de cette deuxième époque au moment où des agences gouvernementales Américaines, grandes utilisatrices des systèmes d'informations géographiques ont ressenti le besoin d'organiser et de cataloguer l'information géospatiale ; ce qui était devenue une nécessité pour faciliter l'accès à cette dernière.

Dés le 19 octobre 1990,une circulaire présidentielle (A16) américaine établissait des missions pour une structure fédérale le *FGDC(Fédéral Geographic Data Commitee)* de développement d'un standard de catalogage de l'information géospatiale. Cela permit le lancement d'un standard de metadata : le *CSDGM* (Content Standard Data for Geospatial Metadata) qui fut basé sur le *SDDS* (Spatial Data Transfert standard)et l'*USMARC*.

Le CSDGM fut rapidement développé et implanté au niveau des agences gouvernementales entre 1993 et 1994.

En parallèle, d'autres standards de metadata se développaient aux USA, dont les plus connus :

- ✓ *IIFA* (Internet Anonymous Ftp Archives) standard conçu par l'IETF (Internet Engineering Task Force) pour cataloguer les archives des sites ftp.
- ✓ Le *GILS* (Gouvernement Information Locator Service) élaboré pour cataloguer et ainsi faciliter l'accès à l'information détenue par les organismes

gouvernementaux. En décembre 1994, le GILS fût adopté comme norme fédérale à laquelle devait se conformer tous les départements d'état.

Ce même standard a été adopté aussi par le Canada en 1995 pour cataloguer toute l'information des organismes gouvernementaux.

✓ Les standards développés par les agences américaines de l'environnement :

❖ *SMSP* (Scientific Metadata Standards Projects) en 1994 par l'EPA
(US Environment Protection Agency)

❖ *CERES*(California Environmental Ressources Evaluation System)
en 1995.

✓ *MRLC* (Multiple Résolution Landscape Characteristics) développé par le NBS (National Biological Survey) structure dépendant du ministère américain de l'intérieur.

2.3)Troisième époque :

Nous situons le début de cette dernière époque en 1995, c'est à dire à l'année où, sous la houlette d'OCLC (Online Computer Library Center) un colloque international fût organisé à DUBLIN, dans l'état de l'OHIO (USA). Il permit d'aboutir au lancement d'un standard plus simple et ouvert : le Dublin Core.

Celui-ci est supporté par une série de workshops qui lui assurent un état d'évolution permanent. Au workshop, qui s'est tenu a l'université de Warwick, en avril 1996 au Royaume-Uni, une syntaxe particulière est proposée pour le Dublin Core, qui est approuvée par le W3C. Cela le raccroche à des développements majeurs du W3C, que sont la RDF(Ressource Description Framework)et XML. Une autre conférence d'OCLC sur les metadata, relative aux images accessibles en réseau, qui s'est tenue en septembre 1996, a rendu possible l'utilisation du Dublin Core dans le domaine des images. Enfin le quatrième workshop qui a eu lieu a Camberra en Australie en mars 1997 fit l'unanimité sur une proposition d'élever le Dublin Core à un standard international.

Plus d'une quarantaine de projets, se basant sur ce standard commencent à s'implanter un peu partout dans le monde. Des standards dérivés ont vu aussi le jour tel le standard de l'EDNA (Educationel Network Australy)ou le

standard du secteur de l'éducation américaine : GEM metadata (**G**ateway to **E**ducational **M**aterials).

Il faut signaler qu'en octobre 1995 fut fondée aussi la '*Metadata Coalition*', association regroupant alors 53 compagnies d'informatique (dont IBM et Microsoft) et d'autres organismes, dans le but de définir un ensemble de spécifications standards pour l'interchangeabilité et la prise en charge des metadata par des outils logiciels. Cela aboutit à un projet dénommé : **MDIS** (Metadata Interchange Spécification Initiative).

Devant ce début de prolifération de standards de metadata, et sur les recommandations de l'ISO (International Standardisation Organisation), s'est tenue une conférence du 8 au 11 juillet 1997, à **BERKELEY** (CALIFORNIE\USA). Celle-ci, sponsorisée par l'*Agence de Protection de l'Environnement U.S.*, l'*Online Computer Library Center* et la *Metadata Coalition*, avait pour but de dégager des recommandations ou des normes pour la mise en place ou le développement de standards de metadata afin d'assurer pour ces derniers :

- ✓ l'interopérabilité ;
- ✓ le mapping ;
- ✓ l'évolution (caractère évolutif) ;
- ✓ la flexibilité .

Avant de terminer ce bref historique, nous proposons ci-après un graphe portant la légende : *Genealogy of Environmental Directories*, établi par Bruno Kestemont (CESSE=ULB\BRUXELLES). Nous y retrouvons certains standards de metadata dans l'ordre chronologique de leurs créations, le tout convergeant vers un probable format d'échange commun... .

3) Les différents standards de metadata :

3.1) Content Standard Digital Geospatial Metadata :(CSDGM)

Le CSDGM est un standard de metadata développé par le FGDC (Federal Geographic Data Committee). La première version fut approuvée le 8 juin 1994. Ce standard comporte 334 éléments, dont 119 sont prévus uniquement pour référencer d'autres éléments. C'est un système ouvert, c'est à dire qu'il est possible de lui rajouter d'autres éléments (ou d'en enlever). Ses utilisations majeures sont :

- ✓ aider a organiser et a maintenir les données géospatiales ;
- ✓ fournir de l'information sur des données pratiques et théoriques, fournir de l'information pour interpréter et traiter des données reçues par un transfert de sources externes.

Les éléments de données du CSDGM fournissent de l'information sur :

- l'identification (information de base sur le lot de données) ;
- la qualité des données ;
- l'organisation des données géospatiales ;
- les références géospatiales ;
- les attributs et les entités (information sur le contenu du lot de données) ;
- la distribution ;
- référence metadata (information sur l'auteur de la metadata).

Les éléments du CSDGM peuvent être créés sous SGML (qui est son format d'échange conformément à la DTD développée par le FGDC) en utilisant un simple éditeur de texte, mais pour éviter les risques d'erreur, il est préférable d'utiliser des outils tels que XtME (Xt Metadata Editor) qui fonctionne sous Unix. Il existe évidemment d'autres outils :

- ARC/INFO GIS Metadata Generator AMLs
- ASCII template
- ASCII template (autre version)
- blmdoc : pour ARC/INFO version 7

- CORPSMET: outil de création de metadata de l'Armée US
(Corps des Ingénieurs)
- Document.aml : ARC/INFO
- Document aml : version de l' USGS pour ARC/INFO 7.0.4.
- Geolineus 3.0: organise les données, visualise la lignée et crée des modèles de flux de données pour ARC/INFO
- Informix SQL template
- mp : Compilateur pour des metadata formalisées (USGS)
- NOAA's FGDC Metadata Toolkit: logiciel
- Word Perfect template

➤ **Standards de metadata dérivés du CSDGM :**

a)Le DENVER CORE :

C'est un standard développé à l'Université de DENVER (USA).

Les champs ou éléments suggérés par le DENVER CORE SONT :

- Thème-Mots clés;
- Position-Mots clés;
- Coordonnées extrêmes;
- Résumé;
- But;
- Temps-période-contenu;
- références;
- Données géospatiales-Présentation-Forme;
- Créateur;
- Titre;
- Langage;
- Description des ressources.

Tous ces champs se retrouvent dans le CSDGM, sauf l'élément '*Langage*'.

Il est à noter que ce standard de metadata est quasiment inconnu. Il n'est cité que par ceux qui critiquent le CSDGM pour le temps que ce dernier nécessite à sa mise en place et sa relative complexité (description trop technique).

b)The Australia New Zealand Land Information Council(ANZLIC) :

Le développement de ce standard rentre dans le cadre de la mise en place d'un plan stratégique (ANZLIC's Strategic Plan pour 1994/1997)d'implantation d'un système d'information géographique global pour ces contrées.

Dés le mois de Décembre 1995,une ébauche de recommandations pour le standard de metadata *ANZLIC* fut remise à 180 représentants d'organismes potentiellement utilisateurs ou en relation avec ce type d'information. Les auteurs de ces recommandations se sont inspirés du CSDGM, tout en reconnaissant être moins ambitieux pour leur standard, dont ils limitent le nombre d'éléments au strict minimum. Le groupe de travail chargé de la définition des champs du standard ANZLIC mit 18 mois de consultations pour proposer un condensé de 220 éléments du CSDGM :

| <u>Catégorie</u> | <u>Elément</u> |
|--------------------|----------------------------------|
| Dataset..... | -Titre |
| | -Conservateur |
| | -Juridiction |
| Description | -Résumé |
| | -Mots de recherche |
| | -Nom de la position géographique |
| | Ou |
| | -Coordonnées géographiques |
| Data Currency..... | - date de début |
| | - date finale |

Dataset Status.....-Progression

- Maintenance et fréquence de mise à jour

Access.....- Format de stockage de données

- Type de format valable
- Contraintes d'accès

Data Quality.....-Lignée

- Position Actuelle
- Attribut Actuel
- Consistance logique

Contact Information.....-Point (ou personne) de contact de l'organisation

- Point(ou personne)de contact de la position
- Mail Address 1 (adresse postale)
- Mail Address 2 (extension de l'adresse postale)
- Place ou Localité
- Etat ou Localité 2
- Pays
- Code postal
- Téléphone
- Fax
- Adresse électronique

Metadata Date.....-Date de création de la metadata

Additional Metadata.....-Metadata additionnelle (références d'autres répertoires ou systèmes contenant davantage d'informations sur les données traitées)

⇒ Les catégories d'éléments *Dataset* et *description* fournissent essentiellement de l'information sur le contenu des données décrites, de

l'agence responsable de leur collecte et organisation et du terrain géographique couvert.

- ⇒ Les catégories *Data Currency* et *Data Status* établissent le cadre temporel des données décrites.
- ⇒ La catégorie *Access* informe l'utilisateur sur les formats des données et les éventuels formats d'échange.
- ⇒ L'inclusion de la catégorie *Data Quality*, constituée des éléments *lineage*, *positional accuracy*, *attribute accuracy*, *logical consistency* (lignée, précision de position, précision d'attribut, consistance logique), a entraîné un grand débat chez la communauté de l'information géographique dans la mesure où cela risque d'être source d'incompréhension de la part des utilisateurs non habitués à ce genre d'information.
- ⇒ La catégorie *Contact Information* fournit les détails d'adresse de l'emplacement des contacts (personne contact) au sein de l'organisation responsable de la distribution des données aux autres utilisateurs.
- ⇒ *Metadata Date* donne la date de création des éléments de metadata relatifs aux données décrites.
- ⇒ *Additional Metadata* fournit un lien vers éventuellement d'autres sources d'informations sur les données décrites.

3.2) Government Information Locator Service : (GILS)

Le GILS a été élaboré pour appuyer les politiques gouvernementales américaines en vertu desquelles les départements et organismes du gouvernement sont tenus de rendre l'information qu'ils détiennent accessible au public et de mettre au point des systèmes d'information utilisables dans un environnement de systèmes ouverts.

En décembre 1994, le *Département of Commerce* a approuvé le profil GILS à titre de norme fédérale de traitement de l'information (FIPS 192) à laquelle tous les départements devaient se conformer. La création et l'exploitation du GILS sont devenues obligatoires aux Etats-Unis lorsque la loi dite *Paper Reduction Act* de 1995 (article 3511) a été adoptée. Dans le contexte du projet

de société d'information mondiale du G7, le GILS a été proposé à titre de modèle de localisateur d'information mondiale. En février 1995, les participants à la conférence ministérielle du G7 ont approuvé un projet pour la gestion de l'environnement et des ressources naturelles, qui comprendra un localisateur d'information mondiale.

Profil GILS :

Le profil est un guide à l'intention des préposés à la mise en application, qui augmente les probabilités que les systèmes GILS élaborés par divers préposés à la mise en application et fournisseurs soient interconnectables et interopérables. Le profil traite de l'intégration des systèmes et des données devant être échangées aux fins du GILS. Par exemple, il précise que la norme d'extraction de l'information ANSI/NISO (Z39.50) est la norme d'échange de l'information. Toutefois, le GILS ne précise pas l'apparence de l'interface d'utilisateur ni la structure interne de la base de données renfermant les enregistrements de localisation GILS. La première version du profil d'application GILS a été approuvée en 1994, à titre de norme fédérale américaine de traitement de l'information (FIPS 192). La deuxième version, rédigée en 1996, tient compte de l'expérience des fournisseurs de services GILS ainsi que des commentaires du sous-groupe GILS du gouvernement canadien.

Enregistrements GILS :

Les enregistrements GILS peuvent être utilisés pour décrire différents genres de ressources en information. Toutefois, la priorité doit toujours être accordée aux genres de ressources en information suivantes :

a- Produits de diffusion de l'information :

Les organismes devraient créer des enregistrements de localisation GILS pour décrire les produits de diffusion de l'information, comme des livres, des

CD-ROM, des publications, des études, des rapports et des brevets, sans égard au support. Ces localisateurs (qu'il ne faut pas confondre avec le GILS lui-même) cataloguent ou décrivent les produits de diffusion de l'information. Par exemple, un enregistrement GILS pourrait exister pour un catalogue de publications ministérielles qui, lui, sert de localisateur pour ces publications.

b- Systèmes d'information automatisés :

Les enregistrements GILS doivent être créés pour décrire les systèmes d'information automatisés, surtout ceux auxquels le public a accès directement ou indirectement.

c- Ressources sur Internet:

Les enregistrements GILS doivent également servir à identifier et à décrire les ressources en information gouvernementale sur Internet, qu'il s'agisse de sites Web ou de documents particuliers. Les enregistrements GILS peuvent appuyer la recherche précise d'information gouvernementale sur Internet et aider les utilisateurs à savoir si les renseignements repérés sont à jour, exacts et authentiques.

GILS ET USMARC :

Le profil d'application GILS fournit des renvois entre les éléments de base du GILS et le format de catalogage lisible par machine (USMARC).

LISTE DES ÉLÉMENTS DE BASE DU GILS :

Tous les éléments sont facultatifs et non répétitifs par défaut, sauf mentions contraires :

TITRE (Obligatoire)

CRÉATEUR (Obligatoire, répétitif)

AUTEUR (répétitif)

DATE DE PUBLICATION(Obligatoire pour les publications ou les ressources possédant des dates de création ou de mise à jour distinctes)

DATE DE PUBLICATION -STRUCTURÉE

LIEU DE PUBLICATION
LANGUE DE LA RESSOURCE(Obligatoire, s'il y a lieu, répétitif)
RÉSUMÉ
INDEX IDÉOLOGIQUE NORMALISÉ (répétitif)
THÉSAURUS IDÉOLOGIQUE
TERMES NORMALISÉS
TERME NORMALISÉ (Répétitif)
TERMES NON NORMALISÉS
DOMAINE SPATIAL
COORDONNÉES DE DÉLIMITATION
THÉSAURUS DE MOTS CLÉS DE LIEU
MOT CLÉ DU LIEU(répétitif)
DURÉE (répétitif)
DATE DU DÉBUT
DATE DE LA FIN
DISPONIBILITÉ (Obligatoire, répétitif)
SUPPORT (Obligatoire)
DISTRIBUTEUR (Obligatoire)
DESCRIPTION DE LA RESSOURCE (répétitif)
TRAITEMENT DE LA COMMANDE (Obligatoire)
INFORMATION SUR LA COMMANDE (Obligatoire)
COÛT
INFORMATION SUR LES COÛTS
PRÉ REQUIS TECHNIQUES
DURÉE DE DISPONIBILITÉ(répétitif)
DATE DU DÉBUT
DATE DE LA FIN
LIEN DISPONIBLE (répétitif)
TYPE DE LIEN
LIEN
SOURCES DE DONNÉES
MÉTHODOLOGIE

CONTRAINTES D'ACCÈS
CONTRAINTES D'ACCÈS GÉNÉRALES
CONTRÔLE DE CRÉATION OU DE DISSÉMINATION
CONTRÔLE DE LA COTE DE SÉCURITÉ
CONTRAINTES D'UTILISATION
POINT DE CONTACT
INFORMATION SUPPLÉMENTAIRE
BUT
PROGRAMME DE L'ORGANISME
RENGVOI (répétitif)
RAPPORT DE RENVOI (répétitif)
LIEN DE RENVOI (répétitif)
TYPE DE LIEN
LIEN (répétitif)
NUMÉRO DE CALENDRIER
IDENTIFICATEUR DE CONTRÔLE (Obligatoire)
IDENTIFICATEUR DE CONTRÔLE INITIAL
SOURCE DE L'ENREGISTREMENT (Obligatoire)
LANGUE DE L'ENREGISTREMENT (Obligatoire)
DATE DE RÉVISION DE L'ENREGISTREMENT

➤ **Application du GILS dans l'administration canadienne :**

En novembre 1995, le Conseil du Trésor du Canada a créé le sous-groupe GILS SGG) du Groupe de travail sur les normes des documents électroniques auquel il a confié la responsabilité d'évaluer la possibilité de faire du GILS une Norme du Conseil du Trésor sur la Technologie de l'Information (NCTTI). Après avoir comparé le GILS à un certain nombre de standards de metadata servant à décrire l'information administrative et les ressources en information en général, le groupe a conclu que le GILS satisfaisait aux exigences des ministères fédéraux qui sont tenus d'identifier clairement leurs ressources en information et d'offrir un bon accès aux utilisateurs. Le Sous-Groupe GILS s'est réuni

périodiquement en 1996 pour passer en revue l'actuel profil d'application GILS, formuler des suggestions pour l'adaptation de ce dernier, rédiger les Lignes directrices relatives au GILS canadien, planifier un projet pilote sur le GILS et préparer l'ébauche de la NCTTI sur le GILS.

3.3)TEXT ENCODING INITIATIVE:

La TEI - Text Encoding Initiative - permet l'échange des données textuelles et d'autres types de données comme les images et les sons. Elle tire son origine d'une part de l'anarchie qui règne dans la communauté scientifique en matière de format, et d'autre part du nombre croissant de traitements que les chercheurs opèrent sur les textes sous forme électronique.

Les Recommandations de la TEI fournissent le moyen de rendre explicites certaines caractéristiques d'un texte, de façon à faciliter le traitement de ce texte par des programmes informatiques pouvant s'exécuter sur des plates-formes différentes. Cette tâche est appelée balisage ou codage.

Les Recommandations s'appuient sur le langage SGML (Standard Generalized Markup Language) pour définir leurs règles de codage. Tous les outils SGML généralistes sont capables de traiter des textes conformes à la TEI.

Les Recommandations de la TEI peuvent être appliquées aussi bien pour créer de nouvelles informations que pour échanger des informations existantes.

La TEI est soutenue par *l'Association for Computers and the Humanities*, *l'Association for Computational Linguistics* et *l'Association for Literary and Linguistic Computing*. Le projet a été en partie financé par *le National Endowment for the Humanities américain*, la DG XIII de la CEE, la fondation Andrew W. Mellon et le *Social Science and Humanities Research Council* du Canada.

Les Recommandations ont été publiées en mai 1994, après six ans de travaux auxquels ont participé des chercheurs de toute nationalité et de toute discipline. Au début de cette entreprise, les objectifs généraux de la TEI ont été définis par la résolution finale de la conférence préparatoire tenue au Vassar Collège de New York en novembre 1987. Cette résolution connue sous le nom de «

Principes de Poughkeepsie » fut peu à peu précisée à travers une série de documents de travail. D'après ces documents les Recommandations devaient :

- ✓ être suffisamment précises pour représenter les propriétés des textes intéressants pour les chercheurs;
- ✓ être simples, claires et concrètes;
- ✓ être utilisables facilement par les chercheurs et ne pas nécessiter l'utilisation de logiciels spécifiques;
- ✓ permettre une définition rigoureuse des textes en vue de traitements efficaces;
- ✓ être modifiables par l'utilisateur;
- ✓ respecter les normes en vigueur ou sur le point d'être adoptées.

Le monde de la recherche est large et divers. Pour que ces recommandations aient une large audience, il était important de s'assurer que :

- 1- les descriptions des caractéristiques fondamentales d'un texte puissent être facilement échangées;
- 2- les descriptions spécialisées puissent être facilement ajoutées (ou supprimées) d'un texte;
- 3- la même caractéristique puisse être encodée, en parallèle, de plusieurs façons;
- 4- la richesse du balisage puisse être déterminée par l'utilisateur de la façon la plus simple possible;
- 5- une documentation relative au texte et à la façon dont il a été codé soit fournie.

La TEI prévoit des mécanismes pour paramétrer le nom des balises et donc, si on le souhaite, utiliser des balises de son choix (par exemple des balises en français).

L'En-tête TEI :

Tout texte conforme à la TEI comporte :

- 1- une en-tête TEI (balisé comme un élément <teiHeader>
- 2- la transcription du texte lui-même (balisé comme un élément <text>).

L'en-tête TEI contient des informations analogues à celles que l'on trouve sur la page de titre d'un texte imprimé. Elle contient quatre parties :

- 1- une description bibliographique du texte électronique;
- 2- une description de la manière dont il a été codé;
- 3- une description non-bibliographique du texte (le « profil » du texte);
- 4- un historique de révision.

Par conséquent, une entête TEI peut être assimilée à une metadata. Sa fonction est d'assurer que l'information nécessaire pour créer une notice catalographique soit facilement repérable. Les recommandations 'TEI' expliquent que les en-têtes TEI n'ont pas le même rôle que les notices MARC. Alors que les enregistrements MARC sont fondamentalement une version électronique d'une fiche de catalogue qui fait référence à un objet physique, les en-têtes TEI fournissent non seulement toute l'information catalographique, mais en plus toute l'information non bibliographique, déterminante dans le traitement du texte électronique.

L'en-tête TEI, avec ses zones descriptives, peut être facilement repérée et analysée par machine et assure un lien direct avec le texte décrit. Cependant les recommandations relatives à ces en-têtes n'ont pas le statut de normes, ce qui limite la généralisation de ces dernières.

3.4) Internet Anonymous Ftp Archives Templates : (IAFA)

IAFA Templates ont été créés pour décrire le contenu et les services d'archives Ftp (File Transfert Protocol), et ainsi, en faciliter l'accès.

A l'origine, IAFA Templates fut développé pour être utilisé avec le protocole Whois++, par le groupe de travail IAFA de l'IETF (Internet Engineering Task Force) et l'ébauche des directives de IAFA fut publiée en juillet 1995.

Des compagnies privées (BUNYIP, NEXOR,...) s'y sont investies et ont développé des outils de navigation et des répertoires de services. Le but visé par la conception d'IAFA Templates était de fournir un moyen que pourraient utiliser les administrateurs des archives Ftp pour décrire les diverses ressources qu'ils détiennent. Ces dernières peuvent être sous différents formats : images, textes, sons, services tels des listes de diffusion ou des bases de données, aussi bien que des archives de listes de diffusion ou de groupes Usenet et des logiciels. L'intention à l'origine était que les administrateurs de sites Ftp soient responsables de l'implantation de IAFA Templates pour chaque fichier contenu dans leurs archives. Ainsi, l'information serait repérable par tout individu, visitant le site. Il existe actuellement plusieurs applications utilisant IAFA Templates. Le système de recherche ALIWEB développé par NEXOR, est le premier outil de recherche qui 'pointe' les archives Ftp décrites avec IAFA Templates. Un outil logiciel de recherche ROADS (Ressource Organisation And Discovery in Subject-based services) est venu utiliser IAFA Templates pour la description de ressources. Les auteurs ont incorporé dans la version de mai 1996 de ce logiciel, le protocole Whois++.

IAFA Templates est un standard de metadata d'implantation facile, avec des éléments de description très simples. Il y a différents types de gabarits définis dans les directives de l'IETF pour décrire les diverses ressources accessibles par le réseau :

- ✓ Documents;
- ✓ Lot de données ;
- ✓ Archives de listes de diffusion ;
- ✓ Archives de groupes de discussion ;
- ✓ Packages de logiciels ;
- ✓ Images.

D'autres types de gabarits ont été conçus dans le contexte des archives Ftp pour fournir de l'information sur des sites Ftp particuliers :

- ✓ Information sur des configurations de sites,
- ✓ Configuration logique des 'archives',
- ✓ Services (catalogue en ligne, informations sur serveurs),
- ✓ Miroir (détails des sites 'miroirs' incluant des informations sur la fréquence des mises à jour du site 'source').

Définition des gabarits :

a) Informations sur le Site :

Champs de ce gabarit :

- Nom du gabarit;
- Nom de l'hôte;
- Alias de l'hôte;
- Admin-(Utilisateur);
- Autre-(Organisation) ;
- Sponsor-(Organisation);
- Ville;
- Etat;
- Pays;
- Latitude-Longitude ;
- Fuseau horaire;
- Fréquence de mise à jour;
- Temps d'accès;
- Règles d'accès;
- Description;
- Mots clés ;
- Notes pour ce gabarit.

b) Information logique d'archive :

- Type de gabarit;
- Admin-(Utilisateur);
- Nom de l'hôte;
- Alias de l'hôte;

- Autre-(Organisation);
- Sponsor-(Organisation);
- Règles d'accès;
- Description;
- Fréquences de mises à jour;
- Mots clés.

c)Fichiers Automatiques de mise à jour de l'Information :

Un certain nombre de ces fichiers doit exister au niveau de l'archive.

d) Information contenue :

Pour cette catégorie, l'information dont il est question, est celle contenue dans l'archive, plutôt que celle disponible sur le serveur Ftp Anonymus.

e) Information Utilisateur :

Le gabarit, qui décrit cette information peut être stocké dans une place assurant un seul usage. Ce type de gabarit est désigné: 'USER'.

f)Information d'organisation :

D'une façon similaire que la précédente, le gabarit 'organisation' fournit une information commune qui doit orienter vers d'autres gabarits qui décrivent la source centrale d'information.

g) Information de service :

Les champs pour ce gabarit sont :

- Type de gabarit;
- Titre ;
- URI ;
- Admin-(Utilisateur) ;
- Autre-(Organisation) ;
- Sponsor-(Organisation);
- Description;
- Authentification;
- Enregistrement;
- Règles de chargement;
- Règles d'accès;
- Temps d'accès;

- Mots clés;
- Sujet-Descripteurs-Schéma;
- Sujet-Descripteurs;

h)Autres gabarits :

Les gabarits relatifs aux Documents, aux lots de données, aux archives de listes de diffusion, aux archives USENET, aux packages de logiciels, aux images et autres objets contiennent les mêmes champs mais des valeurs différentes pour le champs « type de gabarit » :

| <u>Type d'objet:</u> | <u>Type de gabarit:</u> |
|-----------------------------------|--------------------------------|
| Document: | DOCUMENT |
| Image: | IMAGE |
| Package de logiciels: | SOFTWARE |
| Archive des listes de diffusion: | ARCHIVE de MAILS |
| Archive de groupes de discussion: | USENET |
| Fichier son: | SON |
| Fichier Vidéo: | VIDEO |
| Fichier des FAQ: | FAQ |

Donc nous aurons pour chaque gabarit les champs suivants :

Type de gabarit: (voir liste ci-dessus)

| <u>Catégorie:</u> | <u>Type d'objet :</u> |
|--------------------------|---|
| Titre: | titre complet de l'objet |
| URI-V: | description de l'accès à l'objet. |
| Titre court : | partie du titre (s'il est long) |
| Auteur: | description/contact information sur l'auteur/créateur de l'objet. |
| Admin-(Utilisateur): | description/contact information sur l'administrateur de l'objet. |
| Source: | information sur la source de l'objet. |
| Conditions: | conditions d'utilisation de l'objet. |
| Description: | description (résumé) de l'objet. |

| | |
|------------------------------|---|
| Bibliographie: | bibliographie |
| Citation: | citation de l'objet quand il est utilisé dans un autre travail. |
| Status-Publication: | statuts de la publication courante (draft, pré-publication, etc.). |
| Editeur: | éditeur(information) |
| Copyright: | droit de copyright . |
| Date de création: | date de création |
| Discussions: | description (si possible) des forums de discussions relatifs à cet objet. |
| Mots clés: | mots clés |
| Version-v: | version désignée pour cet objet. |
| Format-v: | formats dans lequel cet objet est donné. |
| Taille-v: | taille(en octets). |
| Language-v: | langue dans lequel, l'objet est écrit si c'est un document, ou langage de programmation si c'est un logiciel. |
| Caractères-v: | jeu de caractères(ASCII ou "ISO Latin-1"). |
| ISBN-v: | <i>International Standard Book Number</i> de l'objet. |
| ISSN-v: | <i>International Standard Serial Number</i> de l'objet. |
| Dernière date de révision-v: | dernière date de révision. |
| Sujet-Descripteurs-Schéma-v: | nom de la classification utilisée. |
| Sujet-Descripteurs-v: | classification pour la ressource. |

3.5)Summary Object Interchange Format : (SOIF)

Le standard **SOIF** (Summary Object Interchange Format) a été défini dans le cadre du Projet HARVEST, en Janvier 1994. Il est dérivé du standard **IAFA Templates** et du format bibliographique **BIBTEX**.

Les enregistrements SOIF ont été conçus pour être générés par les outils du Projet HARVEST et permettre la recherche d'information auprès des

brokers(fournisseurs d'information) HARVEST. Depuis le mois de Mars 1996, la compagnie Netscape Communications a annoncé qu'elle allait utiliser les enregistrements SOIF, dans ses produits.

Les gabarits SOIF peuvent être générées à la main par les auteurs ou ceux qui maintiennent les archives, bien que la majorité de ceux qui sont utilisés actuellement, soient générés automatiquement par des robots. SOIF est réellement un format d'enregistrement interne au projet HARVEST. Il a été conçu dans un but très spécifique (Indexation sommaire de ressources)mais le format de base peut-être étendu à d'autres besoins.

Les éléments descriptifs de base d'un gabarit SOIF sont :

- Résumé;
- Auteur;
- Description;
- Mots clés;
- Titre.

3.6)Metadata IMS : (Instructional Management System)

C'est un standard de metadata développé à l'université de l'Etat de Californie, en 1997, pour les besoins des personnes et organismes qui travaillent pour, ou entretiennent des relations avec le monde de l'éducation.

Le standard IMS a un système de contenants (objets) types avec des ensembles metadata bien définis. Les membres de chaque ensemble sont tirés d'un dictionnaire de champ de metadata commun. L'ensemble minimal de metadata, dont tout contenant IMS doit s'accommoder, est le core (IMS core) de 12 champs. Le module 'contenant' comporte 25 champs,et est d'un intérêt particulier pour la création des ressources éducatives . Des outils de création de metadata ont été développés, pour faciliter la création des modules.

Le standard de metadata IMS comprend 35 champs tirés du Dictionnaire des Champs IMS :

- Résumé;
- Auteur;
- Catalogue ID;
- Concepts;
- Type de contenant;
- Crédits;
- Date d'expiration;
- Forme;
- Format;
- Guide;
- Niveau d'interactivité;
- Mots clés;
- Language;
- Niveau d'apprentissage;
- Localisation;
- Version de la metadata;
- Objectifs;
- Pedagogie;
- Plate-forme;
- Pré-requis;
- Presentation;
- Code des prix;
- Relation;
- Rôle;
- Taille;
- Source;
- Distributeur;
- Structure;
- Sujet;
- Titre;
- Droits d'utilisation;

- Support de l'utilisateur;
- Temps d'utilisation;
- Date de la version;
- Version.

La plupart des champs de Metadata sont structurés, permettant des termes multiples, et des hiérarchies. Ceux-ci ont été définis en utilisant le Format de Définition de Ressource du W3C (RDF).

Les types de 'contenant' de metadata IMS :

Le système de metadata IMS possède différents types de contenants (objets) et définit un minimum pour chacun.

Tous les types incorporent un ensemble de cores de metadata, *l'IMS core*. Ce dernier comprend quatre (4) types de contenants:

- L'article
- Le module
- Le profil
- L'outil

Les metadata IMS sont dérivées du Dublin Core. Nous y retrouvons d'ailleurs tous les éléments du Dublin Core, en plus d'autres éléments plus spécifiques.

3.7) Meta Content Framework : (MCF)

Le format d'échange MCF(Meta Content Framework) a été développé dans les laboratoires de la campagne APPLE.

Le but de MCF est de fournir un langage de représentation du contenu d'un large éventail de ressources d'information. Sa particularité réside dans le fait que les metadata ne sont pas placées dans des balises HTML ou SGML, mais elles sont extraites automatiquement et représentées sous un format MCF.

Pour comprendre MCF, il est nécessaire de connaître :

- 1- les objets, les catégories, et les propriétés qui forment un bloc conceptuel dans MCF;
- 2- la syntaxe XML avec laquelle MCF peut-être stockée;
- 3- le modèle mathématique DLG (Directed Linked Graph), qui peut-être employé par des programmeurs informatiques pour développer efficacement des mise-en-oeuvres de MCF.

Propriétés, objets et Catégories :

MCF fournit l'information sur l'information en attachant des propriétés aux objets. Depuis toujours, dans un ordinateur, les objets sont structurés en mémoire, mais ils sont normalement employés pour représenter des choses telles que des pages Web, des entreprises, des peuples, des pays et des évènements. Par exemple, une page Web pourrait avoir une propriété qui donne sa taille, une autre qui donne son URL, et une autre qui identifie la personne qui l'entretient. Les propriétés sont utilisées pour donner de l'information sur ces objets. Nous distinguons entre des types de propriété et propriétés. Un exemple d'un type de propriété est « sizeInBytes », qui pourrait s'appliquer à toute page Web. Quand il est appliqué à un objet particulier, il devient une propriété, un exemple du type de propriété, qui a une valeur: par exemple la page Web :<http://www.textuality.com/> actuellement, a un « sizeInBytes » de 5,676 (propriété dont la valeur est : 5,676).

Il faut signaler que les types de propriété sont aussi des objets; cela signifie qu'ils peuvent eux aussi avoir des propriétés. Par exemple, le type de propriété qui donne la taille d'une page Web pourrait être nommée « sizeOfPage », il pourrait avoir une propriété qui s'appliquera aux pages Web, une deuxième fixera sa valeur (nombre), une troisième donnera le nombre en octets, et une quatrième fournira un texte explicatif.

Un objet peut avoir plus d'un type; par exemple, un objet représentant une personne pourrait être du type Docteur et du type Chercheur. La clé finale du concept est la catégorie - dans l'exemple précédent, "Docteur" et "Chercheur" sont des catégories.

Utilisation de la syntaxe XML :

Les objets, propriétés, et catégories de MCF, nécessitent une syntaxe pour les stocker, avec une facilité de traitement manuelle ou par ordinateur. Cette syntaxe est basée sur XML (Extensible Markup Language). Dans XML, les documents contiennent des éléments, lesquels ont des types et sont l'un ou l'autre vides ou sont délimités par des start-tags et des end-tags (début de balise et fin de balise), et ont des attributs avec des noms et des valeurs, par exemple:

<p secret ="faux">. Cette phrase est dans le contenu d'un élément dont le type est 'p'; le contenu est placé entre le start-tag et le end-tag (début de balise et fin de balise). Le paragraphe a un attribut nommé "secret" dont la valeur est "faux".

Le Formalisme Direct Linked Graphic :

MCF est bâti sur un modèle mathématique. Mathématiquement, MCF est un graphique relié direct (DLG). Les objets sont représentés par des noeuds; les propriétés sont représentées par des arcs, qui sont des flèches reliant deux noeuds; les arcs ont des étiquettes, qui sont des types de propriété.

Les types d'éléments de metadata pris en compte dans MCF sont :Description, Auteur, Affiliation, Date de publication, Hyperliens, Langage, Sujets, Sites miroirs, Types de media(JPEG, MPEG, Poscript, Java Applet).

3.8- Metadata for Interechange of Files on Sequential Storage Media between File Storage Management Systems :(FSMS)

Ce standard, développé par *l'Association for Information and Image Management International (AIIM)*, précise le format et le contenu de metadata pour l'échange de fichiers sur des médias amovibles à stockage séquentiel. Ces derniers comprennent les bandes optiques et magnétiques et autres médias .

Le standard de metadata de FSMS permet, entre autres :

- l'échange de médias à stockage séquentiel entre deux Systèmes de gestion de stockage de fichiers différents;
- l'échange positionnel de médias à stockage séquentiel,
- l'échange de médias à stockage séquentiel entre des environnements manuels et automatisés,
- l'échange de médias à stockage séquentiel entre des plates-formes matérielles différentes,
- l'échange de médias à stockage séquentiel entre des systèmes d'exploitation différents.

Le but de ce standard est de spécifier une voie de description de l'information supplémentaire ajoutée par un FSMS, avec d'autres informations, afin de permettre à un autre FSMS de lire ce que le premier a écrit et de reconstruire le fichier original comme il a été généré par l'application logicielle de départ.

Le standard de metadata du FSMS comporte une collection d'enregistrements. Chaque enregistrement consiste en une séquence de champs. L'enregistrement prend le nom du premier champ et les autres champs sont nommés champs d'informations ou simplement champs. Les champs d'information suivent aussitôt leurs champs de nom. Chaque enregistrement est soit nécessaire (obligatoire) soit sélectable (optionnel).

3.9)Climate and Environmental Data Retrieval and Archive:(CERA)

Le standard de metadata CERA a été conçu pour décrire des données climatologiques et écologiques, des modèles de données numériques, ainsi que des données d'observations. Il peut prendre en charge d'autres types de données spatiales (satellites).

Le standard de metadata CERA est basé sur le modèle de référence de IEEE. Le développement était fondamentalement guidé par l'intention de garder ce standard de metadata aussi simple que possible, mais aussi flexible que nécessaire pour prendre en charge les exigences des utilisateurs pour le système de base de données de climatologie et à incorporer des normes internationales de description de données comme DIF(Directory Interchange Format) et INFOCLIMA(World Climate Data Information Retrieval Service). INFOCLIMA est un thésaurus pour l'environnement et l'écologie réalisé par l'UNESCO.

L'usage de normes internationales dans les systèmes de base de données climatiques est nécessaire pour la communication future avec d'autres bases de données ou avec des systèmes d'information qui sont basés sur ces normes.

Il faut signaler que le standard de metadata CERA a été développé par l'Institut de Climatologie Allemand (DKRZ), et qu'il fait partie d'un ensemble d'outils et systèmes entrant dans la mise en place de la base de données climatiques et écologiques CERA.

3.10)The California Environmental Resources Evaluation System: (CERES)

Le standard de metadata *CERES* est constitué du même contenu de base que le *CSDGM* (standard de metadata du FGDC), mais le format varie dans les schémas de description.

Basé sur le langage SGML, sa DTD ne correspond pas en tout point non plus à la DTD du standard *CSDGM*, dont pourtant il a pris ses sources en partie.

Les auteurs du standard de metadata CERES se sont inspirés aussi du draft(ébauche) du *Content Standards for Non-Geospatial Metadata* du National Biological Survey (NBS) et ont adopté une bonne partie des recommandations de l'American Institut of Biological Sciences relatives à ce standard.

Les rares modifications apportées ont été faites dans le but d'améliorer l'utilité, spécialement dans le contexte de la diffusion et de la recherche des données électroniques. Nous en énumérons les plus importantes :

- Modifications des éléments 'Citation Information', 'Contact Information' et des sous-éléments qui font référence au contenu des metadata;
- L'élément 'Citation Information' est remplacé par 'Citation Definition';
- Attribuer les noms des entrées des éléments définis par le FGDC (CSDGM) pour inclure des données aussi bien que pour limiter le nombre de balises requises pour la représentation (tag HTML);
- Réduction du rôle de quelques noms d'entrée d'éléments du FGDC vers SGML;
- Quelques entrées du FGDC ont été implémentées comme attributs de spécifications d'éléments dans des groupements de données;
- La méthode de création de site est spécifiée, mais un unique identificateur de metadata est publié;
- Introduction des pointeurs NameLoc de Hytime pour simplifier la réutilisation et l'inclusion de metadata;
- La notation des références pour permettre des enregistrements de données cohérentes avec les spécifications du standard CSDGM;
- Utilisation de courtes références pour simplifier la création de documents de metadata individuels;
- Définition d'éléments additionnels non spécifiés dans le CSDGM, et suggestion d'éléments dérivés du Draft Content Standard for Non-Geospatial Metadata;
- Réarrangement de quelques entrées type FGDC pour implémentation par compilateur SGML;
- Enlèvement de quelques entrées types FGDC (exemple : *Cloud_Cover*).

3.11) Multiple Resolution Landscape Characteristics : (MRLC)

Le MRLC Consortium a cherché à développer une approche unifiée sur les metadata, qui est conséquente des exigences de metadata de chacun des

programmes participants, et conforme au standard de metadata développé par le *Federal Geographic Data Committee*. Le projet GAP(Gap Analysis Programm)a permis de développer un standard de contenu de metadata qui a servi de base pour les metadata du MRLC(Organisme rattaché au National Biologic Survey, lui-même rattaché au Ministère de l'intérieur US). Il s'en est suivi un document de base, rédigé le 5 septembre 1994, et intitulé : 'Standard de metadata pour Gap Analysis'....

Avec des centaines de chercheurs à travers les USA contribuant à Gap Analysis Programm, le rôle pour les metadata devient capital: Elles doivent fournir un moyen d'accès sélectif aux données. Par exemple, il peut y avoir un besoin de chercher les enregistrements de GAP pour l'information sur la propriété terrestre dans un emplacement géographique spécifique. Un type d'information disponible dans les metadata est la description de l'étendue spatiale d'une base de données. Une question basée sur la latitude et la longitude peut produire une liste de tous les produits cartographiques de la région décrite.

Le standard de metadata GAP(du MRLC) est dérivé du CSDGM(standard du FGDC) dont il maintient l'axe d'évolution afin de toujours assurer une certaine compatibilité entre eux.

Le format du standard GAP consiste en huit sections majeures de documentation contenant un ou plusieurs éléments de metadata. Chaque élément porte un nom. Le "Type" d'entrée (texte, entier, date, temps) et le champ de l'entrée sont aussi définis. Actuellement 276 éléments du CSDGM ne sont pas décrit dans le document de base qui définit le standard de GAP(standard de metadata MRLC).

Onze nouveaux éléments considérés nécessaires pour le projet GAP ont été ajoutés, sans qu'ils figurent dans le CSDGM.

Les données complètes, incluant des metadata et le dictionnaire des données, doivent être conçues pour le transfert numérique qui permettront aux metadata d'être distribuées isolément, cependant la base de données réelle doit toujours contenir des metadata.

3.12) Content Standard for Non-Geospatial Metadata :

Ce standard de metadata a été développé par le National Biological Survey (NBS) pour le compte du National Biological Information Infrastructure (NBII), pour décrire l'information non géospatiale. Le NBS a utilisé l'approche suivante, dans la consultation avec ses partenaires, pour développer un standard de metadata :

- Le but est d'avoir un standard de metadata NBII, qui est essentiellement dérivé du CSDGM/USMARC, tout en étant le plus étendu possible;
- NBS a sollicité L'Institut Américain de Sciences Biologiques (AIBS) pour organiser un atelier des experts nationaux dans les sciences biologiques pour étudier et recommander des modifications à son projet de standard de metadata;
- NBS a procédé à une revue interne de la version beta (draft recommandé par l'atelier des experts) afin de déterminer l'utilité, la faisabilité et éventuellement apporter des modifications, si nécessaire;
- NBS a sollicité L'Académie Nationale des Sciences/ Conseil National de la Recherche pour un avis sur ce projet de standard de metadata.
- NBS a repassé en revue le standard renvoyé par l'Académie Nationale des Sciences pour d'autres modifications possibles. Le standard final a été présentée au Federal Geographic Data Committee pour étude et adoption formelle comme standard FGDC.

Cette version finale est basée sur :

- Les définitions non spatiales par opposition aux données biologiques géospatiales;
- L'utilisation du CSDGM pour les données géospatiales;
- L'utilisation de USMARC pour cataloguer la portion des données non-spatiales.

3.13) Inter-university Consortium for Political and Social Research : (ICPSR)

Etabli en 1962, le consortium inter-universitaire pour la recherche politique et sociale (ICPSR) est une organisation du type associative, permettant d'accéder aux plus grandes archives du monde, dans le domaine des sciences sociales.

ICPSR fournit des équipements et des services pour la communauté internationale de ses membres.

Les données d'exploitation couvrent un large intervalle de disciplines, telles les sciences politiques, la sociologie, la démographie, les sciences économiques, l'histoire, l'éducation, la gérontologie, le droit, et la santé publique.

ICPSR encourage les sociologues dans tous les domaines à contribuer à l'enrichissement de ses bases de données et à utiliser ses ressources de données.

ICPSR inclut parmi ses membres plus de 325 universités en Amérique du Nord et plusieurs centaines d'établissements répartis en Europe, en Australie, en Asie et en Amérique Latine. Le siège social est situé dans l'Institut de Recherche Sociale de l'Université de l'Etat du Michigan (USA).

ICPSR ne dispose pas à proprement parler d'un standard de metadata, mais uniquement d'une sorte de formulaire, que tout dépositaire d'informations dans leurs bases de données ou d'archives, se doit de remplir pour décrire son document ou ce qui le constitue. Ces formulaires appelées : *Data Deposit Form*, exige de l'auteur ou du producteur de données de fournir des informations sur les caractéristiques techniques et référentielles d'une collection de données. En complétant ce formulaire de 24 champs (simple description de texte à remplir manuellement), l'auteur ou le producteur assure que la collection de données sera exactement et entièrement décrite pour des analystes secondaires potentiels dans la communauté des sciences sociales.

Le *Data Deposit Form* permet au personnel d'ICPSR de préparer une description d'étude (souvent considérée comme une metadata) pour diverses publications d'ICPSR ainsi que pour ses bases de données en réseau. Cette information permet aussi la création de citations bibliographiques d'autorité pour la collection.

3.14) National Environmental Data Referral Service : (NEDRES)

Le National Environmental Data Referral Service, NEDRES, est un programme de la *National Oceanic and Atmospheric Administration (NOAA)*.

Les utilisateurs peuvent accéder à une large gamme d'informations écologiques par des répertoires NEDRES, offerts en réseau(en ligne). NEDRES est un réseau auquel coopèrent des organisations fédérales, étatiques, académiques et privées, pour améliorer l'accès aux données écologiques. La base de données NEDRES est un catalogue informatisé de données écologiques. Ces données concernent le rayonnement du soleil à travers l'atmosphère, sur les terres et les océans, La physique solaire et de l'atmosphère supérieure , l'océanographie, la climatologie, la météorologie, la pollution, les substances toxiques, la géophysique et la géologie, la géochimie.

Le répertoire *NEDRES* contient seulement les descriptions, et non pas les données réelles, et renvoie l'utilisateur à la source pour une information intégrale.

La base de données fournit des documents sur plusieurs types de description d'information écologique. Elle inclut:

- 1- les données des centres, programmes et organisations;
- 2- les fichiers de données non imprimés ;
- 3- les données des publications séries ;
- 4- les données publiées ;
- 5- les atlas ou les données publiées sous la forme graphique;

- 6- les publications contenant des compilations extensives, des analyses extensives ou des applications de données ;
- 7- les manuels, les guides de l'utilisateur, ou la documentation des lots de données ;
- 8- les catalogues de données, les inventaires, ou les bibliographies ;

Une recherche dans la base de données NEDRES fournit une description complète des sources de données disponibles qui satisfont les spécifications de recherche. Les informations obtenues décrivent les données dans le détail. Les champs de définition sont:

- AN Accession Number
- TI Title
- AB Abstract
- DC Data Collection
- DD Data Processing and Quality Control
- PE Period of Record (earliest date and latest date)
- LR Length of Record
- GE Geographic Place Name
- GC Geographic Codes (FIPS, Ecoregion, Water, Resource)
- GL Geographic Grid Locators
- PA Parameter Description
- DE Descriptors of the data
- CO Contact address to obtain the data
- AV Availability Characteristics
- PI Principal Investigators
- PR Program or Project
- PO Processing Organization
- PR Related Publications
- RR Related Records
- DT Date Entered and Date Reviewed
- CC Category Codes

Les chercheurs et scientifiques qui utilisent la base de données NEDRES sont de tous les horizons, incluant des institutions académiques, privées, associatives, et gouvernementales.

Le format du catalogue NEDRES est créé selon le standard CSDGM du FGDC, dont il est finalement une application particulière. D'ailleurs un manuel : 'Content Standard for Digital Geospatial Metadata Workbook (version 1.0)', est prévu à cet effet.

3.15) DUBLIN CORE :

3.15.1) Introduction :

En Mars 1995 s'est tenu un workshop sur les Metadata, parrainé par *Online Computer Library Center (OCLC)* et le *National Center for Supercomputing Applications (NCSA)*, rassemblant 52 chercheurs et professionnels des bibliothèques, de l'informatique, et des spécialités connexes, pour faire avancer l'état de l'art dans le développement des descriptions de ressources (ou metadata). Les buts de ce workshop incluaient une compréhension commune des besoins, des points forts, des défauts, et des solutions à proposer; et l'atteinte d'un consensus sur un ensemble d'éléments de metadata pour décrire des ressources d'informations. La complexité du problème des descriptions de ressources a fait limiter l'étendue des discussions. Sachant que la majorité des informations diffusées sont sous forme de documents, et que les enregistrements de metadata sont requis pour faciliter la recherche de ressources sur Internet, l'ensemble proposé d'éléments de metadata (Le Dublin Core) est destiné à décrire les caractéristiques essentielles de documents électroniques. D'autres enregistrements de metadata tel ceux décrivant des informations de coût ou d'archives n'ont pas été pris en considération. Il fut reconnu, néanmoins que ces éléments pourraient être inclus dans une future version du Dublin Core.

Le Dublin Core n'est pas destiné à supplanter les autres descriptions de ressources, mais plutôt à les compléter. Il y a actuellement deux types de descriptions de ressource pour des documents électroniques diffusés: les index automatiquement générés tels ceux employés par Lycos et WebCrawler; et le catalogage des enregistrements (USMARC). Générés automatiquement, les enregistrements contiennent souvent un minimum d'information pour être utiles, tandis que générés manuellement ces enregistrements sont trop coûteux à créer et à entretenir pour le grand nombre de documents électroniques actuellement disponibles sur Internet. Les enregistrements du Dublin Core sont destinés à arbitrer ces extrêmes, fournissant un enregistrement structuré simple et qui peut être amélioré.

Le travail du workshop de 1995 est une première étape qui devrait être suivi d'autres pour améliorer au fur et à mesure, la description des informations. Un comité, issu du premier workshop, a été formé pour suivre le travail par une série d'activités similaires pour faire évoluer le Dublin Core.

Depuis qu'Internet contient plus d'information complète que de simples résumés professionnels, les indexeurs et catalogueurs tentent de gérer cela en utilisant les méthodes et systèmes existants: il était évident qu'une alternative pour obtenir des metadata utilisables pour des ressources électroniques doit donner aux auteurs et aux fournisseurs d'information un moyen permettant de décrire les ressources eux-mêmes, sans formation intensive et spécifique préalable.

C'est pour atteindre ce but, que la tâche majeure du workshop sur les metadata, était d'identifier et de définir un ensemble simple d'éléments pour décrire des ressources électroniques diffusées(en réseau).

La discussion, lors du workshop, était davantage limitée aux éléments de metadata nécessaires pour la découverte de ce qui était appelé 'Document-Like Objets'(objets documents) ou DLOs.

DLOS n'étaient pas rigoureusement définis, mais étaient compris par l'exemple. Par exemple, une version électronique d'un article de journal ou un dictionnaire est un DLO, tandis qu'une collection de diapositives ne l'est pas. Bien sûr, le cœur du problème est que dans un environnement réseau, les DLOs peuvent être arbitrairement complexes parce qu'ils peuvent consister en un texte avec des images, des clips de vidéo, du son, ou des documents hypertextuels. Les participants au workshop n'ont pas essayé de limiter la complexité du DLOs, sauf de remarquer que le contenu intellectuel d'un DLO est principalement du texte, et que les metadata susceptible de décrire des DLOs ont une forte ressemblance avec celles qui décrivent des textes imprimés traditionnels.

Donc, finalement les participants devaient définir un ensemble d'éléments de metadata qui permettraient aux auteurs et fournisseurs d'information de décrire leur travail et à faciliter l'interopérabilité parmi des outils de recherche d'information.

La première version de la synthèse des travaux du workshop établissait un ensemble minimal de treize éléments de metadata, qui fut nommé : Ensemble d'éléments de core de metadata de Dublin (ou plus simplement : Dublin Core).

La syntaxe était délibérément laissée vague comme un détail de mise-en-oeuvre. La sémantique de ces éléments était assez claire pour être comprise par une large gamme d'utilisateurs.

Les discussions entre les participants au workshop ont permis de révéler plusieurs principes qui devaient guider davantage le développement de l'ensemble des éléments. Ces principes sont : propriété intrinsèque, extensibilité, indépendance de syntaxe, optionalité, répétabilité et modifiabilité.

Propriété intrinsèque:

Le Dublin Core a été dirigé, dès le départ pour décrire des propriétés intrinsèques de l'objet. Par exemple, l'élément "Sujet" est une donnée intrinsèque, tandis que des informations de transaction telle que coût et droit d'accès sont des données extrinsèques.

Extensibilité :

En plus de son emploi en traitant de la propriété intrinsèque des données, le mécanisme d'extension permettra l'inclusion de données intrinsèques pour des objets qui ne peuvent pas être décrits suffisamment par un petit ensemble d'éléments.

L'extensibilité est importante parce que les utilisateurs peuvent vouloir ajouter des descriptions supplémentaires à des champs, même ayant des buts spécifiques. En outre, la spécification du Dublin Core lui-même peut changer dans le temps, et le mécanisme d'extension permet des révisions.

Indépendance de Syntaxe :

Les constructions syntaxiques sont évitées parce qu'il est trop tôt de proposer des définitions formelles et parce que le Dublin Core est destiné à être employé tôt ou tard dans une large gamme de programmes d'application et de disciplines.

Optionalité :

Tous les éléments sont optionnels, pour deux raisons :

La première est que le Dublin Core peut tôt ou tard être appliqué aux objets pour lesquels certains éléments n'ont pas de signification : Qui est l'auteur d'une image de satellite?

La seconde est qu'il semble futile de donner des descriptions complexes lorsque les auteurs du contenu prévoient de fournir la matière descriptive.

Répétabilité :

Tous les éléments du Dublin Core sont répétables. Par exemple, plusieurs éléments 'auteur' seraient employés quand une ressource a plusieurs auteurs.

Modifiabilité :

Chaque élément dans le Dublin Core a une définition qui est sensée être évidente. Cependant, il est aussi nécessaire que les définitions des éléments satisfassent aux besoins de communautés différentes. Ce but est accompli en permettant à chaque élément d'être modifié par un qualificateur optionnel. Si aucun qualificateur n'est proposé, l'élément prend son sens commun.

Les qualificateurs seront typiquement dérivés de conventions connues dans la communauté des bibliothèques ou du domaine des connaissances propres à la ressource. Les qualificateurs sont importants parce qu'ils donnent au Dublin core, un mécanisme pour supprimer ou amoindrir l'écart entre des spécialistes et des utilisateurs quelconques.

Par exemple, les données dans l'élément 'sujet' consiste en mots ou expressions qui décrivent le contenu de l'objet. Cependant, un professionnel du catalogage peut vouloir se référer à la source de laquelle les termes de sujet sont pris. Dans un tel cas, l'élément peut être écrit ainsi :

Sujet(scheme=LCSH), indiquant que les termes de sujet sont pris de la liste:
Library of Congress Subject Headings.

3.15.2) Liste des éléments du Dublin Core :

La liste actuelle des éléments, ainsi que leur définition générale, a été établie en décembre 1996. Elle comporte 15 éléments, alors que le premier workshop avait défini 13 éléments seulement. Il demeure que certains d'entre eux sont encore à l'étape expérimentale et que leur interprétation peut changer d'une implémentation à une autre.

Description des éléments :

1. Titre :

Etiquette: titre

Le nom donné à la ressource par le créateur ou l' auteur.

2. Auteur ou Créateur :

Etiquette: créateur

La personne ou l'organisation principalement responsable de la création du contenu intellectuel de la ressource.

3. Sujet et mots-clef :

Etiquette: sujet

Le sujet 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.

4. Description :

Etiquette: description

Une description textuelle du contenu de la ressource, y compris un résumé, dans le cas d'objets tels que des documents, ou une description du contenu dans le cas de ressources visuelles.

5. Editeur :

Etiquette: editeur

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

6. Autre contributeur :

Etiquette: contributeur

Une personne ou une organisation, non mentionnée dans un élément créateur, qui a fait une contribution intellectuelle significative 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).

7. Date :

Etiquette: date

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 par la norme ISO8601. 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.

8. Type de ressource :

Etiquette: type

La catégorie de la ressource, telle qu'une page personnelle, un roman, un poème, un document de travail, un rapport technique, une dissertation ou un dictionnaire.

9. Format :

Etiquette: format

Le format de la ressource, utilisé pour identifier le logiciel et, éventuellement, le matériel qui peuvent être nécessaires pour afficher ou traiter la ressource.

10. Identificateur de la ressource :

Etiquette: identificateur

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.

11.Source :

Etiquette: source

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 roman 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.

12.Langage :

Etiquette: langage

Langage(s) du contenu intellectuel de la ressource. Si approprié, le contenu de ce champ devrait correspondre à la norme RFC 1766.

13.Relation :

Etiquette: relation

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.

14.Couverture :

Etiquette: couverture

Definit 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.

15. Gestion des droits :

Etiquette: droits

Un lien sur les droits de reproduction, les droits d'utilisation, ou renvoi à un service capable de fournir l'information sur les conditions d'accès à la ressource. Une définition formelle de cette étiquette est en cours.

Utilisateurs et développeurs doivent comprendre que l'utilisation de cet élément est pour l'instant expérimentale.

3.15.3) Les qualificatifs du Dublin Core :

Nous reprenons ici un ensemble de qualificatifs pour le Dublin Core, qui ont été proposés par REBECCA GUENTER (spécialiste d'USMARC à la Bibliothèque du Congrès US). Le document a été mis à jour à la suite de discussions qui ont eu lieu lors du cinquième workshop qui s'est tenu à Helsinki en octobre 1997. Il traite des qualificatifs "shéma" et "type". Un qualificatif "shéma" est employé pour interpréter la valeur du contenu et est généralement basé sur des normes externes. Un qualificatif "type" affine la définition de l'élément de données lui-même.

La présente liste de qualificatifs est une approche intermédiaire entre les 'Minimalistes' (partisans du principe de n'utiliser qu'un minimum de qualificatifs et de laisser une certaine liberté aux auteurs) et l'approche des 'Structuralistes', qui voudraient fixer toutes les règles au départ (définition de tous les qualificatifs possibles et obligatoires). Il faut remarquer que malgré cela, tous les qualificatifs restent encore optionnels.

Liste des qualificatifs du Dublin Core :

1. Titre :

Shéma: non nécessaire

Type:

- DC.TITRE

- DC.TITRE.ALTERNATIF (employé pour tout titre autre que le titre principal incluant le sous-titre, etc.)

2. L'auteur ou Créateur :

Shéma:

- LCNAF (Library of Congress Name Authority File)

TyPE:

- DC.CREATOR
- DC.CREATOR.Nom Personne
- DC.CREATOR.Nom Campagne
- DC.CREATOR.Nom Personne(inclut tout type d'adresse, ou email)
- DC.CREATOR.Nom Compagnie(inclut adresse)

3. Sujet et Mots-clés :

Shéma:

- (Non qualifié:Mots-clés pris par défaut)
- LCSH (Library 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. L'éditeur :

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_presente_forme
- DC.DATE.FORMAL_PUBLI

- DC.DATE.ACCESSIBLE
- DC.DATE.VALIDE (inclut la vérification)
- DC.DATE.ACQUISITION/Accession
- DC.DATE.ACCEPTEE
- DC.DATE.DATAGATHERIN(Collecte de Données)

8. Type de Ressource :

Shéma: non 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 Resource 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

- URL
- 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. La relation :

Shéma :

- texte libre est pris par défaut
- URL
- URN
- ISBN

Type:

- Créatif (e.g. traduction, annotation)
- Mécanique (copier, changement de format etc...)
- Version (edition, draft)
- Inclusion (collection, part)
- Référence (citation)

14. La couverture :

Shéma:

Sera déterminé par le Groupe de travail sur Coverage
(Couverture)

Type:

La liste suivante a été déterminée par le Groupe de travail sur Coverage :

- DC.COVERAGE.PERIODNAME
- DC.COVERAGE.PLACENAME
- DC.COVERAGE.T
- DC.COVERAGE.X
- DC.COVERAGE.Y
- DC.COVERAGE.Z
- DC.COVERAGE.POLYGON
- DC.COVERAGE.LINE
- DC.COVERAGE.3D

15. gestion des droits :

Shéma:

- Texte libre est pris par défaut
- URL
- URN

Type: Non nécessaire.

3.15.4) Syntaxe du Dublin Core :

Pour transcrire les éléments du Dublin Core en un document électronique, il est évident qu'il est nécessaire de disposer d'une syntaxe spécifique.

Actuellement, le web constitue l'outil stratégique d'Internet. La communauté du Dublin Core a commencé donc par préconiser l'utilisation du langage HTML (voir le site <http://www.oclc.org:5046/~weibel/html-meta.html>).

La convention proposée est la suivante:

```
<META NAME = "schema_identifieur.element_name.qualifieur" CONTENT =  
"string data">
```

ou :

NAME: Cet attribut précise un nom de propriété.

CONTENT: Cet attribut précise la valeur d'une propriété.

HEMA: Cet attribut désigne un schéma de description à employer pour interpréter la valeur de la propriété.

HTTP=: Cet attribut peut être employé à la place de l'attribut de nom.

L'élément META peut être employé pour décrire des propriétés d'un document (auteur, date d'expiration, une liste de mots-clés, etc.) et attribuer des valeurs à ces propriétés. Cette spécification ne définit pas un ensemble normatif de propriétés.

Exemples :

1-Titre:

```
<META NAME="DC.title" CONTENT="Standards de metadata">
```

```
<LINK REL=SCHEMA.dc HREF=
```

```
"http://purl.org/metadata/dublin_core_elements#title">
```

```
<META NAME="DC.title.subtitle" CONTENT="Dublin Core">
```

2-Auteur or Créateur :

```
<META NAME="DC.creator" CONTENT="Amerouali Y">
```

```
<META NAME="DC.creator.email" CONTENT="ameroual@enssib.fr">
```

3-Sujet et mots-clés :

```
<META NAME="DC.subject" CONTENT="Dublin Core, metadata, ressource électronique">
```

```
<META NAME="DC.subject" CONTENT="(scheme=LCSH) Catalogage de ressources électroniques">
```

```
<META NAME="DC.subject" CONTENT="(scheme=UDC) 518.118">
```

4-Description:

```
<META NAME="DC.description" CONTENT="Description textuelle du contenu de la ressource, incluant éventuellement un résumé.">
```


5-Editeur:

<META NAME="DC.publisher" CONTENT="NORDINFO">

6-Autres Contributeurs:

<META NAME="DC.contributors" CONTENT="Monia Lind">

7-Date:

<META NAME="DC.date" CONTENT=
"(scheme=ANSI.X3.30-1985) 19980220">

<META NAME="DC.date.current"
CONTENT="(scheme=IETF.RFC-822) Thu, 11 Mars 1998 21:12:34 +0100">

8-Type de ressource:

<META NAME="DC.type" CONTENT="Note de synthèse">

9-Format:

<META NAME="DC.format" CONTENT="(SCHEME=imt) text/html">

<LINK REL=SCHEMA.imt HREF=
"http://sunsite.auc.dk/RFC/rfc/rfc2046.html">

10-Identificateur de la ressource :

<META NAME="DC.identifieur"CONTENT="http://www.ncl.ac.uk/~napm1/
dublin_core/index.html">

<META NAME="DC.identifieur.ISBN" CONTENT="1-56884-452-2">

11-Source :

<META NAME="DC.source" CONTENT="Levine and Baroudi: Internet
secrets">

<META NAME="DC.source.ISBN" CONTENT="1-56884-452-2">

12-Language :

<META NAME="DC.language" CONTENT="(SCHEME=NISOZ39.53) FRE">

13-Relation :

<META NAME="DC.relation.IsDerivedFrom"
CONTENT="http://www.oclc.org:5046/oclc/research/conferences/
metadata2/">

14-Couverture :

<META NAME="DC.coverage.local" CONTENT="Scandinavie">

15-Gestion des droits :

<META NAME="DC.rights" CONTENT="Domaine public">

3.15.5)Structure de Description de Ressources : (RDF)

La Structure de Description de Ressources est un cadre pour les metadata; qui permet l'interopérabilité entre des applications qui échangent l'information compréhensible par machine sur le Web. Elle améliore les facilités de traitement automatisé des ressources d'internet. La RDF metadata peut être employée dans des applications diverses; par exemple: dans la recherche d'information en permettant de meilleures capacités aux outils de recherche; dans le catalogage pour décrire le contenu et les rapports avec les contenus disponibles dans un site Web particulier, une page Web, ou une bibliothèque numérique; par des agents logiciels intelligents, pour faciliter le partage et l'échange de connaissances.

Le groupe de travail qui suit les développements de la RDF est sous les auspices du W3C. La RDF a commencé comme une extension des éléments de description définis par le W3C pour PICS(Plateform for Internet Content Selection). Elle intervient maintenant aussi dans le langage XML(produit XML de Microsoft et XML/MCF de Netscape). Les orientations du Dublin Core discutées à WARWICK ont eu une certaine influence aussi sur les récents développements de la RDF(version du 16 Fevrier 1998). Il y a actuellement un groupe de réflexion du Dublin Core qui travaille sur la manière d'intégrer entièrement la syntaxe RDF dans le Dublin Core. Cela prépare sûrement à une future syntaxe sous XML pour le Dublin Core(et peut-être une liberté de définir soi-même ses propres balises).

3.15.6) Les projets utilisant le Dublin Core :

Projets Australiens :

1-DSTC :

Page d'accueil : <http://www.dstc.edu.au/RDU/>

Le DSTC participe dans le Groupe de travail du W3C sur la RDF. Le DSTC compte développer une spécification et des outils pour l'utilisateur pour créer des éléments de Dublin Core compatibles avec la RDF et fournir des interfaces de recherche.

2-AGCRC : (Australian Geodynamics Cooperative Research Center)

Page d'accueil : <http://www.agcrc.csiro.au/>

L'AGCRC, qui est une collaboration entre deux organismes publics de recherche et deux universités, utilise le Web comme un premier système de publication des résultats de ses recherches. Le projet utilise deux standards de metadata différent pour le texte et les données numériques, avec un lien entre eux.

3-Projet de Metadata de la Bibliothèque d'état de Queensland :

Page d'accueil: <http://www.slq.qld.gov.au/meta/overview.htm>

Le but du Projet est la prise en compte de metadata dans la page Web de la Bibliothèque D'état de Queensland. C'est aussi une tentative initiale pour établir des normes de déploiement de metadata dans les bibliothèques du Queensland.

4-Le projet PANDORA :

Page d'accueil: <http://www.nla.gov.au/politique/pandje97.html>

La Bibliothèque Nationale D'Australie (NLA) a engagé un projet de développement d'un système de gestion d'archives électroniques appelée PANDORA, pour fournir l'accès en ligne, à long terme, à de significatives publications Australiennes. Le projet intégrera un système de description des documents archives, basé sur le Dublin Core.

5-EdNA (Education Network Australia) :

Page d'accueil:<http://www.otfe.vic.gov.au/edna/dc5edna.htm>

EdNA est un projet collaboratif entre tous les territoires et Etats Australiens et tous les secteurs de l'éducation et de la formation. EdNA utilise un standard de metadata basé sur le Dublin Core.

6-Le secteur de l'environnement Australien :

Page d'accueil: <http://www.environment.gov/>

Le secteur de l'environnement en Australie utilise le Dublin Core pour les informations qu'il met sur le Web et dans son intranet. Le service d'information en réseau du secteur de l'environnement Australien propose plus de 8000 documents en ligne, répartis sur plusieurs bases de données.

Projet Canadien :

7-SearchBC: Vancouver Webpages

page d'accueil:<http://vancouver-webpages.com/VWBOT/searchBC.html>

Ce projet consiste en un robot pour parcourir le web, un moteur de recherche de base de données, et un générateur de scripts de metadata.

Projet Français :

8-MedExplore & Metadata :

Page d'accueil:<http://www.loria.fr/%Educlloy/PUB/Publi97/DC5/DC5.HTM>

L'objectif de MedExplore est de permettre à un utilisateur de naviguer grâce à un concept de graphes et à manipuler divers morceaux d'information de grandes bases de données internationales, des documents de sources locales et des informations tirées d'Internet.

Projets Allemands :

9-Metadaten-Projekt : (Projet de metadata)

Page d'accueil :<http://www2.sub.uni-goettingen.de>

Ce projet explore l'emploi des metadata d'un point de vue des bibliothèques. Il vise la transmission du savoir-faire accumulé dans la recherche de source

d'information en réseau dans le monde, vers les bibliothèques Allemandes. Il constitue une partie d'un plus grand projet impliquant plusieurs bibliothèques Allemandes.

10-SSG-FACHINFORMATION (SSG-FI) Mathematick :

Page d'accueil:<http://www.sub.uni-goettingen.de/ssgfi>

Les metadata sont générées pour l'évaluation de l'information relative aux mathématiques. Les sources incluent des serveurs Internet, des CD-ROMS, et des références de livres.

11-SSG-FACHINFORMATION (SSG-FI) Geowissenschaften

Page d'accueil:<http://www.sub.uni-goettingen.de/ssgfi/>

Ce service d'information est hébergé par le même serveur que le précédent, mais il est orienté vers les sciences de la terre. Le projet de metadata leur est commun.

12-Deutscher Bildungs - Serveur : (secteur de l'éducation)

Page d'accueil:<http://dbs.schule.de/indexe.html>

Ce site contient actuellement près de 2000 documents sur l'éducation.

13-Math-Net :

Page d'accueil:<http://elib.zib.de/math-net/>

Ce projet concerne les metadata créés par les auteurs eux-mêmes, pour des documents relatifs aux mathématiques.

14-Gestion de l'information électronique et Metadata en physique :

Page d'accueil:<http://www.physik.uni-oldenburg.de/EPS/>

EurophysNet/PhysDep/dep-links.html

Le but de ce projet est de développer un système distribué d'information électronique en physique. La première version de ce système est basée sur Harvest(modèle US).

15-Bibliothèque Electronique De visualisation :

Page d'accueil:<http://visinfo.zib.de/Bibliothèque/>

La Bibliothèque Electronique de visualisation (EVlib) est un service distribué de publications électroniques.

16-Bibliotheksservice-Zentrum(BSZ)Baden-Wuerttemberg (Sudwestdeutscher Bibliotheksverbund-SWB-VERBUND) : (Réseau de Bibliothèques dans le Sud-ouest Allemand)

Page d'accueil: http://www.swbv.uni-konstanz.de/wwwroot/s71800_d.html

Ce réseau de bibliothèque fonctionne avec une base de données bibliographiques pour la région, de 16000 000 de références, et incluant des documents électroniques avec l'accès en ligne.

Les Pays-Bas :

17-Koninklijke Bibliotheek :Bibliothèque Nationale des Pays-Bas

Page d'accueil: <http://www.konbib.nl:8000>

La Bibliothèque Nationale des Pays-Bas est en train de développer une nouvelle version de son service d'information sur le Web. C'est une volonté de changement avec des nouvelles caractéristiques de fonctionnement et l'incorporation des éléments de metadata du Dublin Core dans les pages HTML.

La Scandinavie :

18-Le Projet de Metadata Nordique :

Page d'accueil:<http://linnea.helsinki.fi/meta/>

Dans les pays Nordiques, il y a un besoin spécial pour un système de création de metadata, pour une meilleure exploitation des documents en commun. Le Dublin Core est employé pour fournir et améliorer les services pour l'utilisateur final, en permettant une recherche plus efficace et un meilleur accès aux documents numériques.

La Suède :

19-Le projet EnviroNet Suédois :

Page d'accueil:<http://smn.environ.se/smnproj/proj/summary.htm>

EnviroNet est un projet du gouvernement Suédois qui est perçu comme une voie d'accès à l'information et aux données électroniques sur l'environnement en Suède. Il devra fournir des liens, des descriptions de données par le Dublin Core et autres services aux sites web des grandes agences publiques et compagnies privées travaillant dans le domaine de l'écologie.

Danemark :

20-Netpublikationer :

Page d'accueil:<http://www.fsk.dk/fsk/publ/+connecté-pub/>

Dans le but de rendre l'information publique plus efficace et plus accessible sur le web, toutes les nouvelles publications des ministères Danois ont commencées à être mises sur le web, en parallèle avec les éditions imprimées, à partir de 1997.

21-INDOREG: (INternet Document REGistration)

Page d'accueil:<http://www.purl.dk/rapport/html.uk/>

Projet du Centre Danois des bibliothèques (DBC) pour fournir des enregistrements de toutes les publications d'internet et l'accès à ces documents par DanBib (un système conjoint de superstructure pour le système de bibliothèque Danois)

Le Royaume-Uni :

22-ADAMIGVADS : (Art,Design,Architecture & Media Information Gateway and the Visual Arts Data Service)

Page d'accueil:<http://adam.ac.uk/>

'Art,Design, Architecture & Media Information Gateway' et 'Arts Data Service' sont deux services dont l'objectif est de fournir la communauté anglaise de l'éducation, d'un accès fiable, rapide, par réseau à des ressources d'information dans les arts visuels.

23-AHDS Arts & Humanities Data Service :

Page d'accueil:<http://ahds.ac.uk/>

Le AHDS est une organisation fédérale , consistant en un Exécutif central et cinq fournisseurs de service incluant l'archéologie, l'histoire, l'étude de textes et les arts visuels. L'objectif de cette organisation est de développer un système intégré capable de fournir à l'utilisateur des ressources électroniques disponibles chez chaque fournisseur de service.

24-Projet BIBLINK :

Page d'accueil:<http://www.ukoln.ac.uk/metadata/BIBLINK/>

Le projet est commun à plusieurs bibliothèques nationales de l'Union Européenne. Il vise à établir un lien entre les éditeurs et les Agences Bibliographiques Nationales (NBA'S), pour l'échange des enregistrements de metadata des articles nouvellement publiés. la phase de démonstration de ce projet a été prévue entre Novembre 1997 et Mars 1999.

25-Projet 'DESIRE' :

Page d'accueil : <http://www.nic.surfnet.nl/surfnet/projette/désir/desire.html>

LE projet DESIRE aborde deux approches pour la recherche de ressource d'information: Un service basé sur la sélection manuelle et la description de ressources de haute qualité, et un service régional de recherche basé sur les metadata générées par des agents de recherche du Web. Les objectifs du projet sont de contrôler et d'incorporer de nouveaux développements dans la gestion des metadata.

26-SCRAN :(Scottish Cultural Ressources Access Network)

Page d'accueil:<http://www.scran.ac.uk>

SCRAN est un projet pour créer une base de données en réseau, de ressources multimédia, pour les études, l'enseignement et les travaux d'histoire en Ecosse. Les partenaires sont les Musées Nationaux D'Ecosse, le Comité Royal sur Les monuments Historiques et anciens D'Ecosse, et le Conseil Ecossais des musées. Il est prévu de permettre un accès facile à 1.5 millions d'enregistrements de texte d'objets et monuments historiques et 100,000 ressources multimédia connexes seront disponibles pour l'an 2001.

27-NewsAgent : (pour les bibliothèques)

Page d'accueil:<http://www.sbu.ac.uk/litc/newsagent/>

L'objectif du projet NewsAgent est de créer un service de news et un service actualisé de sensibilisation pour le personnel de bibliothèque avec un mélange de contenus, fournissant jusqu'à des descriptions de date de documents pour les utilisateurs finaux(basé sur les préférences configurées de l'utilisateur).

28- ELISE II : (Electronic Library Image Service for Europe)

Page d'accueil:<http://severn.dmu.ac.uk/elise/>

Le service ELISE opérera sur le modèle client/serveur, en intégrant l'emploi de la norme Z39.50 et le Dublin Core.

Dans le prototype ELISE II, les données de catalogue fournies par les institutions participantes, sont reprises selon la syntaxe du Dublin Core.

Les Etats-Unis :

29- Bibliothèque Electronique de Monticello :

Page d'accueil: <http://www.solinet.net/monticello/monticel.htm>

La fonction fondamentale de la Bibliothèque Electronique de Monticello est de relier et distribuer des ressources régionales indépendamment de la source ou du type d'information. Le Dublin core est employé pour fournir l'interopérabilité sémantique entre plusieurs bases de données de médias électroniques et des types d'enregistrements incluant SGML, EAD(Encoded Archival Description), USMARC et des collections GILS.

30-Médical Metadata Project :

Page:<http://medir.ohsu.edu/~maletg/MedMetadata.HTM>

L'Université des sciences de la santé d'Oregon, l'Association Américaine d'informatique Médicale (Groupe de travail sur Internet) et l'institut National du Cancer fournissent un ensemble de tests de la base nationale de données génétiques.

31-FIUDL : (Florida International University Digital Library)

Page d'accueil:<http://www.fiu.edu/~diglib/>

Ce projet de bibliothèque numérique portera sur des images, du son et de la vidéo, incluant des modules multimédias de cours et de présentations. Il soutient les sujets des chercheurs et enseignants de l' Université de l'état de Floride.

32- Bibliothèque Numérique de L'université de Washington :

Page d'accueil:<http://content.engr.washing/>

La Bibliothèque Numérique de l'université de Washington contribue au développement et à l'adoption de descripteurs de ressource d'information; ces efforts initiaux portent sur des collections d'image. Le projet utilise maintenant les éléments du Dublin Core pour la description des collections d'images.

33-Everglades Information Network & Digital Library :

Page d'accueil:<http://everglades.fiu.edu/>

La collection de documents disponibles sur ce réseau inclue des rapports techniques, des articles scientifiques, des résumés et des programmes de conférences, des données sur la qualité de l'eau, des cartes, des diapositives et des photos, des documents légaux et des documents d'archives couvrant les sujets des recherches en écologie, en hydrologie, en écologie marine, et en restauration de marécages.

34-UMDLRD : (University of Michigan Digital Library Registry Database)

Page d'accueil:<http://dns.hti.umich.edu/enregistrement/>

Ce projet créera une base de données à partir des ressources du Web, choisies pour leurs valeurs académiques et institutionnelles. Il fournira des éléments de metadata, assez simples d'utilisation pour des non spécialistes.

35-Digital Library Catalog :

Page d'accueil:<http://sunsite.berkeley.edu/Cataloguer>

Ce projet concerne la description d'un fonds comprenant des livres, des dissertations, des discours, et autres textes, écrits en HTML, des rapports

techniques (dans divers formats), des photographies, des gravures et autres images, des clips vidéo et des clips audio.

36-GEM : (Gateway to Educational Materials)

Page d'accueil:<http://gem.syr.edu>

Le projet GEM est une initiative du département de l'éducation des USA et de la Bibliothèque Nationale de l'éducation.

Son but est d'améliorer l'organisation et l'accessibilité des importantes collections de documents de l'éducation, non catalogués et qui sont déjà disponibles sur divers sites Internet.

37-SRS : (Scout Report Signpost)

Page d'accueil:<http://www.signpost.org/signpost/index.html>

Développée par 'Internet Scout Project', avec le soutien de la NSF(National Sciences Fondation), Signpost est une base de données permettant la 'navigation' et la recherche d'information. C'est un projet de recherche utilisant une structure de Dublin Core modifiée dans le but de décrire des ressources Internet appropriées.

38- SiteSearch : (Universite de l'Arizona)

Page d'accueil: <http://dizzy.library.arizona.edu/sitesearch/welcome.html>

La Bibliothèque de l'université de l'Arizona a créé ce site pour fournir l'accès à diverses bases de données multimedia, sur Internet.

39-CIMI : (Consortium for the computer Interchange of Museum Information)

Page d'accueil:<http://www.cimi.org/documents/metafina/PD.html>

CIMI est un groupe d'institutions et d'organisations qui encourage une approche de gestion de l'information électronique des musées, par des standards simples et ouverts. C'est dans ce sens que ce consortium a adopté le Dublin Core pour cataloguer ses ressources d'information.

40- Plusieurs autres projets impliquant le développement de bibliothèques numériques ont adoptés le Dublin Core :

Exemple: *La bibliothèque numérique de la sécurité aérienne US*

Cette bibliothèque a mis en œuvre un site web contenant principalement de l'information choisie pour l'aviation en générale.

4) Metadata et outils de recherche :

Les metadata ne sont pas prises en compte encore par tous les moteurs de recherche. Certains ont commencé à en tenir compte succinctement dans l'indexation des documents mis sur Internet. Nous pouvons citer :

- Infoseek :Les balises Meta sont repérées dans le document HTML, mais aucune priorité ne leur est donnée.
- Altavista :L'indexation se fait d'abord sur le titre et les premiers mots du document, puis les champs « description » et « keyword » définis par les balises Meta.
- Hotbot :Les balises Meta sont prises en compte dans l'indexation, mais avec une moindre importance que les mots du titre.
- Ecila :Les balises Meta « keywords » sont prises en compte avant les mots du titre et du document, en l'absence de mots clés.
- HotMeta :Moteur de recherche(encore en période d'évaluation) orienté entièrement vers les documents décrits par des metadata (Dublin Core et SOIF).

5) CONCLUSION :

La metadata est somme toute une valeur ajoutée à l'information. Elle en permet la compilation et le repérage. Pour leur part, les bibliothèques produisent ce type d'information depuis longtemps. Par contre Internet en est à ses débuts. Indexer la masse d'information qui y circule, pour en faciliter l'accès aux utilisateurs nécessite un certain travail de préparation de l'information. L'ajout d'élément de metadata devient une action nécessaire.

Nous venons de passer en revue les principaux standards de metadata existants, même s'ils diffèrent souvent dans leurs structures, il n'en demeure qu'ils poursuivent un objectif principal commun : offrir des éléments de description de l'information pour en faciliter l'accès.

Il faut signaler que les standards de metadata du domaine géospatial avait pour but de fournir toute l'information sur les données décrites, et pas seulement l'information bibliographique. De plus, des éléments de description trop nombreux et trop technique dont l'emploi peut échapper au simple utilisateur, les éloignent de tout parallèle avec des fiches catalographiques classiques.

C'est peut-être cette relative complexité qui a fait que le Dublin Core, en offrant une structure simple, ait fédéré autour de lui beaucoup d'organismes et d'institutions. Au bout de trois ans, le Dublin Core est déjà opérationnel dans plus de 40 projets, de par le monde. Il est en constante évolution.

Une série de workshops lui assurent un suivi constant dans les développements. Des propriétés d'ouverture, d'extensibilité et d'interopérabilité, le rendent applicable à tous les domaines.

Des schémas de conversion des autres standards vers le Dublin Core, ou inversement, ont été développés, pour montrer sa portabilité. Il faut rappeler qu'au départ, les promoteurs du Dublin Core voulaient relever un défi important : offrir une aide améliorée pour le repérage de ressources d'information. Du moment qu'Internet renferme plus d'information que tout ce que les professionnels des bibliothèques peuvent prendre en charge par les méthodes classiques, il semblait logique de donner des outils aux auteurs et fournisseurs d'information

électronique pour leur permettre de décrire eux-même leurs documents. Cela a certes suscité quelques critiques, mais il n'empêche que depuis qu'une syntaxe spécifique, basée sur SGML a été approuvée par le W3C en 1996, le Dublin Core s'est mis sur la voie d'une norme internationale. C'est aussi en 1996, qu'a été adopté le *Warwick Framework*, au troisième workshop qui s'est tenu à la Warwick University, au Royaume-Uni. Ce dernier est une architecture ayant le potentiel de mettre ensemble des metadata à structures syntaxiques diverses, qui sont accessible et maintenues séparément.

En se conformant aux développements du W3C (RDF et XML), le Dublin Core s'affirme directement comme un futur standard de metadata international.

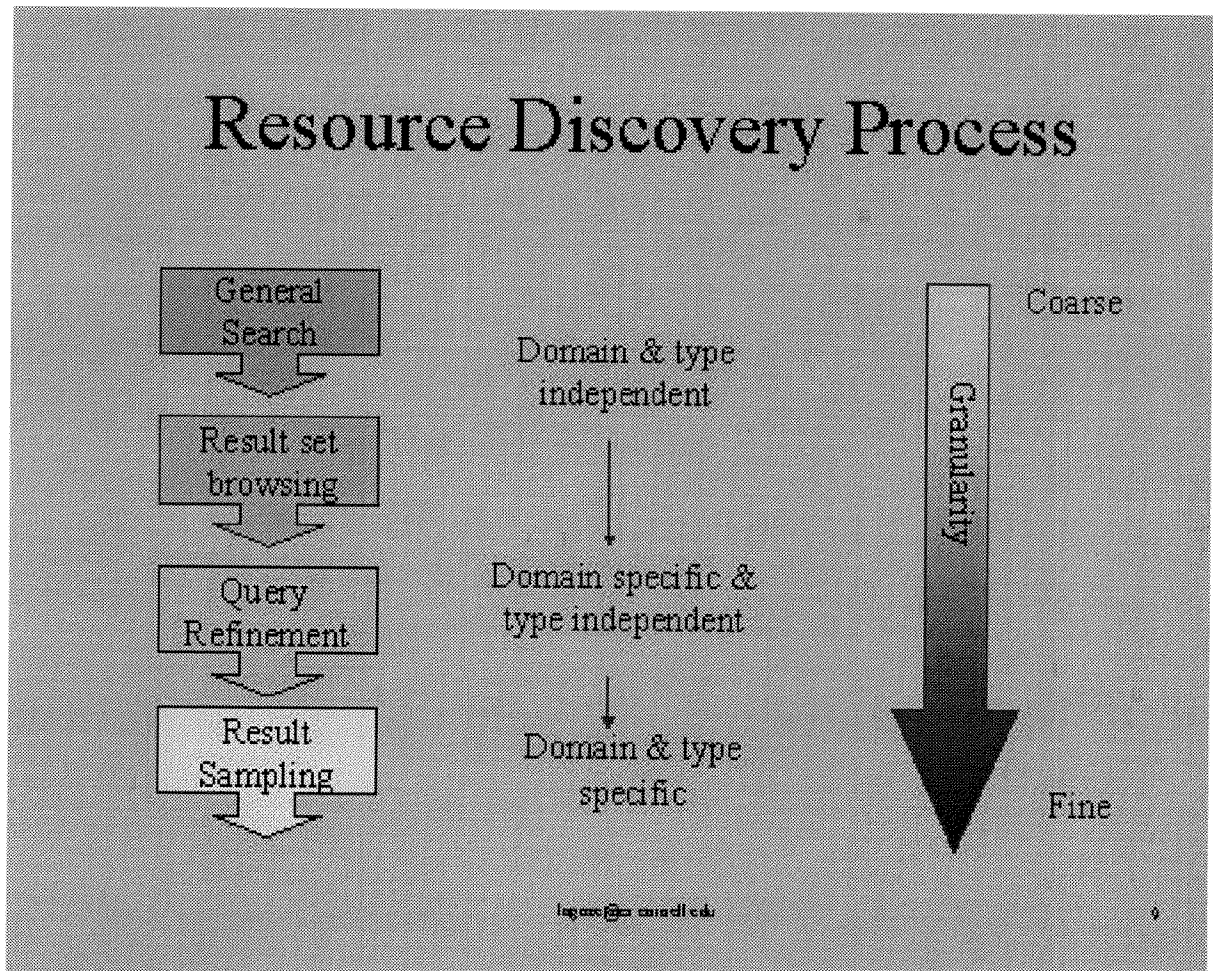
DEUXIEME PARTIE

1- Introduction au projet Profildoc:

1.1-Introduction :

Une production d'information en évolution permanente et une incroyable prolifération des ressources augmentent considérablement le volume d'information à consulter pour obtenir une réponse pertinente à une requête donnée. . En fait, dans les systèmes de recherches documentaires en texte intégral, il est toujours possible de retrouver des documents contenant un des termes de la question, sans pour autant qu'ils soient pertinents. Le fait de retrouver de pareils documents est appelé « *bruit* ». Celui-ci peut constituer une gêne certaine dès que le volume des réponses dépassent un seuil tolérable, surtout si on récupère l'intégralité des documents, alors que l'information effectivement utile dans ces documents, peut être très courte. Evidemment, ces systèmes tout en proposant presque toujours une réponse à la demande de l'utilisateur, n'arrivent en réalité à répondre que partiellement à ses besoins. En effet, la particularité de la recherche d'information provient du fait que l'utilisateur collecte des données pour un besoin précis, ou pour la résolution d'un problème bien défini. Dans ce cas-là, le système ne possède aucune information sur le contexte dans lequel il effectue sa recherche, ou sur les buts qu'il poursuit. Rappelons que pour réaliser une classique diffusion sélective de l'information (DSI), qu'on peut assimiler à un profil utilisateur, le bibliothécaire établit une équation de recherche (souvent constituée d'un ensemble structuré de descripteurs) en procédant à un entretien détaillé avec l'utilisateur. Les points d'accès ou de recherche recueillis sont les diverses caractéristiques d'une information ou d'un document à partir desquels peut s'opérer la recherche ou la sélection. Ils sont exprimés par l'utilisateur dans sa question par les indications qu'il donne sur le sujet, les dates, l'aire

géographique, le type de document recherché, la langue du document, l'auteur, l'éditeur, etc. De la même manière, pour qu'un système documentaire automatisé fournisse des réponses satisfaisantes, il faut qu'il ait une certaine connaissance du problème que l'utilisateur se pose. Si l'on cherche l'information sur un réseau, un autre fait vient amplifier le bruit. Sur un réseau comme Internet où circulent des informations issues de multiples serveurs répartis un peu partout dans le monde, des documents hétérogènes, sont présentés au même niveau, sans distinction particulière de domaine (la physique, la chimie, l'économie...), de nature (on retrouve pêle-mêle des images, du texte, du son), de contenu (pages personnelles, catalogues publicitaires, publications scientifiques, ...), ou de format (HTML, Postscript, texte,...). Pour pallier aux insuffisances de l'indexation de texte intégral et avoir une meilleure connaissance du fonds, les systèmes documentaires ont rajouté dans la description des documents, des critères externes à leurs contenus. Ainsi en bibliothéconomie classique, la dimension d'un ouvrage, son nombre de pages, etc..., sont autant de critères supplémentaires permettant d'améliorer la gestion du fonds, mais il est rare qu'un utilisateur se serve de ces critères pour sélectionner des documents. Grâce aux systèmes de gestion de fichiers ou aux systèmes de gestion de bases de données, la recherche d'une notice par l'ensemble des champs (zones) la décrivant est devenue possible ; des champs définissant des caractéristiques externes au contenu ont ainsi pu être rajoutés : le pays et le champ disciplinaire de l'auteur, le nom du laboratoire, etc. Ces systèmes documentaires permettent aujourd'hui de guider l'utilisateur, par un processus de recherche d'information affiné (souvent en plusieurs étapes) de remonter au document pertinent, qu'il pourrait ne pas retrouver par une recherche simple. Ce processus est très bien illustré par le schéma de CARL LAGOZE (Digital Library Scientist / Cornell University / USA), que nous proposons ci-dessous :



En réalité, le plus souvent, dans une opération de recherche documentaire classique, l'utilisateur se contente de formuler une requête donnée, puis le système apparie les mots de celle-ci avec ceux du dictionnaire qu'il possède et génère ainsi une réponse. Dans le cas des systèmes documentaires en texte intégral, il est toujours possible de trouver des documents contenant un des termes de la question, mais cela ne veut pas dire qu'ils seront vraiment pertinents pour l'utilisateur. Une étude approfondie sur un certain nombre de textes, livres, thèses, articles de revues scientifiques, a montré qu'on pouvait trouver, pour chacun d'eux, une structure générique facilement identifiable. En effet, dans la majorité des cas, un texte (article, conférence, rapport, ouvrage, etc.) a une structure générale, il forme une unité car il est construit pour faire passer un message : résultats de synthèse, nouvelles pistes de recherche, etc. Cette unité matérielle et intellectuelle est le résultat d'un lien parfaitement établi entre ses différentes parties, celles-ci

pouvant former à leur tour des unités indépendantes remplissant une fonction bien déterminée. Ainsi, par exemple, la bibliographie est utilisée généralement pour étayer les propos cités dans les différentes parties du texte et pour donner au lecteur une idée plus ou moins exhaustive de tout ce qui a été écrit sur le sujet.

L'équipe du laboratoire RECODOC de l'université LyonI, a lancé le projet Profildoc en s'appuyant sur l'hypothèse que l'insatisfaction de l'utilisateur par rapport aux réponses à une requête donnée présentée à un système documentaire de type textuel, résulte essentiellement d'une surabondance de documents non pertinents. La solution proposée par l'équipe est de prendre en considération d'une part le profil de l'utilisateur, et d'autre part la fonction d'usage des documents qui permet de leur affecter des propriétés. Cette constatation, a conduit l'équipe à admettre que " l'éclatement " du document en unités documentaires permet, tout en préservant l'unité globale du document (le lien entre l'unité documentaire et le document auquel elle appartient), de présenter à l'utilisateur une information plus affinée et plus facile à saisir. Evidemment, il faut considérer aussi les différents types de textes (publicitaires, scientifiques), le mode d'organisation du discours (narratif, argumentatif, etc.) ou même encore la structure physique (attributs typographiques, polices, espaces, etc.) comme des caractéristiques propres à discriminer une partie du document. Le découpage du document en unités documentaires est basé sur la fonction remplie par ces parties du document et non sur leur contenu. Au niveau de l'utilisateur, ces propriétés seront autant d'outils supplémentaires utilisables lors de la requête, pour sélectionner l'information " pertinente ". En effet, on peut remarquer que l'utilisateur, face à un système en texte intégral qui lui fournit généralement trop d'information, va développer une stratégie de recherche empirique. Il va par exemple se limiter à certaines bases de données, selon la discipline ou le type de revues répertoriées, ou encore, selon la langue, pays ou année. Toutes ces stratégies ont deux caractéristiques : elles portent sur des critères (la forme, le support, le style, etc...) autres que le contenu du document, elles sont très fortement individualisées et permettent une personnalisation de la recherche [Lain 01].

Ces propriétés permettront de sélectionner un corpus " personnalisé " suivant les caractéristiques de l'utilisateur, corpus sur lequel portera la question. En d'autres termes, ces propriétés, appariées avec le profil de l'utilisateur, nous permettent de présélectionner un ensemble de documents.

1.2- Principe de Profildoc :

Profildoc s'appuie sur trois composantes fondamentales :

- Un découpage des documents en unités documentaires, en se basant sur des fonctions d'usage, c'est à dire, sur l'utilité qu'elles peuvent avoir, ou sur le type d'utilisateur qui pourrait en avoir besoin ;
- Une caractérisation du profil de l'utilisateur, de ses compétences et habitudes, mais aussi de l'objectif de sa recherche ;
- Un système d'aiguillage qui à partir du profil de l'utilisateur extrait un sous-corpus personnalisé d'unités documentaires.

Le profil de l'utilisateur est défini par des caractéristiques telles que son niveau éducationnel, son champ disciplinaire, le type de recherche qu'il souhaite (recherche exploratoire, pointue, etc.), le cadre de recherche (réalisation d'un projet, mise à jour des connaissances, etc.). Cette caractérisation permet de définir avec précision les attentes de l'utilisateur. La fonction qui permet de définir l'ensemble des propriétés des unités documentaires relatives à un profil donné, est appelée : système d'aiguillage.

1.3- La phase d'interrogation :

Dans le système Profildoc, le processus d'interrogation comporte cinq étapes :

✓ Le découpage des documents :

Les unités appartenant au même document héritent de ses propriétés et se distinguent aussi par des propriétés qui leur sont propres. L'attribution des propriétés des unités documentaires n'exige pas nécessairement une étape de compréhension du contenu. Ces propriétés peuvent être facilement repérées à

l'intérieur du document ; comme le champ disciplinaire de l'auteur ou le style de l'unité documentaire ; ou alors repérées automatiquement par certains marqueurs (linguistiques ou autres), comme la forme discursive de l'unité documentaire, par exemple.

⇒ Les propriétés propres au document global (document père) sont les suivantes :

□ Les caractéristiques usuelles :

- titre du document,
- titre de la revue,
- auteur du document,
- coauteurs,
- affiliation de l'auteur,
- pays,
- année,

□ L'environnement de production :

- champ disciplinaire,
- communauté,
- profession de l'auteur,

□ Le support de diffusion :

- environnement éditorial,
- type d'article.

⇒ La seconde partie des propriétés est propre aux unités documentaires :

□ Le type de l'unité logique :

- résumé,
- introduction,
- description de contexte,
- description de thème,
- description de la méthode,
- environnement,

- expérimentation,
- résultat,
- discussion,
- développement,
- conclusion,
- bibliographie,
- table des matières,
- annexes.

□ La forme discursive du document :

- argumentatif,
- descriptif,
- narratif.

□ Le style du document :

- littéraire,
- littéraire contenant des données numériques,
- équations et formules de calculs,
- schémas ou figures.

✓ **Description du profil de l'utilisateur :**

Dans un contexte documentaire, la connaissance du profil de l'utilisateur permet au bibliothécaire ou au système d'information de retrouver, pour une requête donnée, des documents plus pertinents. Une même question peut induire des réponses différentes en fonction d'un grand nombre de facteurs objectifs ou subjectifs tels que le niveau éducationnel de l'utilisateur, sa profession, la ou les langues comprises, l'objectif précis de la recherche d'information et tout un contexte personnel dont, d'une façon implicite, le professionnel de l'information tient compte pour effectuer sa recherche.

Le système Profildoc propose quatre propriétés pour caractériser le profil de l'utilisateur :

- ❖ *Niveau éducationnel* : maîtrise ; DEA ; recherche.

- ❖ *Champ disciplinaire* : SIC (Sciences de l'Information et de la Communication) ; informatique ; agronomie ; pharmacie ; etc....
- ❖ *Etapas de recherche* : constitution d'une bibliographie ; définition du sujet ; faisabilité ; expérimentation ; interprétation des données ; rédaction ; repérage des approches expérimentales ; plan de travail ; compréhension de la problématique ; état de l'art ; synthèse bibliographique ; dégagement des nouveaux axes de recherche ; mise à jour des connaissances .
- ❖ *Type de recherche* : recherche pointue ; recherche généraliste.

✓ **Mise en place d'une fonction d'aiguillage :**

La fonction d'aiguillage est définie comme le «moteur » du système qui permet d'extraire les unités documentaires du corpus, en fonction du profil de l'utilisateur, donc de l'usage qu'il va en faire. La méthode est basée sur des observations tirées d'une enquête effectuée sur les usages et habitudes des chercheurs en SIC, sciences pharmaceutiques et sciences physiques; et sur une étude systématique des publications réalisées autour de ces questions. Une matrice " profil utilisateurs " a été construite à partir des résultats obtenus. La sélection, dans cette matrice, des colonnes décrivant le profil de l'utilisateur permet d'obtenir un ensemble de propriétés, utilisées pour présélectionner un ensemble d'unités documentaires, ensemble sur lequel portera la requête de l'utilisateur.

✓ **Requête de l'usager :**

La requête de l'usager se fait à partir du logiciel documentaire SPIRIT. Ce dernier permet la génération automatique de bases de données textuelles, et leurs interrogations en langage naturel.

✓ **Visualisation de la réponse :**

L'utilisateur choisira de visualiser telle ou telle partie du ou des documents auxquels correspondent les unités documentaires sélectionnées. Une lecture de type navigationnelle est suggérée car la lecture d'un article scientifique s'y prête du fait de sa structure et de "l'indépendance sémantique" de ses différentes parties.

2- Eléments descriptifs de Profildoc:

Dans le projet Profildoc, les éléments de description représentent les propriétés propres au document lui-même et celles propres à l'unité documentaire. Nous les regroupons dans deux tableaux distincts:

2.1-Propriétés propres au document:

| <i>Eléments de description</i> | <i>designations</i> |
|---------------------------------------|---|
| Référence de l'UD | Référence identifiant l'unité documentaire |
| Référence du document | Référence identifiant le document |
| Titre de la revue ou de la collection | Titre de la revue d'où est extrait l'article |
| Titre du document | Titre de l'article ou du document à traiter |
| Auteur du document | Auteur principal mentionné en premier |
| Co-auteur du document | Autres auteurs; préfaciers |
| Affiliation de l'auteur | Adresse complète de l'auteur |
| Pays | Pays de production du document et où les études ou expériences ont été faites |
| Année de publication | Année de publication du document |
| Type d'environnement éditorial | - M3 pour les thèses de 3 ^{ème} cycle; - PPROF pour les publications destinées à des professionnels(articles, ouvrages, etc....); - PGP pour les publications orientées grand public (documents de vulgarisation) ou à un public averti (articles, ouvrages, etc....); - PFOND pour les publications scientifiques et les articles extraits d'une revue internationale à comité de lecture; - DIVERS pour tout autre type non répertorié ci-dessus. |
| Profession de(s) l'auteur(s) | - E3 pour les étudiants; - SPE pour les spécialistes du secteur |

| <i>Eléments de description</i> | <i>designations</i> |
|---------------------------------|--|
| | <p>concerné;</p> <p>-J pour les journalistes;</p> <p>-DIVERS pour les autres.</p> |
| Champ disciplinaire de l'auteur | <p>-SIC pour les sciences de l'information et de la communication ,la bibliothéconomie et le journalisme;</p> <p>-INFO pour l'informatique;</p> <p>-MATHS pour les mathématiques;</p> <p>-ECO pour l'économie;</p> <p>-SOCIO pour la sociologie;</p> <p>-AGRO pour l'agronomie;</p> <p>-AUTRES à préciser selon besoin.</p> |
| Communauté de l'auteur | <p>-ETUD pour les étudiants;</p> <p>-UNIV pour les universitaires;</p> <p>-INDUS pour les grands groupes industriels;</p> <p>-PME pour les PME/PMI.</p> <p>-PUB pour le secteur public et para-public;</p> <p>-INDIV pour les auteurs indépendants.</p> |

2.2-Propriétés propres à l'unité documentaire:

| <i>Eléments de description</i> | <i>Désignations</i> |
|--------------------------------|--|
| Type de l'unité documentaire | <p>-Résumé et éventuels mots clés;</p> <p>-Table de matière (index, sommaire);</p> <p>-Introduction;</p> <p>-Description du contexte (état de l'art, enjeux, etc....)</p> <p>-Description du thème (thème général de l'étude et non seulement de l'article);</p> <p>-Environnement (outils disponibles);</p> |

| <i>Eléments de description</i> | <i>Désignations</i> |
|--|---|
| | <ul style="list-style-type: none"> -Description de la méthode; -Développement (divers...); -Expérimentation (mesures, enquêtes); -Résultats; -Discussion; -Conclusion; -Bibliographie; -Annexes. |
| Forme discursive de l'unité documentaire | <ul style="list-style-type: none"> -Descriptive ; - Narrative ; - Argumentative ; - Discours rapporté (où l'auteur se limite à la citation des travaux des autres) |
| Langage (style) de l'unité Documentaire | <ul style="list-style-type: none"> -LITT pour littéraire pur; -LITT+NUM pour le style littéraire avec des données numériques de formalisation; -QUANT pour les données numériques; -CAL pour les formules et équations; -SCHEMA pour des schémas et figures; -REP pour formalisation, représentation, symboles, programmes. |

3-Structuration de Profildoc selon le Dublin Core:

Les propriétés d'extensibilité et de répétabilité nous permettent de reprendre parfaitement les éléments de description de Profildoc selon la structure des éléments de description du Dublin Core. Les deux systèmes, n'ayant pas été conçus pour les mêmes finalités ne partagent évidemment pas les mêmes éléments de description.

Sur ce point de vue, nous pouvons reprendre le parallèle qu'ont fait Stuart Weibel et Erik Miller (Chercheurs à l' Online Computer Library Center et parmi les plus actifs du Groupe qui suivent les développements du Dublin Core) entre le système PICS (Platform for Internet Content Selection) et les metadata, d'un côté, et AACR2 (Anglo-American Cataloguing Rules 2) et USMARC, d'un autre côté. AACR2 s'intéresse à la sémantique et au sens du document; tandis que USMARC s'intéresse à la structure et à la syntaxe du document; de même, si PICS s'intéresse à la sémantique et au sens du document (pour un objectif évident de filtrage d'information), les metadata prennent en compte la structure et la syntaxe (pour un objectif de simple description de l'information sans tenir compte de son sens). Toutefois, clarifier ces nuances ne l'a pas empêché pour autant de conclure que l'association de ces deux concepts (et objectifs) dans le futur, seront la meilleure solution pour une meilleure information sur le réseau.

Nous nous retrouvons donc avec la même approche en ce qui concerne Profildoc et Dublin Core. Nous rappelons que Profildoc possède des éléments de description de l'information avec une orientation vers un objectif de filtrage (préférence utilisateur); alors que le Dublin Core s'arrête à la simple description de l'information. La faculté, liée aux propriétés d'extensibilité et de répétabilité, qu'a ce dernier, de permettre la duplication du même élément ou la création de sous-éléments pour prendre en charge la description d'information spécifique, nous a grandement facilité la prise en charge des éléments de Profildoc selon la structure du Dublin Core.

⇒ Structuration de Profildoc suivant l'architecture du Dublin Core:

Nous allons reprendre tous les éléments de description de Profildoc selon la structure des éléments du Dublin Core. Il est bien clair dans notre esprit que nous ne procédons pas à une simple opération de conversion classique telle qu'elle a été faite entre le Dublin Core et d'autres standards. Certaines de ces conversions ont été critiquées pour la relative perte d'information, qu'elles induisaient. Nous préférons repartir d'un principe qui a commencé à se

dégager des développements du Dublin Core, après le deuxième workshop qui s'est tenu à l'université de Warwick, en avril 1996, en Grande Bretagne.

Sandra D. Payette (Digital Library Research Group / Cornell University) décrit cela, dans une communication donnée lors d'un séminaire en novembre 1997, par ces quelques mots :

- ◆ *“Les enregistrements du Dublin Core sont l'ancre de tout système de description.*
- ◆ *Le Dublin Core est le bon choix pour les lignes de base des paquets de description qui peuvent être complétés par d'autres paquets spécialisés, au sein d'un même container.*
- ◆ *Les 15 éléments fondamentaux du Dublin Core sont un dénominateur commun pour la recherche d'information dans des ressources disparates.*
- ◆ *Les enregistrements du Dublin Core sont flexibles par leurs qualificatifs.”*

Nous rajouterons pour notre part que les propriétés de répétabilité et d'extensibilité des éléments de description du Dublin Core permettent à toute communauté de les utiliser tous ou en partie pour une description de ressource; et d'adjoindre d'autres éléments à portée plus spécifiques.

A partir de ce moment là, nous pensons que nous ne devons plus parler de conversion, mais plutôt de structuration d'éléments de description propres à un standard donné suivant l'architecture du Dublin Core. Les différents standards gagneraient en flexibilité, extensibilité, mais surtout, une syntaxe commune qui les rendraient intéropérables et facilement lisibles par machine.

Pour les éléments de description de ressources de Profildoc, nous aurons:

✓ ***Titre du document:***

Pour cet élément, nous retrouvons dans le Dublin Core un élément analogue, ayant les mêmes objectifs de description: nom donné à la ressource par l'auteur ou le créateur.

✓ ***Auteur:***

De même, pour cet élément, le Dublin Core propose un élément équivalent appelé auteur ou créateur et qui désigne la personne ou l'organisation principalement responsable de la création du contenu intellectuel de la ressource.

✓ ***Co-auteur:***

Pour cet élément, le Dublin Core ne prévoit pas de correspondant direct, mais La propriété de réputation et les usages dans la communauté du Dublin Core, nous permet de le proposer comme deuxième auteur.

✓ ***Affiliation de l'auteur:***

Non prévu dans le Dublin Core comme élément; mais Rebecca Guenter (Bibliothèque du Congrès US) l'a évoqué comme éventuel sous-élément de l'élément *auteur ou créateur*, dans la proposition de la liste des qualificatifs qu'elle a présentée au cinquième workshop qui s'est tenu à Helsinki, en Finlande.

(voir le document : <http://www.loc.gov/marc/dcqualif.html>)

✓ ***Profession de l'auteur:***

Non prévu également dans le Dublin Core; mais nous pouvons l'inclure comme sous-élément, rattaché à l'élément *auteur* de Profildoc .

Il désignera la profession de l'auteur ou créateur de la ressource.

✓ ***Champ disciplinaire de l'auteur:***

Élément non prévu aussi dans le Dublin Core. Nous le proposons en tant que sous-élément, rattaché à l'élément *auteur de Profildoc* pour nous renseigner sur sa spécialité universitaire, donc sur son domaine de savoir.

✓ ***Communauté de l'auteur:***

Celui-ci est aussi un élément non prévu dans le Dublin Core. Nous le proposons, en tant que sous-élément rattaché bien sûr à l'élément *auteur* de Profildoc; et désignant la communauté d'appartenance du créateur de la ressource décrite.

✓ **Type d'unité documentaire:**

Cet élément est spécifique au système Profildoc. Il définit le type de partie discrétionnaire, qui est le résultat d'un découpage de document suivant le principe de Profildoc. Il n'y a pas d'élément correspondant dans le DublinCore; mais sur le plan sémantique, il se rapproche de l'élément *type de ressource*.

✓ **Forme discursive de l'unité documentaire:**

C'est un élément spécifique à Profildoc, qui définit le type de contenu (descriptif, narratif, argumentatif, ou discours rapporté). Nous le rapprochons aussi, sur le plan sémantique de l'élément *type de ressource* du Dublin Core.

✓ **Type d'environnement éditorial:**

Élément de description non pris en charge par le Dublin Core. Nous le considérons sur le plan sémantique comme un type de production éditoriale; et le proposons par conséquent comme sous-élément rattaché à l'élément *éditeur* de Profildoc; élément dont nous proposons la création.

✓ **Date de publication:**

Nous retrouvons le même élément dans le Dublin Core. La juxtaposition du mot publication lève même une certaine ambiguïté incidemment introduite dans le Dublin Core, entre la date de publication ou de création du document ou de l'objet décrit, avec des dates de faits décrits ou en relation avec le contenu; lesquelles sont prises en charge par l'élément *couverture(coverage)*.

✓ **Style de l'unité documentaire:**

Elément de Profildoc qui désigne le style de l'unité documentaire (littéraire, données numériques, calculs et équations, schéma, représentation).

L'élément *format* du Dublin Core prend en charge ce genre de description. Nous rappelons qu'il est utilisé pour identifier le logiciel et, éventuellement, le matériel qui peuvent être nécessaires pour afficher ou traiter la ressource. Nous pensons que les deux types de description se rejoignent sur le plan sémantique.

✓ ***Référence du document:***

Le même élément existe dans le Dublin Core. Il est appelé *identificateur de la ressource* et est constitué d'une chaîne de caractères ou de nombres, utilisée pour identifier de façon unique, la ressource. Les exemples pour des ressources en réseau incluent les URLs et les URNs (si implémenté). D'autres identificateurs globaux et uniques, tels que l'ISBN (International Standard Book Number), ou d'autres non formellement définis, sont des candidats potentiels pour cet élément, dans le cas de ressources privées.

✓ ***Référence de l'unité documentaire:***

Pour cet élément, nous proposons la création d'un sous-élément rattaché à l'élément défini précédemment (référence du document), comme l'équivalent de l'élément *identificateur de ressource*, du Dublin Core. Pour cet élément, à l'opposé de la référence du document définie précédemment, nous utiliserons un identificateur propre à Profildoc.

✓ ***Titre de la revue:***

Cet élément correspond entièrement à l'élément du Dublin Core appelé *Source* défini comme "une chaîne de caractères utilisée pour identifier de façon unique le travail d'où la ressource est tirée..."; par conséquent, nous reprendrons la même structure pour Profildoc.

✓ **Pays:**

Dans le Dublin Core, l'élément *couverture (coverage)*, qui définit les caractéristiques spatiales ou/et temporelles du document ou de l'objet décrit, prend en compte ce genre d'information. Dans Profildoc, ce champ prend en charge uniquement l'information concernant la dénomination du pays. Il s'agit évidemment du pays dans lequel, le document a été produit et les études et expériences faites.

J'espère, en opérant de cette façon, ne pas froisser les tenants des conversions, s'il en existe. Je pense qu'elles ont été faites surtout pour comparer les types de descriptions propres à chaque standard, ce qui suppose un certain respect des spécificités de chacun et donc nécessairement une perte d'information à l'arrivée, aussi minime soit-elle.

De ce que nous avons compris jusqu'à maintenant des développements du Dublin Core et de l'essence même de son évolution, nous sommes complètement convaincus, évidemment en dehors des interprétations subjectives quant à ses objectifs, qu'il constitue la base nécessaire à tout lancement de projet de création de standard de metadata propre à une communauté donnée, qui pourra bénéficier d'une structure et d'une syntaxe toutes prêtes; avec des propriétés capitales telles l'optionnalité, la répétabilité et l'extensibilité; et des atouts majeurs tels l'interopérabilité et la lisibilité par machine. Cela permettra aussi un rattachement à des développements majeurs, pour le moment, du W3C, que sont La RDF et XML et un état de permanente évolution. La meilleure preuve de tout cela est qu'en l'espace de trois ans, plus de quarante projets se sont basés sur le Dublin Core, avec des spécificités très différentes les unes des autres, des fois des langues différentes. La plupart sont des standards de metadata dérivés, avec des nombres et des types d'élément de description tout à fait différents, mais tous ont adopté le formalisme du Dublin Core, avec les mêmes propriétés et utilisent la même syntaxe....

4- Définition des éléments de metadata de Profildoc:

A partir des définitions établies au chapitre précédent, nous avons repris tous les éléments descriptifs de Profildoc selon la structure du Dublin Core. Pour illustrer cela, nous avons dressé un tableau des équivalences:

| Eléments simples de Profildoc | Eléments structurés selon le Dublin Core |
|--|---|
| -Titre du document | -Titre Etiquette:Titre • DC.Titre |
| -Auteur -Co-auteur(s) -Affiliation de l'auteur -Profession de l'auteur -Champ disciplinaire de l'auteur -Communauté de l'auteur | -Auteur ou créateur Etiquette:Créateur • DC.Créateur -Sous-éléments(type): • DC.Créateur • DC.Créateur. Affiliation • DC.Créateur.Profession • DC.Créateur. Discipline • DC.Créateur.Communauté |
| -Type d'unité documentaire | -Type d'unité documentaire Etiquette:Type • DC.Type |
| -Forme discursive de l'unité documentaire | -forme discursive de l'unité documentaire -Etiquette : Forme • DC.Forme |
| -Editeur | -Editeur Etiquette:Editeur -Sous-élément (type): |

| Eléments simples de Profildoc | Eléments structurés selon le Dublin Core |
|--|---|
| -Type d'environnement éditorial | <ul style="list-style-type: none"> • DC.Editeur.TypeProduction |
| -Date de publication | -Date Etiquette:Date <ul style="list-style-type: none"> • DC.Date |
| -Style de l'unité documentaire | -Style de l'unité documentaire Etiquette:Style <ul style="list-style-type: none"> • DC. Style |
| -Référence du document -Référence de l'unité documentaire | -Identificateur de la ressource Etiquette:Référence -Sous-élément (type): <ul style="list-style-type: none"> • DC.Référence • DC.Référence.UDRéférence |
| -Titre de la revue où de la collection | -Source Etiquette:Source <ul style="list-style-type: none"> • DC.Source |
| -Pays | -Pays Etiquette:Pays <ul style="list-style-type: none"> • DC. Pays |

5- Définition de qualificateurs propres à Profildoc:

5.1-Définitions:

Nous allons repasser en revue tous les éléments de metadata de Profildoc, définis précédemment et proposer pour chacun d'eux des qualificateurs selon la procédure utilisée dans le Dublin Core et conformément à la définition de ces derniers, que nous rappellons ci-après:

➤ *Sous-élément (ou type):*

Il affine et clarifie la définition de l'élément auquel il est rattaché. Par cette fonction, il exprime aussi la propriété d'extensibilité du système.

➤ *Schéma:*

Il permet à la valeur d'un élément d'être identifiée en tant que donnée extraite d'un système de classification, d'un code, d'un glossaire ou d'un thésaurus existant.

➤ **Langage:**

Ce dernier qualificateur, introduit au quatrième workshop sur le Dublin Core, qui s'est tenu à Camberra en Australie, en Mars 1997, indique le langage du contenu **d'un élément** de metadata. Il faut bien comprendre ici que chaque élément de metadata peut être écrit dans une langue, différente ou non du contenu décrit; mais surtout qu'il peut être aussi écrit dans une langue différente de celle d'autres éléments qui entrent dans la même description.. Exemple: Le titre d'un document en anglais peut être reporté dans une metadata en français; avec une répétition de l'élément "*titre*" contenant le titre original; et dans la même description de ce document (description par des éléments de metadata en français), nous pouvons retrouver un résumé en français et des descripteurs en anglais.

5.2-Qualificateurs proposés pour les éléments de Profildoc:

❖ **Titre du document:**

Etiquette : Titre

-Sous-élément:

- DC.Titre
- DC.Titre.Alternatif (pour tout autre titre ou sous-titre)

-*Shéma* : texte libre par défaut

-Langage : indiquée généralement par deux lettres (RFC 1766)

REMARQUE:

Le qualificateur "langage" sera identifié par le même contenu (RFC 1766) pour tous les éléments; donc nous nous abstenons de le répéter à chaque fois.

❖ **Auteur :**

Etiquette : Créateur

-Sous-élément:

- DC.Créateur
- DC.Créateur.affiliation
- DC.Créateur.Profession
- DC.Créateur.Discipline
- DC.Créateur.Communauté

-Schéma : d'après liste définie dans Profildoc:

| | |
|---------------------------------|---|
| Profession de(s) l'auteur(s) | - E3 pour les étudiants; - SPE pour les spécialistes du secteur concerné; - J pour les journalistes; - DIVERS pour les autres. |
| Champ disciplinaire de l'auteur | - SIC pour les sciences de l'information et de la communication , la bibliothéconomie et le journalisme; - INFO pour l'informatique; - MATHS pour les mathématiques; - ECO pour l'économie; - SOCIO pour la sociologie; - AGRO pour l'agronomie; - AUTRES à préciser selon besoin. |
| Communauté de l'auteur | - ETUD pour les étudiants; - UNIV pour les universitaires; - INDUS pour les grands groupes industriels; - PME pour les PME/PMI. - PUB pour le secteur public et para-public; - INDIV pour les auteurs indépendants. |

❖ *Type de l'unité documentaire :*

Etiquette : type

-Sous-élément :

Non défini

- Schéma : liste définie dans Profildoc que nous donnons ci-après:

| <i>Eléments de description</i> | <i>Désignations</i> |
|--------------------------------|---|
| Type de l'unité documentaire | <p>-Résumé et éventuels mots clés;</p> <p>-Table de matière (index, sommaire);</p> <p>-Introduction;</p> <p>-Description du contexte (état de l'art, enjeux, etc....)</p> <p>-Description du thème (thème général de l'étude et non seulement de l'article);</p> <p>-Environnement (outils disponibles);</p> <p>-Description de la méthode;</p> <p>-Développement (divers...);</p> <p>-Expérimentation (mesures, enquêtes);</p> <p>-Résultats;</p> <p>-Discussion;</p> <p>-Conclusion;</p> <p>-Bibliographie;</p> <p>-Annexes.</p> |

❖ *Forme discursive de l'unité documentaire :*

Etiquette : Forme

-Sous-élément : non défini

-Schéma : liste définie dans Profildoc:

| | |
|--|--|
| Forme discursive de l'unité documentaire | <p>-Descriptive ;</p> <p>- Narrative ;</p> <p>- Argumentative ;</p> <p>- Discours rapporté (où l'auteur se limite à la citation des travaux des autres)</p> |
|--|--|

❖ *Type d'environnement éditorial:*

Etiquette : Editeur

-Sous-élément :

- DC.Editeur
- DC.Editeur.TypeProduction

-Schéma : d'après liste Profildoc donnée ci-dessous:

| | |
|---------------------------------------|--|
| <p>Type d'environnement éditorial</p> | <p>-M3 pour les thèses de 3^{ème} cycle</p> <p>-PPROF pour les publications destinées à d professionnels(articles, ouvrages, etc....)</p> <p>-PGP pour les publications orientées grand public (documents de vulgarisation) ou à un public averti(articles, ouvrages, etc....)</p> <p>-PFOND pour les publications scientifiques et les articles extraits d'une revue internationale à comité de lecture.</p> <p>-DIVERS pour tout autre type non repertorié ci-dessus.</p> |
|---------------------------------------|--|

❖ *Date de publication:*

Etiquette : Date

-Sous-élément : non défini

- Schéma :

- ISO 8601
- ANSI X3.30
- IETF RFC 822

❖ *Style de l'unité documentaire:*

Etiquette : Style

-Sous-élément :

non défini

- Schéma : d'après liste Profildoc donnée ci-dessous:

| | |
|--|---|
| <p>Style de l'unité Documentaire</p> | <p>-LITT pour littéraire pur; -LITT+NUM pour le style littéraire avec des données numériques de formalisation; -QUANT pour les données numériques; -CAL pour les formules et équations; -SCHEMA pour des schémas et figures; -REP pour formalisation, représentation, symboles, programmes.</p> |
|--|---|

❖ *Référence du document :*

Etiquette : Identificateur

-Sous-élément :

- DC.Identificateur
- DC.Identificateur.UDRéférence

- Schéma : à définir dans Profildoc

❖ *Titre de la revue ou de la collection:*

Etiquette : Source

- Sous-élément : non défini

- Schéma : texte libre

❖ *Pays de l'auteur :*

Etiquette : Pays

-Sous-élément : non défini

- Schéma : texte libre

6- Syntaxe de transcription en HTML :

6.1- Définitions:

Nous allons adopter pour la transcription en HTML, des éléments de description de Profildoc, les conventions proposés pour le Dublin Core lors du workshop du W3C, tenu sur ce sujet les 28-29 Mai 1996 (W3C Distributed Indexing and Searching Workshop) ; auquel ont assistés des représentants du Dublin Core, de Lycos, de Microsoft, de WebCrawler, de IEEE, de Verity Software et du W3C. Le but était de convenir d'une convention simple pour coder une information de metadata, structurée, d'une variété de types. Le choix d'un codage en HTML permet d'avoir un format ou une forme de représentation interprétable aussi bien par l'homme que par la machine. Les metadata sont extraites automatiquement et ne doivent pas apparaître à la visualisation du document dans le navigateur ou à sa sortie sur imprimante.

En HTML, chaque définition d'élément d'enregistrement commence avec "<META" et fini par ">".

Pour transcrire des éléments de metadata en HTML, la balise <META> est placée entre le <HEAD> et le </HEAD> de fin de paragraphe en utilisant la syntaxe suivante:

```
<META NAME = "DC.NomElément" CONTENT = "Valeur">
```

Dans l'écriture ci-dessus, 'NomElément' et 'Valeur' contiennent une désignation de l'élément et sa 'valeur'.

Exemple :

```
<META NAME = "DC.Créateur" CONTENT = "William Shakespeare">
```

6.1.1- Metadata autonome :

Une metadata autonome peut exister dans toute sorte de base de données. Cet exemple décrit une photo dans un autre dossier qui a un emplacement donné par une URL (Uniform Resource Locator) :

```
<META NAME="DC.Title" CONTENT="Kita Yama (Japan)">
```

```
<META NAME="DC.Creator" CONTENT="Kertesz, Andre">
```

```
<META NAME="DC.Date" CONTENT="1968">
```



```
<META NAME="DC.Type" CONTENT="e/photograph">
<META NAME="DC.Format" CONTENT="image/gif">
<META NAME="DC.Identifieur" CONTENT=
"http://foo.bar.zaf/kertesz/kyama">
```

6.1.2- Metadata Contenu dans une Ressource :

L'exemple ci-dessous est un enregistrement de metadata contenu dans un dossier le long du document qu'il décrit. Le document lui-même est un court poème exprimé dans le format HTML :

```
<HTML>
<HEAD>
<TITLE>Que serais-je sans toi</TITLE>
<META NAME="DC.Title" CONTENT="Que serais-je sans toi">
<META NAME="DC.Creator" CONTENT="Aragon">
<META NAME="DC.Type" CONTENT="e/document">
<META NAME="DC.Date" CONTENT="1959">
<META NAME="DC.Format" CONTENT="text/html">
<METANAME="DC.Identifieur"CONTENT="http://www.enssib.fr/ po.html">
</HEAD>
<BODY><PRE>
Que serais-je sans toi qui vins à ma rencontre
Que serais-je sans toi qu'un cœur au bois dormant
Que cette heure arrêtée au cadran de la montre
Que serais-je sans toi que ce balbutiement
</PRE></BODY>
</HTML>
```

5.1.3- Principes de base des éléments de description :

✓ La notation que nous décrivons est basée sur les balises META. Dans le cas d'utilisation de caractères non-ASCII, il faudrait utiliser la même disposition des mots que dans le corps du document.

✓ Chaque élément est optionnel et répétable. Les éléments de Metadata peuvent apparaître dans n'importe quel ordre. La disposition de multiples occurrences du même élément peut avoir une signification voulue par le fournisseur de l'information ; mais cette disposition n'est pas assurée d'être préservée dans chaque environnement utilisateur.

✓ La Convention proposée pour la transcription des Metadata dans HTML prévoit leur identification et leur groupement. Cette convention compte sur l'utilisation de préfixe pour indiquer que l'élément utilisé appartient au Dublin Core ou à un autre standard. Pour le Dublin Core, le préfixe "DC" doit être écrit avant l'étiquette de l'élément et en lettres capitales.

exemple:

META NAME="DC.Titre"

META NAME="DC.Créateur"

⇒ Pour le cas de Profildoc, nous proposons de ne pas utiliser de préfixe pour le moment, comme c'est le cas pour la plupart des standards de metadata dérivant du Dublin Core :

Exemple :

META NAME=" Titre"

META NAME=" Créateur"

6.1.4- Contenu des éléments de metadata et Vocabulaire contrôlé :

Les données du contenu de quelques éléments doivent être sélectionnées à partir d'un vocabulaire contrôlé.

6.2- Transcription des éléments de Profildoc en HTML :

1. Référence de la ressource :

Etiquette : Référence

Description :

Cet élément est défini par une chaîne de caractères ou un nombre, utilisé pour uniquement identifier le document en tant que ressource. Il peut être une référence locale, attribuée dans le cadre de Profildoc ; ou un autre

type d'identificateur tel l'ISSN, l'ISBN, l'URI, l'URL, l'URN, le DOI, etc...

Transcription :

<META NAME="Référence" CONTENT="identificateur" >

Exemple :

<META NAME="Référence" CONTENT=

"http://recodoc.univ-lyon1.fr/doc.html/">

<META NAME="Référence" CONTENT="0385424728" [ISBN] >

- La référence de l'unité documentaire est définie de la même manière, dans un sous-champ, pour identifier celle-ci en tant que ressource aussi. Elle constitue ainsi un sous-élément, dont la valeur est définie uniquement dans le cadre de Profildoc (identificateur local).

Transcription :

<META NAME="Référence.UDoc" CONTENT="identificateur">

Exemple :

<META NAME="Référence.UDoc" CONTENT="PR607.A26W574 .1998">

2. Titre du document :

Etiquette : Titre

Description :

Nom donné à la ressource par l'auteur ou le créateur

Transcription :

<META NAME=" Titre" CONTENT="Nom de la ressource">

Exemples:

<META NAME=" Titre" CONTENT="Les fleurs du mal">

<META NAME=" Titre" CONTENT="L'albatros">

3. Titre de la revue ou de la collection :

Etiquette : Source

Description: Ce champ doit comporter l'information sur la source à partir de laquelle a été tiré le document décrit. Il est généralement recommandé

que cette information identifie clairement la source.

Transcription :

<META NAME=" Source" CONTENT="donnée descriptive">

Exemple :

<META NAME=" Source" CONTENT="adonis">

*Adonis : base de donnée textuelle en sciences médicales

4. Auteur :

Etiquette: Créateur

Description :

Personne ou organisme principalement responsable de la création du contenu intellectuel de la ressource. Les auteurs doivent être repris séparément, dans le même ordre qu'ils apparaissent dans la publication. Le champ *auteur* peut être répété dans le cas de l'existence de coauteurs.

Transcription :

<META NAME=" Créateur" CONTENT="Nom de l'auteur">

Exemples:

<META NAME=" Créateur" CONTENT=" Baudelaire, Charles">

<META NAME=" Créateur" CONTENT=" Marx, Karl">

<META NAME=" Créateur" CONTENT="Engels, Friedrich">

*Nous pouvons regrouper ces deux derniers coauteurs dans un seul champ :

<META NAME=" Créateur" CONTENT=" Marx, Karl ; Engels, Friedrich">

*Nous pouvons aussi introduire les sous-éléments rattachés à l'élément *auteur* :

<META NAME="Créateur.Affiliation" CONTENT=" donnée descriptive">

<META NAME=" Créateur. Profession" CONTENT=" donnée descriptive">

<META NAME=" Créateur. Discipline" CONTENT=" donnée descriptive">

<META NAME=" Créateur. Communauté" CONTENT=" donnée descriptive">

Exemples :

```
<META NAME="Créateur.Affiliation" CONTENT="
Université Lyon1. Laboratoire Recodoc">
<META NAME=" Créateur. Profession" CONTENT=" enseignant">
<META NAME=" Créateur. Discipline" CONTENT=" informatique">
<META NAME=" Créateur. Communauté" CONTENT=" universitaire">
```

6. Pays :

Etiquette : Pays

Description :

Ce champ permet d'identifier le ou les pays dans lequel, le document a été produit et les études ou expériences, qui lui ont données lieu, ont été faites.

Transcription :

```
<META NAME=" Pays" CONTENT="donnée descriptive">
```

Exemple :

```
<META NAME=" Pays" CONTENT="France">
```

```
<META NAME=" Pays" CONTENT="France ; Italie">
```

Remarques :

- Dans le cas où il s'agirait de plus d'un pays, il est recommandé de les citer dans un ordre chronologique tel l'exemple ci-dessus.
- Le nom du pays est écrit en texte libre.

7. Date de publication :

Etiquette : Date

Description :

Cette date est associée à la création ou à la validité de la ressource décrite.

Il est recommandé de toujours l'écrire dans le format ISO8601.

La note technique du W3C :

[<http://www.w3.org/TR/NOTE-datetime>] inclue le format suivant : YYYY et YYYY-MM-DD. Ainsi, par exemple, la date 1998-06-15 correspond au 15 juin 1998. Si des éléments de la date ne sont pas connus, écrire seulement le mois et l'année, ou uniquement l'année ; comme par exemple :

<META NAME=" Date" CONTENT="1998-06-15">

<META NAME=" Date" CONTENT="1998-06">

<META NAME=" Date" CONTENT="1998">

8. Editeur:

Etiquette : Editeur

Description :

Entité responsable de la publication de la ressource dans sa forme actuelle, telle une maison d'édition, un département universitaire, ou une société savante. Le but, en spécifiant ce champ est de permettre l'identification de l'entité fournissant l'accès à la ressource décrite. Nous rattachons à cet élément, le sous-élément environnement éditorial qui définit le type de publication ou le type de production éditoriale, parmi une liste définie dans le système Profildoc, qui comprend : Les mémoires et thèses de 3^o cycle ; la presse professionnelle ; la presse grand public ; la presse fondamentale ; et divers autres types.

Transcription :

<META NAME="Editeur" CONTENT=" donnée ou attribut">

<META NAME="Editeur.Type" CONTENT=" donnée ou attribut">

Exemples:

<META NAME="Editeur" CONTENT="

"Université Lyon I.Laboratoire recodoc">

<META NAME="Editeur.Type" CONTENT=" thèse">

8. Type d'unité documentaire

Etiquette : Type

Description :

Ce champ décrit la catégorie de la ressource ; et en l'occurrence, dans notre cas, de l'unité documentaire. Les schémas de description ont été détaillés au chapitre consacré aux qualificatifs de Profildoc.

Transcription :

```
<META NAME=" Type" CONTENT="donnée de description">
```

Exemple:

```
<META NAME=" Type" CONTENT="introduction">
```

9. *Forme discursive de l'unité documentaire :*

Etiquette : Forme

Description :

Cet élément décrit la forme discursive de l'unité documentaire, définie par les attributs : Descriptif ; Narratif ; Argumentatif ; Discours rapporté.

Transcription :

```
<META NAME="Forme" CONTENT="donnée ou attribut">
```

Exemple:

```
<META NAME="Forme" CONTENT="narratif">
```

10. *Style de l'unité documentaire :*

Etiquette : Style

Description :

Cet élément décrit le style ou le langage du contenu de l'unité documentaire, qui doit être perçu comme un type de ressource.

Les schémas de description s'inspireront de la liste établie dans le cadre du système Profildoc ; et qui comprend les qualificatifs suivants :

Littéraire pur ; littéraire contenant des données numériques ; données numériques ; calculs équations ; schémas ; représentation (algorithme ; diagramme ; image).

Transcription :

```
<META NAME="Style" CONTENT="donnée de description">
```

Exemple :

```
<META NAME="Style" CONTENT="littéraire pur">
```

6.3- Syntaxe HTML 4.0 :

Cette nouvelle syntaxe a été adoptée par le W3C en 1997 ; et sera utilisée dès la généralisation de HTML 4.0. Elle se présente comme suit :

<META NAME="DC.NomElément.SousElement"

SCHEME="valeur ou donnée descriptive"

LANG="attribut"

CONTENT="valeur ou contenu">

où:

- 'NomElément' représente le nom de l'un des 15 éléments du Dublin Core.
- 'SousElément' représente le nom du sous-élément rattaché à l'élément considéré.
- 'Valeur ou donnée descriptive' représente un identifiant extrait d'une classification, d'un vocabulaire contrôlé ou d'une liste d'autorité.
- 'attribut' représente un identifiant pour le langage dans lequel la metadata est écrite.
- 'Valeur ou contenu' représente la valeur de l'élément de metadata considéré.

*Toutes les valeurs associées aux qualificateurs sont optionnelles et doivent être évitées si leur usage est inapproprié ou risque d'entraîner une ambiguïté quelconque.

7- Metadata et profil utilisateur :

7.1. Le profil utilisateur « classique » :

Depuis toujours, les bibliothécaires savent qu'une recherche d'information ne vise jamais à obtenir des renseignements pour eux-mêmes ; elle est faite en vue de leur exploitation extérieure dans des conditions précises. Il faut donc que ces conditions soient connues. En particulier, il convient de savoir :

- Qui est le demandeur.
- Quelle utilisation compte-t-il faire des informations.
- De quel délai il dispose.
- Quels documents connaît-il déjà sur la question et, d'une manière générale, ce qu'il sait déjà sur le sujet.
- Quelles langues peut-il lire.
- Sous quelle forme, préfère-t-il obtenir les informations.
- Quelle période et quelle aire géographique la question couvre-t-elle exactement.

La formulation des questions par les utilisateurs risque d'être imprécise ou ambiguë à plus d'un titre :

- ✓ d'abord au niveau de la description du sujet, qui peut être trop large ou trop restreint ;
- ✓ ensuite au niveau de l'utilisation envisagée des informations recueillies. Le même sujet peut être traité différemment par des documents de différents types, dont chacun peut être mieux adapté à une utilisation qu'à une autre. Par exemple, un article résumant les principales orientations d'un plan de développement économique peut en donner une vue d'ensemble, mais ne pourra pas permettre d'entamer un travail spécifique d'analyse économique. Il y faudrait dans ce cas le document du plan même.
- ✓ Enfin au niveau des conditions dans lesquelles les informations devront être employées. Faire établir une bibliographie d'une centaine de références et rechercher les documents correspondants n'a guère d'utilité quand le demandeur doit produire une note de synthèse sur le sujet en un temps très court (vingt-quatre heures par exemple).

Autrement dit, la question la plus fréquente : « Quelles informations avez-vous sur tel sujet ? » devrait être formulée par une phrase du type : « Avez-vous sur tel sujet tel type d'information me permettant de faire tel travail, dans telles conditions ? ».

Cela nous amène à la notion de profil utilisateur classique tel qu'il est défini en bibliothéconomie ; c'est à dire comme une équation de recherche (ensemble structuré de descripteurs), exprimant les informations que l'utilisateur désire recevoir régulièrement d'un service de diffusion sélective de l'information (DSI). L'utilisateur peut construire lui-même son propre profil, en s'aidant d'un manuel que lui fournit le service de DSI ; mais le plus souvent, le profil est construit par un spécialiste de l'information, qui procède d'abord à un entretien détaillé avec l'utilisateur et à des essais qu'il soumet à son appréciation. Chaque envoi de résultats est accompagné d'une demande d'évaluation qui permet de corriger rapidement les insuffisances éventuelles du profil établi. L'utilisateur est invité à demander des modifications de son profil, au fur et à mesure que ses centres d'intérêt évoluent.

Les points d'accès ou de recherche sont les diverses caractéristiques d'une information ou d'un document à partir desquels peuvent être opérées à la fois, la recherche et la sélection. Ils sont exprimés par l'utilisateur dans sa question, par les indications qu'il donne sur le sujet, les dates, l'aire géographique, le type de document recherché, la langue, etc... Ils sont aussi fonction, d'une part du détail plus ou moins important de la description bibliographique et de la description de contenu (c'est à dire de leur présence dans la base d'information) et, d'autre part de la finesse du système de recherche, c'est à dire de la possibilité ou non de faire des tris dans les fichiers ou index, en fonction de ces caractéristiques. Ils portent en général sur les sujets traités, la date des informations ou des documents, l'aire géographique, l'auteur, le type de document (et donc de traitement des sujets) ; et ils peuvent aussi parfois porter sur la langue, le volume, l'accessibilité, le numéro de rapport ou brevet, le lieu de publication, etc...

Nous venons de faire ce bref rappel pour tout juste montrer que depuis très longtemps, si ce n'est pas depuis toujours, les spécialistes de l'information ont compris l'intérêt d'associer le profil utilisateur au processus de recherche de l'information pour améliorer nettement la pertinence des documents retrouvés. Cela nous permet aussi d'introduire un projet, en Grande Bretagne, associant ces principes avec une description de contenu par des éléments de metadata du Dublin Core.

7.2. Profil utilisateur et Dublin Core : le projet NewsAgent

L'objectif de ce projet est de créer un service de préférence utilisateur basée sur un système de news (groupe de discussions Usenet) et un service actualisé de sensibilisation pour le personnel de l'information et des bibliothèques avec un mélange de contenus courants, incluant des metadata (descriptions de documents). Le contenu sera constitué essentiellement d'articles et rapports tirés de revues, telles *Programm*, *Vine*, *Library Technology*, *Ariadne* et le *Journal of Librarianship and Information Science*. Les news et les fiches analytiques transmis aux utilisateurs sont fournis par The Library Association, The Institute of Information Scientists, UKOLN, The British Library, et LITC. Le projet a démarré en avril 1996 et la Phase I a pris fin en mars 1998. Les partenaires continuent maintenant avec davantage de travail pour réaliser un service de préférence utilisateur, à pleine échelle. Les partenaires principaux du consortium impliqués dans ce projet, sont:

- LITC, South Bank University, Londres (Coordinateur de projet)
- CERLIM, Manchester Metropolitan University (Evaluation)
- Dept of Information and Library Studies, UWA, Aberystwyth
- Fretwell-Downing Informatics Ltd (développement technique)
- UK Office for Library and Information Networking (UKOLN)

Finalement le principe qui permet au projet NewsAgent de prendre en charge le profil utilisateur, est le même que celui de la DSI (Diffusion Sélective de

l'Information). Les changements par rapport à la méthode classique, connue chez les bibliothécaires sont :

- L'intégration des éléments de metadata du Dublin Core dans la description du contenu ;
- Les éléments qui définissent le profil utilisateur sont décrits par l'utilisateur lui-même ; sur un formulaire mis en réseau ;
- Les résultats sont reçus dans la boîte de messagerie électronique de l'utilisateur.

Il faut noter que le projet n'est qu'à sa première phase. Il y a tellement d'institutions anglaises prestigieuses (du monde de l'information), impliquées dans ce projet que nous préférons attendre la fin de la deuxième phase ; date à partir de laquelle, est prévue une généralisation de son emploi à plus grande échelle, pour oser porter un œil critique dessus.

Nous avons tenu à citer ce projet pour le principal fait que c'est le premier, à notre humble connaissance a avoir associé la notion de profil utilisateur à une élaboration de requêtes dans un système d'information, basé sur une description de contenu par des éléments de metadata du Dublin Core.

7.3.-Profil utilisateur et format d'affichage :

Beaucoup d'auteurs ont relevés l'inutilité de certains champs qui se retrouvent dans les formats d'affichage, auprès de la plupart des utilisateurs. Crawford (1992) se réfèrent à « l'arcana » des données montrées dans les enregistrements bibliographiques. Montrer des éléments inutiles et non voulus peut entraîner de la confusion et rebuter les utilisateurs pour poursuivre la suite de l'affichage (plusieurs auteurs cités par Wallace, 1984). Le coût de ces enregistrements inutiles a été également soulevé (Wallace 1984 ; IFLA, 1992). Les études de Palmer (1972) ont montré que 5000 utilisateurs à l'université du Michigan ont utilisés, en moyenne, 4,5 éléments des 20 éléments d'un affichage : il a conclu que plusieurs des éléments pouvaient être supprimés.

Des études antérieures (citées par Palmer, 1967) ont soulevées aussi l'utilisation de champs, dans des formats catalographiques, qui n'étaient consultés qu'une fois sur dix...

Ces remarques que nous venons de reprendre ont été citées par deux professeurs de la faculté des sciences de l'information de l'université de Toronto, Lynne C. Howarth et Joseph P.Cox, dans un article qui fait la synthèse d'une étude de comportement d'utilisateurs devant des formats d'affichage bibliographiques. Des enregistrements de dix articles ont été rendus accessibles par internet, dans des OPACs de dix bibliothèques canadiennes (enregistrements conformes à l'AACR2). En même temps que la collecte des résultats, des interviews d'utilisateurs ont été organisés. Ces derniers ont été invités à classer les champs suivant leurs préférences. Ainsi a été mis en place un format d'affichage construit sur des préférences utilisateurs. L'article écrit, compare ce format avec les formats habituels et en tire des observations.

Nous venons de voir une autre façon d'associer la notion de profil utilisateur à un processus de recherche d'information, à travers cette fois-ci, un élément cadre qu'est le format bibliographique. Il faut noter que Lynne C. Howarth est spécialiste du catalogage (d'après ses écrits) d'où l'orientation de cette approche.

Cela nous intéresse pour le fait que cette dernière travaille actuellement sur l'association des metadata et du profil utilisateur, selon le point de vue des formats d'affichage également.

Les trois manières de tenir compte des préférences des utilisateurs dans le processus de recherche de l'information que nous venons de passer rapidement en revue nous font rappeler que les spécialistes de l'information ont pris conscience très tôt de cette nécessité. Certes, il est tout à fait trivial de reconnaître que le meilleur moyen pour récolter des documents pertinents, est de tenir compte du profil du demandeur, dans le processus de recherche de l'information ; mais il n'en demeure que les voies suivies jusqu'à maintenant n'ont apportées que des solutions partielles. Il est utile de rappeler que :

- ✓ La diffusion sélective de l'information tient compte du profil de l'utilisateur uniquement dans l'élaboration des requêtes ;
- ✓ Le projet NewsAgent , dans la mesure où il fonctionne sur le principe de la DSI, n'utilise lui aussi le profil utilisateur que pour affiner et automatiser son système de requêtes documentaires ;
- ✓ Associer des préférences utilisateurs à des formats d'affichage constitue une nouvelle étape dans cette approche, dans la mesure où l'étude rapproche des éléments de description de ces préférences.

Il est tout à fait logique que l'étape qui devait suivre, associerait complètement un profil utilisateur à des éléments de description : c'est l'objectif même de Profildoc. Ce dernier propose une caractérisation de l'utilisateur et une caractérisation de la ressource, avec un partitionnement en unités documentaires.

Dans ce mémoire de DEA, nous nous sommes attelés à proposer une structuration des éléments de description de Profildoc selon l'architecture du Dublin Core, dans un premier but de bénéficier des propriétés et développements d'un standard en pleine évolution, mais surtout parce que nous pensons que cela pourrait constituer les premières assises d'un système de description de ressources, prenant en compte une caractérisation de profil utilisateur.

Certes, associer dans une description de ressources, une caractérisation de ces dernières avec une caractérisation du profil de l'utilisateur entraînera nécessairement un système d'interrogation à relance. Dans un système de recherche d'information classique, l'interrogation est basée sur une élaboration d'une équation de recherche, qui permet d'avoir un premier résultat, lequel pourrait éventuellement être affiné par une deuxième requête ; mais cette dernière opération constitue une option non obligatoire. Dans ce que nous appelons un système d'interrogation à relance, nous supposons que la première interrogation ne constitue en elle-même qu'une ouverture de session interactive avec le système, avec qui s'ensuivra un dialogue. Nous imaginons que les relances permettraient de rentrer dans le système, la caractérisation de l'unité

documentaire recherchée, la caractérisation du document qui contient cette dernière, et la caractérisation de l'utilisateur. En Australie, un organisme de recherche, le RDN-CRC rattaché au DSTC (Distributed Systems Technology Center) ; a mis en place un outil de recherche d'information (hotmeta) qui utilise les interrogations à relance pour tenir compte des préférences de l'utilisateur, dans un processus de recherche d'information. Nous pensons explorer plus profondément cette voie, dans le cadre de notre thèse. Nous pensons aussi au PICS (Platform for Internet Content Selection), développé par le W3C, à l'origine dans un but de filtrage d'un certain type d'information (violence, pornographie) ; qui associe une fonction de description à une fonction de filtrage. Des auteurs (Stu Weibel et Eric Miller d'OCLC) ont déjà fait remarquer la nuance qu'il y a dans le type de description du système PICS, tenant compte de la sémantique du contenu ; et dans le type de description de contenu, réalisé avec des éléments de metadata. Des équipes de recherche tentent d'associer le PICS avec le Dublin Core pour des aspects spécifiques. Nous avons pensé explorer nous aussi une éventuelle association avec des éléments de description de metadata de Profildoc.

Evidemment, il faudrait d'abord voir dans quel cadre, un système conçu pour « éviter » l'affichage ou le télé-déchargement de certains types de documents, pourraient être utilisé pour en retrouver d'autres ? ...

9-CONCLUSION :

En abordant la préparation de la note de synthèse sur les standards de metadata, nous n'imaginions pas l'importance de ces dernières dans le contexte du traitement de l'information. C'est au fil de la consultation des documents y afférents, retrouvés pour la plupart sur Internet, que nous changeâmes peu à peu d'avis. Nous découvrîmes leurs applications à tous les domaines disciplinaires et leur nécessité absolue dans la description de l'information numérique. Les metadata prirent leur véritable essor avec le développement de grands standards fédéraux américains, tels le CSDGM (Content Standard Digital Geospatial Metadata) pour toute l'information géospatiale ; et le GILS (Government Information Locator Service) pour l'information administrative.

D'autres standards de metadata furent développés pour d'autres domaines d'application. Un standard s'est distingué de par sa simplicité et ses qualités ; et a, dès le départ, suscité l'engouement d'une multitude de communautés ou organismes, répartis de par le monde : le Dublin Core. Des communautés diverses se sont basées sur ce dernier, pour développer des standards conformes à leurs besoins. Nous avons alors entrepris d'approfondir nos connaissances sur celui-ci, en nous inscrivant à des listes de discussion très restreintes et à suivre de très près les nouveaux développements du Dublin Core. Nous voulions arriver ainsi à pouvoir deviner par anticipation l'évolution future du Dublin Core. Cela étant, nous entreprîmes de structurer les éléments de description de Profildoc selon l'architecture de Dublin Core ; afin de bénéficier de toutes ses propriétés, notamment l'extensibilité, l'interopérabilité et la lisibilité par machine. Nous avons ainsi structuré les éléments de description de Profildoc en éléments de metadata ; partageant, pour le moment, la même syntaxe html ; et pouvant être rattaché, dans le futur, aux mêmes développements du W3C que sont la RDF et XML ; lesquels sont en voie d'aboutissement.

Nous venons de schématiser très sommairement le travail réalisé dans le cadre du mémoire de DEA ; et nous voudrions maintenant situer le « background » de cette étude.

Le système Profildoc a été conçu avec l'objectif de diminuer le bruit et d'améliorer la pertinence des documents recueillis lors de recherches d'information dans des ressources orientées principalement IST (Information Scientifique et Technique). Basé sur le principe de la caractérisation de l'utilisateur, du document et de sa partie discrétionnaire appelée : unité documentaire ; il apporte l'originalité d'associer une description de contenu, à une définition de profil utilisateur. En réalité cette approche constitue une étape, suite logique de toutes les autres tentatives (des plus anciennes telle la DSI, aux plus récentes telles le projet NewsAgent, les formats d'affichage et même dans le principe, la technologie Push) et qui ont eu pour but d'améliorer, faciliter ou automatiser les recherches d'information. Il est admis depuis tout le temps que quelles que soient les performances du système de recherche d'informations utilisé, les résultats à une requête donnée seront d'autant plus pertinents qu'il est tenu compte du profil du demandeur...

Nous avons commencé par faire une étude rétrospective des principaux standards de metadata, dans le cadre d'une note de synthèse, pour voir comment chacun d'eux prend en charge une description de document ; tout en gardant en tête de fond, la manière dont Profildoc la prend en charge ; afin de retrouver d'éventuelles équivalences. Finalement nous avons opté pour structurer les éléments de ce dernier en éléments de metadata, selon l'architecture du standard qui nous a paru le plus proche et le plus ouvert : le Dublin Core. Ce deuxième travail rentre dans le cadre du mémoire de DEA ; et constitue pour nous une deuxième étape qui ouvre la voie à la prochaine ; car nous n'avons fait maintenant que définir les principes d'une description de ressources par des éléments de metadata de Profildoc qui, rappelons le, repose sur une caractérisation du document, une caractérisation des unités documentaires et une caractérisation de l'utilisateur. Il nous restera à étudier comment un système de recherche d'information pourrait prendre cela en charge. C'est la troisième et plus importante étape, qui constituera le cadre de notre thèse.

BIBLIOGRAPHIE :

1- Ouvrages :

- * V.S.SUBRAHMANIAN et SUSHIL JAJODIA, 'Multimedia database systems', Springer 1995, Berlin, Newyork.

2- Articles de periodiques:

- (1997) M. BEN ROMDHANE, S. LAINE-CRUZEL
Prise en compte de la structure des articles en sciences agronomiques pour la navigation dans un corpus scientifique électronique
Journées SFSIC-ENSSIB, "Une nouvelle donne pour les revues scientifiques", Villeurbanne 19-20 novembre 1997. (Résumé)
- (1997) M. BEN ROMDHANE, S. LAINE-CRUZEL
Poster : Réorganisation de l'information scientifique en vue d'une exploitation non linéaire sous format électronique : éléments de navigation dans un document scientifique électronique.
Premières journées du chapitre français de l'ISKO : "organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information", Lille 16-17 octobre 1997.
- (1997) C. MICHEL & T. LAFOUGE
Profil-doc : un système personnalisé de requête à des bases de données en texte intégral
Journées d'études sur les systèmes d'information élaborée, île Rousse, 12-16 Mai 1997
- (1997) N. BEN ABDALLAH, C. MICHEL, S. LAINE-CRUZEL
Caractérisation et découpage de textes scientifiques pour la construction de systèmes de requêtes personnalisées
Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la langue de l'AUPELF-UREF, Avignon 15-16 avril 1997
- (1997) T. LAFOUGE & S. LAINE-CRUZEL
Liens entre distribution d'usage et de structures des documents
Journées d'études sur les systèmes d'information élaborée, île Rousse, 12-16 Mai 1997

- (1997) T. LAFOUGE & S. LAINE-CRUZEL
A new explanation of the geometric law in the case of library circulation data *Information Processing & Management*, vol. 33, n° 4, Juillet 1997, pp. 523-527
- (1997) Roman S.Panchyshyn et France Bouthillier
Cataloguer le cyberspace :le defi des ressources electroniques
Documentation et bibliotheques, juillet/septembre 1997
- (1997) Desai, B.C.
Supporting discovery in virtual libraries
dans: "Journal of the American Society for Information Science", 48:3, pp.190-204
- (1997) Lynch, C.
Searching the Internet
In: "Scientific American", 276:3, pp.52-56
- (1997) Younger, J.A.
Resource description in the digital age
dans: "Library trends", 45:3, pp.462-487
- (1996) S. LAINE CRUZEL, T. LAFOUGE, J.P LARDY & N. BEN ABDALLAH.
Improving information retrieval by combining user profile and document unit.
Information Processing and Management, v. 32, n° 3, p. 305- 315.
- (1996) S. LAINE CRUZEL, T. LAFOUGE, C. MICHEL & N. BEN ABDALLAH.
Poster : Un système de requête à des bases de données en texte intégral.
Colloque "L'information scientifique et technique : nouveaux enjeux documentaires et éditoriaux", INRA, Tours, 21-23 octobre 1996.
- (1996) Jan Smits
Digital Metadata,Standards for Communication and Preservation
European Research Libraries Cooperation :The Liber Quaterly
6(1996),383-406
- (1996) Dempsey, L. et Weibel, S.L.
The Warwick Metadata Workshop: A framework for the deployment of resource description
dans: "D-lib Magazine", July/August 1996.
- (1996) Dempsey, L.
ROADS to Desire: Some UK and Other European Metadata and Resource Discovery Projects dans: "D-lib Magazine", July/August 1996.

- (1996) Smith, Terence R.
The Meta-information Environment of Digital libraries
dans: "D-lib Magazine", July/August 1996.
- (1996) Heery, Rachel
Review of Metadata Formats
dans: "Program" 30:4, pp.345-373. October 1996
- (1996) Weibel, S. et al
Advances in metadata. (1996),
dans: "D-lib Magazine, July/August 1996"
- (1995) Weibel, Stuart
Metadata: The Foundations of Resource Description
dans: "D-Lib Magazine", July 1995
- (1995) Mangan Elizabeth U.
The making of a standard
Information technology and libraries, 2(14), 1995
- (1995) Aldous Kj
A system for the automatic retrieval of information from a specialist database
Information processing & management, 2(32), 1996, 0306-4573
- (1995) Dempsey L.
RADAR reflexions :internet ressource,access,discovery and retrieval systems and libraries.
Library networking in Europe.
Conference europeenne/Bruxelles/12-10-1994, Londres :TFPL Publishing 1995.
- (1994) S. LAINE-CRUZEL
Vers de nouveaux systèmes d'information prenant en compte le profil des utilisateurs
Documentaliste-sciences de l'information, 1994, vol. 31, n° 3, pp. 143-147.
- (1994) Riekert Wf,
Management of data and services for environmental applications.
Environmental knowledge organisation and information management :
Bratislava, 14-16 Septembre 1994
Knowledge Organisation in Subject Areas, DEU 1994

3- Documents retrouvés sur Internet :

3.1- Documents généraux:

- [1] Distributed Systems Technology Centre.
Resource Discovery Unit (RDU).
<http://www.dstc.edu.au/RDU/>
- [2] Web Developers Virtual Library.
META Tagging for Search Engines.
<http://WWW.Stars.com/Search/Meta/Tag.html>
- [3] Ahronheim, Judy.
Judy and Magda's List of Metadata Initiatives.
<http://www-personal.umich.edu/~jaheim/alcts/bibaccess.htm>
- [4] Cameron, Robert D.
Towards Universal Serial Item Names.
Rapport technique 97-16, Ecole d'informatique, Simon Fraser University, 3 Decembre, 1997.
<http://elib.cs.sfu.ca/USIN/USIN.html>
- [5] Dempsey, Lorcan.
ROADS to Desire: Some UK and Other European Metadata and Resource Discovery Projects.
D-Lib Magazine, Juillet/Aout 1996.
<http://www.dlib.org/>
- [6] Lynch, Clifford.
Searching the Internet.
Scientific American, Mars 1997
<http://www.sciam.com/0397issue/0397lynch.html>
- [7] IEEE.
The Metadata and Data Management Information Page.
http://www.llnl.gov/liv_comp/metadata/metadata.html
- [8] W3 Consortium.
WWW Names and Addresses, URIs, URLs, URNs, URCs.
<http://www.w3.org/hypertext/WWW/Addressing/Addressing.html>
- [9] UKOLN.
Metadata.
<http://www.ukoln.ac.uk/metadata/>
- [10] UKOLN.
Metadata: Mapping between metadata formats.
<http://www.ukoln.ac.uk/metadata/interoperability/>
- [11] EPA (Environmental Protection Agency)

Scientific Metadata Standards Project.
<http://www.lbl.gov/~olken/epa.html>

3.2- DUBLIN CORE :

- [12] *DC-5 Dublin Core Metadata Workshop*
<http://linnea.helsinki.fi/meta/DC5.html>
- [13] Miller, Paul and Tony Gill.
DC5: The Search for Santa.
Ariadne, Issue 12, Novembre 1997.
<http://www.ariadne.ac.uk/issue12/metadata/>
- [14] Weibel, Stu, et. al.
The 4th Dublin Core Metadata Workshop Report.
D-Lib Magazine, Juin 1997
<http://www.dlib.org/dlib/june97/metadata/06weibel.html>
- [15] DC-4: *NLA/DSTC/OCLC Dublin Core Down Under /
The 4th Dublin Core Metadata Workshop.*
<http://www.dstc.edu.au/DC4/>
- [16] Heery, R., et. al.
*The 4th Dublin Core Workshop: Notes from UK
participants.*
[http://www.ukoln.ac.uk/metadata/resources/dc4-
notes.html](http://www.ukoln.ac.uk/metadata/resources/dc4-notes.html)
- [17] Miller, P. and Gill, T.
Down Under with the Dublin Core.
Ariadne. Avril, 1997.
[http://www.ukoln.ac.uk/ariadne/issue8/canberra-
metadata/](http://www.ukoln.ac.uk/ariadne/issue8/canberra-metadata/)
- [18] *CNI/OCLC Metadata Workshop: Workshop on Metadata
for Networked Images.* (24-25 Septembre, 1996)
<http://purl.oclc.org/metadata/image>
- [19] Weibel, Stuart and Eric Miller.
*"Image Description on the Internet: A Summary of
the CNI/OCLC Image Metadata Workshop."*
D-Lib Magazine. Janvier 1997.
<http://www.dlib.org/dlib/january97/oclc/01weibel.html>
- [20] Dempsey, Lorcan and Weibel, Stuart L.
*The Warwick Metadata Workshop: A Framework for the
Deployment of Resource Description.*
D-Lib Magazine, Juillet/Aout 1996.
<http://www.dlib.org/dlib/july96/07weibel.html>

- [21] Burnard, L., et. al.
A Syntax for Dublin Core Metadata: Recommendations from the Second Metadata Workshop. (Avril 1996)
<http://www.uic.edu/~cmsmcq/tech/metadata.syntax.html>
- [22] Hakala, Juha H., et. al. (1996)
Warwick framework and Dublin core set provide a comprehensive infrastructure for network resource description.
<http://www.ub2.lu.se/tk/warwick.html>
- [23] **OCLC/NCSA Metadata Workshop Report.**
URL: http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html
- [24] **OCLC/NCSA Metadata Workshop: The Essential Elements of Network Object Description.** (1995)
<http://www.oclc.org:5046/conferences/metadata/metadata.html>
- [25] **Meta2 Mailing List Archive.**
<http://weeble.lut.ac.uk/lists/meta2/>
- [26] Weibel, S., Kunze, J. and Lagoze, C.
Dublin Core Metadata for Simple Resource Description.
<ftp://ds.internic.net/internet-drafts/draft-kunze-dc-01.txt>
- [27] **Dublin Core Metadata.**
http://purl.org/metadata/dublin_core
- [28] **Dublin Core Metadata Element Set: Reference Description.**
http://purl.org/metadata/dublin_core_elements
- [29] Baker, Thomas.
Metadata Semantics Shared Across Languages: Dublin Cores in languages other than English
<http://www.cs.ait.ac.th/~tbaker/Cores.html>
- [30] Beckett, Dave.
Proposed Encodings for Dublin Core Metadata.
<http://www.hensa.ac.uk/pub/metadata/dc-encoding.html>
- [31] Gill, Tony, Grout, Catherine and Smith, Louise.
Visual Arts, Museums and Cultural Heritage Information Standards: a domain specific review of relevant standards for networked information discovery.
<http://vads.ahds.ac.uk/standards.html>

- [32]Grout, Catherine and Tony Gill.
Visual Arts, Museums & Cultural Heritage Metadata Workshop Report.
<http://vads.ahds.ac.uk/Metadata1.html>
- [33]Guenther, Rebecca.
Dublin Core Qualifiers/Substructure : a proposal.
<http://www.loc.gov/marc/dcqualif.html>
- [34]Hakala, Juha.
Dublin Core Metadata Element Set and it's applications.
<http://linnea.helsinki.fi/meta/present.html>
- [35]Knight, Jon and Martin Hamilton.
Dublin Core Standard Resource Types.
<http://www.roads.lut.ac.uk/Metadata/DC-ObjectTypes.html>
- [36]Knight, Jon and Martin Hamilton.
Dublin Core Qualifiers.
<http://www.roads.lut.ac.uk/Metadata/DC-SubElements.html>
- [37]Kunze, John.
Metadata User Guide group report.
<http://weeble.lut.ac.uk/lists/meta2/0220.html>
- [38]Lagoze, Carl.
The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. Juillet/Aout 1996.
<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>
- [39]Lagoze, Carl, Lynch, Clifford A., and Daniel, RonJr.
The Warwick Framework:A Container Architecture for Aggregating Sets of Metadata. TR96-1593, 21 Juin, 1996.
<http://www.ifla.org/documents/libraries/cataloging/metadata/tr961593.pdf>
- [40]Lasher, R.
Dublin Core Bibliography. TR96-1593, 21 Juin, 1996.
<http://www-diglib.stanford.edu/diglib/pub/dublin.html>
- [41]Library of Congress.
Metadata, Dublin Core and USMARC: A Review of Current Efforts.
<gopher://marvel.loc.gov/00/.listarch/usmarc/dp99.doc>
- [42]Miller, Eric.
An Approach for Packaging Dublin Core Metadata in HTML2.0. Aout 1996.
<http://www.oclc.org:5046/~emiller/publications/metadata/minimal.html>

- [43] Miller, Paul.
An application of Dublin Core from the Archaeology Data Service.
10 Juillet 1996.
<http://www.ncl.ac.uk/~napml/ads/metadata.html>
- [44] Miller, Paul and Daniel Greenstein.
Discovering Online Resources Across the Humanities: A Practical Application of the Dublin Core; Oct. 1997
<http://ahds.ac.uk/public/metadata/discovery.html>
- [45] Miller, Paul.
Metadata for the masses.
Ariadne, Issue 5, Septembre 1996.
<http://www.ukoln.ac.uk/ariadne/issue5/metadata-masses/>
- [46] Nordic Metadata Project
Dublin Core Metadata Template.
http://www.ub.lu.se/metadata/DC_creator.html
- [47] Powell, Andy.
Dublin Core Management; Ariadne. Juillet 1977.
<http://www.ariadne.ac.uk/issue10/dublin/>
- [48] Smith, Terence R.
The Meta-Information Environment of Digital Libraries.
D-Lib Magazine, Juillet/Aout 1996.
<http://www.dlib.org/dlib/july96/new/07smith.html>
- [49] Sperberg-McQueen, C.M. (Avril 1996)
On Information Factoring in Dublin Metadata Records.
<http://www.uic.edu/~cmsmq/tech/metadata.factoring.html>
- [50] Tennant, Roy.
Dublin Core Resource Types.
<http://sunsite.berkeley.edu/Metadata/types.html>
- [51] UKOLN.
Metadata Resources - Dublin Core.
<http://www.ukoln.ac.uk/metadata/resources/dc.html>
- [52] USMARC Advisory Group.
Mapping the Dublin Core Metadata Elements to USMARC.
<http://lcweb.loc.gov/marc/dccross.html>
- [53] Warwick, Cathro. (Aout 1997)
Metadata: An Overview.
<http://www.nla.gov.au/nla/staffpaper/cathro3.html>
- [54] Weibel, Stuart.
A Proposed Convention for Embedding Metadata in HTML.

[http://www.oclc.org:5046/~weibel/html- meta.html](http://www.oclc.org:5046/~weibel/html-meta.html)

3.3- Projets sur le Dublin Core :

Australie

[55] ***DSTC/Resource Discovery Unit.***

<http://www.dstc.edu.au/RDU/>

[56] ***Australian Geodynamics Cooperative Research Centre (AGCRC) .***

<http://www.agcrc.csiro.au/>

Canada :

[57] ***searchBC: Vancouver Webpages.***

<http://vancouver-webpages.com/VWbot/searchBC.html>

Allemagne :

[58] ***Metadaten-Projekt = Metadata Project.***

<http://www2.sub.uni-goettingen.de>

[59] ***SSG-Fachinformation (SSG-FI) Mathematick = Subject Area Information for Mathematics.***

<http://www.sub.uni-goettingen.de/ssgfi/>

[60] ***The German Educational Resources Server / Deutscher Bildungs-Server.***

<http://dbs.schule.de/indexe.html>

[61] ***Math-Net.***

<http://elib.zib.de/math-net/>

[62] ***Electronic Information Management and Metadata in Physics.***

<http://www.physik.uni-oldenburg.de/EPS/EurophysNet/PhysDep/dep-links.html>

Les Pays-Bas :

[63] ***Koninklijke Bibliotheek/ The National Library of the Netherlands.***

<http://www.konbib.nl:8000/>

Scandinavie :

[64] ***The Nordic Metadata Project.***

<http://linnea.helsinki.fi/meta/>

Suede :

[65] ***Swedish EnviroNet.***

<http://smn.environ.se/smnproj/proj/summary.htm>

Danemark :

- [66] **Netpublikationer.**
<http://www.fsk.dk/fsk/publ/online-pub/>

Royaume-Uni :

- [67] **Art, Design, Architecture & Media Information Gateway and the Visual Arts Data Service.**
<http://adam.ac.uk/>
<http://vads.ahds.ac.uk/>
- [68] **AHDS Arts & Humanities Data Service.**
<http://ahds.ac.uk/>
- [69] **Project BIBLINK.**
<http://www.ukoln.ac.uk/metadata/BIBLINK/>
- [70] **Project DESIRE.**
<http://www.nic.surfnet.nl/surfnet/projects/desire/desire.html>
- [71] **SCRAN (Scottish Cultural Resources Access Network).**
<http://www.scran.ac.uk>
- [72] **NewsAgent for Libraries.**
<http://www.sbu.ac.uk/litc/newsagent/>
- [73] **Electronic Library Image Service for Europe (ELISEII)**
<http://severn.dmu.ac.uk/elise/>
- [74] **Resource for Urban Design Information.**
<http://rudi.herts.ac.uk/catsrch/catsrch.html>
- Etats-Unis :**
- [75] **Monticello Electronic Library.**
<http://www.solinet.net/monticello/monticel.htm>
- [76] **Medical Metadata Project.**
<http://medir.ohsu.edu/~maletg/MedMetadata.HTM>
- [77] **Florida International University Digital Library.**
<http://www.fiu.edu/~diglib/>
- [78] **Internet Scout Project's Signpost.**
<http://www.signpost.org/signpost/index.html>
- [79] **University of Washington Digital Library.**
<http://content.engr.washington.edu/>

[80] *Everglades Information Network & Digital Library.*
<http://everglades.fiu.edu/>

[81] *University of Michigan Digital Library Registry Database.*
<http://dns.hti.umich.edu/registry/>

[82] *Digital Library Catalog.*
<http://sunsite.berkeley.edu/Catalog>

3.4-ENCODED ARCHIVAL DESCRIPTION (EAD):

[83] *Encoded Archival Description (EAD) DTD.*
<http://lcweb.loc.gov/loc/standards/ead/>

[84] *Development of the Encoded Archival Description Document. Type Definition.*
<http://lcweb.loc.gov/loc/standards/ead/eadback.html>

[85] *Finding Aids for Archival Collections.*
<http://sunsite.berkeley.edu/FindingAids/>

[86] *Berkeley Finding Aids Conference.*
4-6 Avril, 1995.
<http://sunsite.berkeley.edu/FindingAids/EAD/bfac.html>

[87] *Society of American Archivists. Committee on Archival Information Exchange. Encoded Archival Description Working Group.*
<http://sunsite.berkeley.edu/FindingAids/EAD/eadwg.html>

3.5-STANDARDS DE METADATA GEOSPATIALS :

[88] American Society for Testing and Materials.
ASTM Section D18.01.05 Draft Specification Content Specification for Digital Geospatial Metadata.
<http://info.er.usgs.gov/research/gis/standard/index.htm>

[89] *ERIN Directories and Metadata.*
<http://kaos.erin.gov.au/technical/retrieval/directory/directory.html>

[90] *ERIN Australian Glossary of Geographic Information Systems and Metadata Terms.*
http://kaos.erin.gov.au/gis/gis_gloss.html

[91] *Federal Geographic Data Committee (FGDC).*
<http://fgdc.er.usgs.gov/fgdc2.html>

- [92]Federal Geographic Data Committee.
Metadata Standards Development.
<http://www.fgdc.gov/Metadata/metahome.html>
- [93]Federal Geographic Data Committee.
**Frequently Asked Questions concerning the FGDC's
Content Standard for Geospatial Metadata.**
<http://www.its.nbs.gov/nbs/meta/faq.htm>
- [94]**Directory Interchange Format (DIF) Writer's Guide.**
<http://gcmd.gsfc.nasa.gov/difguide/difman.html>
- [95]**Global Change Master Directory (GCMD).**
<http://gcmd.gsfc.nasa.gov/difguide/difman.html>
- [96]**MetaData and WWW Mapping Home Page.**
<http://www.blm.gov/gis/nsdi.html>

3.6-GOVERNMENT INFORMATION LOCATOR SERVICE (GILS):

- [97]U.S. Geological Survey.
Government Information Locator Service.
<http://www.usgs.gov/public/gils/>
- [98]GILS Forum Archives. (Depuis Aout 1994)
<gopher://gopher.cni.org:70/11/cniftp/pub/forums/gils>
- [99]**Application Profile for GILS - Version 2.**
http://www.usgs.gov/gils/prof_v2.html
- [100]National Archives and Records Administration(NARA)
Guidelines for the Preparation of GILS Core Entries.
<http://www.ifla.org/documents/libraries/cataloging/metadata/naragils.doc>
<http://www.dtic.dla.mil/gils/documents/naradoc/>
- [101]Moen, William and McClure, Charles.
**An Evaluation of the Federal Government's
Implementation of the Government Information Locator
Service.** (Préparé par les Services Généraux de
l'administration U.S. June 30, 1997.
<http://www.unt.edu/slis/research/gilseval/gilseval.htm>
- [102]Turner, Fay.
The U.S. Government Information Locator Service:
Description and Status.
http://gils.gc.ca/gils/backg_e.html
- [103]U.S. Defense Information Technical Center.

Technology Transfer Center GILS Toolbox.
<http://skydive.ncsa.uiuc.edu/toolbox/>

3.7- Implémentations internationale de GILS :

Australia:

- [104] *Information Management Steering Committee (IMSC).
Architecture For Access To Government Information.*
<http://www.adfa.oz.au/DOD/imsc/imstctg/imstctgla.htm>

Canada :

- [105] *Government Information Locator Service.*
<http://gils.gc.ca>
- [106] *Guidelines for the Preparation of GILS Records.*
http://gils.gc.ca/gils/guide_e.html
- [107] *Creating GILS Records in an SGML Environment.*
http://gils.gc.ca/gils/creatingg_e.html

3.8- IAFA/WHOIS++ TEMPLATES :

- [108] Beckett, David.
IAFA Templates in use as Internet Metadata.
<http://www.w3.org/pub/Conferences/WWW4/Papers/52/>
<http://www.hensa.ac.uk/tools/www/iafatools/slides/>
- [109] *Field Descriptions for Document, Software, Image,
Sound, Video, Mailarchive, USENET and FAQ IAFA
Template Types.*
[http://www.man.ac.uk/MVC//SIMA/MMFFDB/IAFA-
help/document.html](http://www.man.ac.uk/MVC//SIMA/MMFFDB/IAFA-help/document.html)
- [110] ROADS.
*Resource Organisation And Discovery in Subject-
based services.*
<http://ukoln.bath.ac.uk/roads/>
- [112] Heery, R. ROADS:
*Resource Organisation and Discovery in Subject-
based Services.* Ariadne, No. 3, 1996.
<http://www.ukoln.ac.uk/ariadne/issue3/roads/>
- [113] Heery, R. (September 1996)
ROADS templates: how they are used.
<http://www.ukoln.ac.uk/metadata/templates.html>
- [114] Hiom, D., Dempsey, L. and Norman, F.

Road to resource discovery.

http://ukoln.bath.ac.uk/roads/lar_arti.html

3.9- MARC (MACHINE-READABLE CATALOGUE) :

[115]Library of Congress.

Library of Congress MARC Office.

<http://lcweb.loc.gov/marc/marc.html>

[116]Library of Congress.

Machine-readable cataloging (MARC).

<http://lcweb.loc.gov/marc/>

[117]Library of Congress.

USMARC formats and documentation.

<gopher://marvel.loc.gov:70/11/services/usmarc>

[118]OCLC.

Bibliographic formats and standards.

<http://www.oclc.org/oclc/bib/about.htm>

3.10-META CONTENT FORMAT (MCF) :

[119]Guha, R.V.

Meta Content Framework. An overview of MCF.

<http://mcf.research.apple.com/hs/mcf.html>

[120]MCF Tutorial.

<http://www.textuality.com/mcf/MCF-tutorial.html>

[121] *Meta Content Framework Using XML.*

<http://www.textuality.com/mcf/NOTE-MCF-XML.html>

3.11-PLATFORM FOR INTERNET CONTENT SELECTION (PICS) :

122 Armstrong, Chris. (1997)

Metadata, PICS and Quality.

<http://www.ariadne.ac.uk/issue9/pics/>

[123]W3 Consortium.

PICS.

<http://www.w3.org/pub/WWW/PICS/>

[124]Resnick, Paul.

Filtering Information on the Internet.

<http://www.sciam.com/0397issue/0397resnick.html>

[125]Resnick, Paul and James Miller. (Octobre 1996)

PICS: Internet Access Controls Without Censorship.
<http://www.w3.org/pub/WWW/PICS/iacwcv2.htm>

[126] Miller, James S.
W3C and Digital Libraries.
Dlib Magazine, November 1996.
<http://www.ukoln.ac.uk/dlib/dlib/november96/11miller.html>

[127] PICS-SE -
A Proposed Standard for the Annotation of Internet Documents using a String Extension to PICS.
<http://www.kulturbox.de/aid/pics-se/>

3.12-HARVEST SUMMARY OBJECT INTERCHANGE FORMAT (SOIF):

[128] **The Harvest Home Page.**
<http://harvest.transarc.com/>

[129] **The Harvest Summary Object Interchange Format**
<http://harvest.transarc.com/Harvest/brokers/soifhelp.html>

[130] **Harvest User's Manual.**
<http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/doc.html>

[131] University of Colorado.
The Harvest Information Discovery and Access System.
<http://harvest.transarc.com/>

3.13-TEXT ENCODING INITIATIVE (TEI):

[132] **The TEI Header.**
<http://etext.virginia.edu/bin/tei-tocs?div=DIV1&id=HD>

[133] Burnard, Lou. (Juillet 1995)
Text encoding for information interchange: an introduction to the Text Encoding Initiative.
<ftp://info.ox.ac.uk:80/~archive/teij31/WHAT.html>

[134] Burnard, Lou and C. M. Sperberg-McQueen.
TEI Lite: An Introduction to Text Encoding for interchange.
Document No: TEI U 5, Juin 1995.
<http://sable.ox.ac.uk/ota/teilight/>

[135] Sperberg-McQueen, C.M. and Lou Burnard.

Guidelines for Electronic Text Encoding and Interchange.

Chicago / Oxford, 1994.

<http://etext.lib.virginia.edu/TEI.html>

3.14-AUTRES DOCUMENTS :

[136] Bearman, David and Sochats, Ken.

Metadata Requirements for Evidence.

<http://www.lis.pitt.edu/~nhprc/BACartic.html>

[137] Berners-Lee, Tim.

Document Naming.

<http://www.w3.org/pub/WWW/DesignIssues/Naming.html>

[138] Burnard, Lou and Richard Light.

Three SGML metadata formats: TEI, EAD, and CIMI .

BIBLINK Work Package 1.1. December 1996.

<http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/>

[139] CNIDR.

The Scientific and Technical Attribute Set (STAS).

<http://stas.cnidr.org/STAS.html>

[140] Daniel, Jr., Ron.

A Global Distributed Directory Service for the World Wide Web.

http://www.acl.lanl.gov/URI/URC_proposal/index.html

[141] Day, Michael. (1997)

Extending metadata for digital preservation.

<http://www.ariadne.ac.uk/issue9/metadata/>

[142] Dempsey, L. (1996)

Meta Detectors.

<http://www.ukoln.ac.uk/ariadne/issue3/metadata/>

[143] Dempsey, L. and Heery, R.

A review of metadata: a survey of current resource description formats.

<http://www.ukoln.ac.uk/metadata/DESIRE/overview/>

[144] Dempsey, L. (1996)

ROADS to Desire: some UK and other European metadata and resource discovery projects.

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

[145] Desai, Bipin C.

The Semantic Header and Indexing and Searching on the Internet.

[http://www.cs.concordia.ca/~faculty/bcdesai/
cindi-system-1.0.html](http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html)

[146]Green, Ann Gerken, and Dionne JoAnn. (20 Dec 1996)
***Preserving the Whole: A two-track Approach to
Rescuing Data and Metadata.***

<http://www-cpa.stanford.edu/cpa/misc/preswhol.html>

[147]Heery, R. (Juillet 1996)
***Resource description: initial recommendations for
metadata formats.***

<http://www.ukoln.ac.uk/metadata/DESIRE/recommendations/>

[148]Heery, R.
Metadata Formats.

<http://www.ukoln.ac.uk/metadata/review.html>

[149]Heery, R.
Review of metadata formats.

<http://www.ukoln.ac.uk/metadata/review.html>

[150]Hockey, Susan.
***Describing Electronic Texts: The Text Encoding
Initiative and SGML.***

<http://lcweb.loc.gov/catdir/semdigdocs/hockey.html>

[151]Iannella, Renato and Waugh, Andrew.
Metadata: Enabling the Internet.

<http://www.dstc.edu.au/RDU/reports/CAUSE97/>

[152]Koch, Traugott and Day, Michael.
***The role of classification schemes in Internet
resource description and discovery .***

Projet DESIRE, February 1997.

<http://www.ukoln.ac.uk/metadata/DESIRE/classification/>

[153]Lasher, Rebecca and Cohen, Danny.
A Format for Bibliographic Records.

RFC 1807. June 1995.

<http://www.dsg.cs.tcd.ie:1995/rfc1807.html>

[154]Lynch, Clifford, et. al.
***CNI White Paper on Networked Information Discovery
and Retrieval.***

<http://www.cni.org/projects/nidr/www/toc.html>

[155]Lynch, Clifford. (March 1997)
Searching the Internet.

<http://www.sciam.com/0397issue/0397lynch.html>

[156]Miller, Eric J.

Issues of Document Description in HTML.

<http://www.oclc.org:5046/~emiller/tmp/issues.html>

[157] OCLC.

Bibliographic Formats and Standards.

<http://www.oclc.org/oclc/bib/about.htm>

3.15-CONFERENCES ET WORKSHOPS :

[158] ***Second IEEE Metadata Conference.***

Sep 16-17, 1997: NOAA, Silver Spring MD.

http://www.llnl.gov/liv_comp/metadata/md97.html

Proceedings :

<http://computer.org/conferen/proceed/meta97/>

[159] Research Libraries Group.

Metadata Summit.

July 1, 1997: Mountain View, California.

<http://www.rlg.org/meta9707.html>

[160] ***Joint Workshop on Metadata Registries.***

July 8-10, 1997: Universite de Californie, Berkeley,

<http://www.lbl.gov/~olken/EPA/Workshop/index.html>

Rapport du Workshop :

<http://www.lbl.gov/~olken/EPA/Workshop/report.html>

[161] ***Report of the Joint Workshop on Standards for the Use of Models that Define the Data and Processes of Information Systems.***

September 9-12, 1996: Bellevue, Washington.

<http://www.mel.nist.gov/workshop/jtc1-96/report.htm>

3.16-EXEMPLES D'OUTILS DE METADATA :

[162] Miller, P. and McDonald, T.

The Dublin Core Metadata creator.

http://www.ncl.ac.uk/~napml/dublin_core/

[163] UKOLN.

DC-dot.Dublin Core Generator.

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

[164] UKOLN.

Metadata Software Tools.

<http://www.ukoln.ac.uk/metadata/software-tools/>

[165] NASA. Global Change Master Directory.

***Web-based DIF and Supplementary Information
Editing Tools.***

<http://gcmd.gsfc.nasa.gov/difweb.html>

[166] Nordic Metadata Project.

Dublin Core Metadata Template.

http://www.ub.lu.se/metadata/DC_creator.html

[167] OCLC.

Dublin Core Metadata System.

<http://orc.rsch.oclc.org:9016/dcms/index.html>

[169] SafeSurf.

Meta Generator.

<http://www.safesurf.com/classify/index.html>

[170] Vancouver Webpages.

META builder.

<http://vancouver-webpages.com/VWbot/mk-metas.html>

[171] Vancouver Webpages.

***A script to build META tags using the Dublin Core
types.***

<http://vancouver-webpages.com/VWbot/mk-dublin.html>