



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 053-F
Meeting: 157. Bibliographic Control - **Workshop**
Simultaneous Interpretation: -

Schémas de métadonnées pour les répertoires par sujet

Lynne C. Howarth

Faculty of Information Studies, University of Toronto
Toronto, CANADA
howarth@fis.utoronto.ca

Résumé :

Les répertoires par sujet sur le Web donnent accès aux collections de haute qualité pour un sujet précis. Les descriptions de documents, d'objets et de sites choisis avec soin et évalués, identifient de façon unique et relient au contenu numérique. Comme cette communication l'examinera, la définition de la cible et du contenu correspondant pour un répertoire par sujet influencera la sélection ultérieure des structures de métadonnées et des schémas. Certaines autres considérations reliées à la granularité des métadonnées et à l'interopérabilité seront également abordées. Une discussion conclura sur l'analyse des défis et tendances futurs reliés à l'utilisation des schémas de métadonnées dans les répertoires par sujet.

1.0 Introduction

Comme une simple recherche à l'aide de Google™ le confirmera, les définitions opérationnelles de l'expression « répertoires par sujet » (*subject gateways*) sont presque aussi nombreuses que les services qu'elles décrivent. Cependant, reconnaissant le besoin d'un point de départ, Traugott Koch (2000, 2) propose la définition suivante comme approche de la définition du concept :

Les répertoires par sujet sont des services Internet supportant le repérage systématique de ressources. Ils fournissent des liens vers les ressources (documents, objets, sites de services), accessibles d'une manière prédominante via Internet. Le service est basé sur la description des ressources. L'accès aux ressources par furetage à l'aide de la structure par sujet est une caractéristique importante.

Koch (2000,1) note plus loin que « Un effort manuel considérable est entrepris pour acquérir une sélection de ressources qui rencontrent des critères de qualité et pour démontrer une description riche de ces ressources à l'aide de métadonnées de base. » Avec cette explication supplémentaire du procédé, le lien entre les concepts de « répertoires par sujet » et « schémas de métadonnées » est éclairci et les assises sont posées pour l'exploration de la manière dont ces derniers peuvent être utilisés de façon la plus efficace possible pour la création et le support des répertoires. Les défis particuliers et les tendances

futures associés à l'application des métadonnées dans les répertoires par sujet seront également abordés plus loin dans cette communication.

2.0 Considérations et applications

En général, une distinction doit être faite entre le format simple de métadonnées, tel que représenté par la syntaxe d'un langage de balisage (par exemple : XML, HTML, SGML) et intégré à l'intérieur de la structure de l'objet numérique, et entre les formats riches de métadonnées. Dans le premier cas, les inforobots Web ou « robot » peuvent amasser des balises spécifiques de métadonnées (par exemple : <Title>) pour extraire des valeurs particulières et, avec l'aide d'une intervention humaine minimum (ou sans aide), présenter les résultats à l'intérieur d'index inversés prédéfinis ou de répertoires par sujet. Les catégories de sujets proposées dans les sites de moteurs de recherches comme Yahoo!™, AltaVista™ ou Google™ sont des exemples bien connus. Les problèmes qui peuvent survenir avec les langages de vocabulaires naturels, avec les textes non structurés et avec ce que la communauté de contrôle bibliographique reconnaît comme une lacune du « contrôle d'autorité », sont endémiques dans l'environnement des métadonnées « format simple ». Cependant, la résistance à la structure, aussi bien imposée que volontairement adoptée par les méthodes libres d'Internet, continue de favoriser les initiatives décisionnelles telle que le *Open Directory Project* où les créateurs des ressources peuvent assigner une terminologie de sujets qu'ils ont conçue indépendamment de tout standard (voir : <http://www.dmoz.org>).

Par opposition, les formats riches de métadonnées sont conçus, appliqués et gérés conformément aux standards (internationaux) clairement établis. Ce sont les schémas formels de métadonnées qui sont utilisés dans le « contrôle de qualité » des répertoires par sujet. En créant des répertoires par sujet de « qualité contrôlée » à l'aide de métadonnées, trois aspects-clés doivent être pris en considération, c'est-à-dire les ressources (documents, objets, sites), y compris quels genres de structures de métadonnées devraient être utilisées pour décrire et permettre l'accès à ces ressources et quel schéma de métadonnées doit être appliqué lors de la création d'enregistrements reliant au contenu. Chacun de ces aspects sera examiné, à son tour, dans les sections qui suivent.

2.1 Définition de la cible et du contenu d'un répertoire par sujet

Déterminer la cible et le contenu correspondant d'un répertoire par sujet influencera nécessairement la sélection ultérieure des *structures* et *schémas* de métadonnées.

Tableau 1
Critère d'inclusion pour déterminer la portée et le contenu d'un répertoire par sujet

| Critère d'inclusion | Simple (exemples) | Multiple (exemples) | Universel (exemples) |
|---------------------------------|--------------------------|--|---|
| Sujet/thème | Leucémie | Maladies cardiaques et neurologiques | Toutes les maladies |
| Langue | Anglais | Allemand, Français, Japonais, Grec | Toutes les langues |
| Situation géographique | Canada | Europe et Asie | Tous les pays, régions, etc. |
| Limitation chronologique | 2003 | 1900-1999 | Toutes les périodes historiques enregistrées |
| Type de ressource | Sites Web | Sites Web, archives de données et archives photographiques | Toutes les ressources du Web |
| Groupes/Associations | Femmes | Enfants et jeunes adultes | Humanité |
| Format matériel | Texte électronique | Documents Word, cartes numériques, DVD | Tous les formats textuels et médias (analogique et numérique) |

Quel est le véritable dessein d'un répertoire par sujet, quels objectifs spécifiques doivent être atteints dans sa conception et quels sont les livrables ou résultats pouvant lui être attribués ? La portée et la couverture, les délimitations et limitations du répertoire par sujet exigent des critères tels que le sujet ou thème, la langue, la situation géographique, la limitation chronologique, le type de ressources, les groupes ou associations, le format matériel, etc. Le tableau 1 résume, avec des exemples, comment chacun de ces critères peut être combiné ou évalué pour déterminer la conception finale du répertoire par sujet. Il faut noter ici que de nombreuses approches peuvent être utilisées et que ce qui vient d'être énuméré n'offre qu'une seule des illustrations possibles où plusieurs permutations ou combinaisons peuvent s'appliquer.

La définition d'«un», «plusieurs» ou «tous» de chaque critère énuméré dans la première colonne est également relative et clairement ouverte à interprétation.

Une fois ces questions posées, les critères de sélection des ressources peuvent être déterminés et les sources ou cibles de ces ressources peuvent être identifiées. Koch (2000,6) suggère que les modèles les plus fréquents pour les répertoires incluent :

- Sujet spécifique national (un sujet; un pays; une langue – par exemple GEM)
- Sujet croisé national (sujets multiples; un pays; une langue – par exemple DutchESS)
- Sujet spécifique mondial (un sujet; mondial; une langue – par exemple EEVL)
- Sujet croisé mondial (sujets multiples; mondial; une langue – par exemple ADAM)
- Universel (tous les sujets; mondial; plusieurs langues – par exemple CORC)

On pourrait même ajouter que la typologie précédente n'est ni définitive ni exhaustive et, à l'instar du tableau 1, identifie un point de vue supplémentaire ou une perspective des façons possibles qui existent de combiner les critères de sélection pour concevoir les répertoires par sujet basés sur les cibles proposées ou sur les objectifs spécifiques devant être atteints.

2.2 Déterminer les structures de métadonnées applicables

Comme il a été mentionné précédemment, la définition de la portée, de la couverture et de la sélection des critères de contenu d'un répertoire par sujet influencera quels éléments de métadonnées et quels schémas seront choisis pour supporter l'identification et les liens vers les ressources cibles. En règle générale, les *types* ou *structures* de métadonnées qui doivent être nécessaires pour supporter un répertoire par sujet prennent modèle sur n'importe quelle configuration de sélection de critères exposés dans le tableau 1, incluant ce qui suit :

- Métadonnées administratives : informations de gestion interne à propos de l'enregistrement lui-même – création, modification, relations avec les autres enregistrements, etc. Des exemples d'éléments se rapportant aux métadonnées administratives incluent, mais pas de façon exhaustive, les éléments suivants :
 - Le numéro de l'enregistrement
 - La date de création de l'enregistrement
 - La date de la dernière modification
 - L'identification du créateur/réviseur de l'enregistrement
 - La langue de l'enregistrement
 - Les notes
 - Les relations de l'enregistrement avec l'(es) autre(s) enregistrement(s)
- Métadonnées descriptives : décrivent les propriétés physiques et intellectuelles, le contenu d'un item numérique ou d'un objet avec des éléments comme :
 - Le titre (aussi bien le titre alternatif que parallèle, les sous-titres, les titres courts, etc.)
 - Le créateur (auteur, compositeur, cartographe, artiste, etc.)
 - La date
 - L'éditeur
 - Les identificateurs uniques et les liens dynamiques (URI; URL; etc.)
 - Le résumé, les notes descriptives, le compte-rendu, etc.
 - Le niveau de l'auditoire
 - Le média physique, le format, etc.

- Métadonnées analytiques : informations analysant et améliorant l'accès aux contenus des ressources. Parfois connus sous le terme « métadonnées de sujet », ces éléments peuvent inclure :
 - Les vedettes-matière
 - Le thésaurus
 - Les termes d'indexation de sujets/thèmes
 - Le résumé; la table des matières (TOC)
 - Les cotes de classification dérivées des systèmes de classification
 - Les autres éléments ayant une importance locale, par exemple les affiliations départementales; les liens vers d'autres contenus électroniques reliés; etc.
- Métadonnées de gestion des droits : informations reliées aux restrictions (juridiques, financières, etc.) sur l'accès ou l'utilisation des items ou des objets numériques. Des éléments comme ceux qui suivent peuvent s'appliquer :
 - Les restrictions d'utilisation
 - Les déclarations de permis
 - Les frais d'abonnement, d'autorisation d'exploitation ou payables à la pièce
 - Les remerciements
 - Les avis de droit d'auteur
 - Les calendriers de conservation
 - Les cotes de qualité
 - La non-responsabilité des utilisateurs
- Métadonnées techniques : matériel et logiciel particuliers utilisés pour la conversion des items/objets au format numérique, ou pour l'archivage, l'affichage, etc. pouvant exiger l'utilisation d'éléments comme :
 - Les spécifications d'équipement de numérisation
 - Les positions de la caméra
 - Les conditions de tournage
 - Les paramètres de codage
 - La reconnaissance vocale et/ou le matériel et logiciel de re-lecture
 - Les spécifications de lecteur optique
 - Les équipements de traduction de l'image
 - Les exigences de logiciels de conversion pour le type de dossier
- Autres métadonnées : par exemple les éléments particuliers de métadonnées basés sur les exigences locales, régionales, organisationnelles ou en conformité avec des standards de métadonnées autorisés à l'échelle nationale, et non énumérés parmi les métadonnées ci-haut mentionnées.

2.3 Choisir un ou des schémas

Le choix d'un schéma ou de schémas de métadonnées à utiliser dans la création d'enregistrements substitués pour identifier de façon unique ou relier aux ressources accessibles via le répertoire par sujet dépendra de la visée particulière des services et des types de métadonnées devant être supportées. Cependant, un répertoire par sujet, crée et géré par un réseau de distribution d'organisations nationales avec un contenu englobant des sites Web de haute qualité (textes et images seulement) et limité par un champ de sujet d'un domaine technique peut exiger une combinaison de métadonnées administratives, descriptives et analytiques. Le *Réseau canadien de la santé* offre un exemple d'une telle configuration. Une « exposition virtuelle » contenant les liens vers une variété d'objets numériques à l'intérieur d'un consortium de galeries d'art publiques et privées et de musées nécessiterait l'utilisation de métadonnées techniques et de gestion des droits, en plus des métadonnées requises à des fins administratives, descriptives et analytiques.

Ce qui peut aider à finalement déterminer le schéma de métadonnées est le degré désiré de *granularité* ou, la quantité de détails devant être repérés et représentés dans les enregistrements de métadonnées. Un « enregistrement-noyau », créé grâce à l'utilisation d'un schéma de métadonnées, comme par exemple le Dublin Core avec ses quinze éléments de base (lesquels sont optionnels, répétables et extensibles), couvre adéquatement les métadonnées administratives, descriptives, analytiques et de gestion des droits, et peut s'adapter à l'information reliée aux spécifications techniques.

Cependant, dans plusieurs domaines spécialisés, un schéma de métadonnées tel que le Dublin Core, ne dispose pas de la granularité (détail) nécessaire pour représenter adéquatement les ressources ou les objectifs particuliers reliés au répertoire par sujet. Le standard de métadonnées ONIX conçu pour les éditions et les éditeurs ou encore le *Content Standard for Digital Geospatial Metadata* sont deux exemples de schémas riches de métadonnées, détaillés et hautement techniques, conçus spécifiquement pour traiter des contenus complexes et des applications uniques à l'intérieur du domaine.

En plus de décider du niveau de détails devant être repérés dans les enregistrements basés sur les métadonnées, le choix d'un schéma peut se limiter aux réponses apportées à des questions telles que les suivantes :

- Est-ce que le répertoire par sujet se rapporte à un (sujet ou discipline) domaine dans lequel les standards de format riches de métadonnées ont été développés ?
- Quels champs parmi les plus utiles pour la communauté de chercheurs le répertoire par sujet a-t-il l'intention de servir ? Combien de détails ces champs devrait-ils comporter ?
- Quels champs seraient les plus utiles à ceux qui créent et/ou gèrent le répertoire par sujet ? Combien de détails ces champs devrait-ils comporter ?
- Quels champs seraient nécessaires pour supporter les services spécifiques que le répertoire se propose d'offrir ?
- Est-ce que l'utilisation ou l'accès du répertoire par sujet sera limité d'une façon ou une autre ? Comment cela devra-t-il (devrait-il) être noté dans les enregistrements de métadonnées ?
- Y-a-t-il des exigences reliées à la langue, au format de matériel ou au type de média pour lesquels des champs particuliers (ou additionnels) doivent être fournis ?
- Y-a-t-il un besoin pour créer ou partager des ressources entre un réseau de collaborateurs responsables du répertoire par sujet ? Y-a-t-il des champs (additionnels) de métadonnées nécessaires à la gestion du répertoire ?
- Est-ce que l'utilisation de plus d'un schéma de métadonnées est envisagée ou requise (partage des ressources à travers les réseaux), est-ce que l'appariement des passerelles est facilement et immédiatement disponible pour faciliter et gérer l'interopérabilité ? Est-ce que les ressources représentées par un schéma de métadonnées (ou un standard) peuvent être échangées par les collaborateurs d'un répertoire par sujet qui utilisent un schéma (ou standard) différent ?
- Jusqu'à quel point un schéma particulier est utilisé et quels sont les applications ou environnements comparables à celui présentement proposé ? Combien ce schéma est-il robuste et/ou flexible à l'intérieur de différents contextes ?
- Comment faciliter la migration à partir de ce schéma particulier vers un autre schéma, la conversion des données devrait-elle être nécessaire à un moment donné ?
- Comment et jusqu'à quel point un schéma particulier se conforme aux standards organisationnels (locaux), nationaux ou internationaux, s'ils existent ?
- Quelles ressources humaines (quantité, éducation, formation) techniques, financières ou autres sont nécessaires pour supporter l'application d'un schéma de métadonnées et est-ce que mon organisation ou opération possède déjà ces ressources et sont facilement et suffisamment disponibles ? Y-a-il d'autres contraintes pratiques à l'implantation et la gestion d'un ou plusieurs schémas ?

Après avoir répondu à une ou plusieurs des questions précédentes, le choix d'utiliser un ou plusieurs schémas de métadonnées standardisés peut se confirmer. Alternativement, un individu, une organisation ou un consortium décidant de créer un répertoire par sujet peut déterminer qu'une solution locale ou « maison », c'est-à-dire un ensemble d'éléments de métadonnées conçus et supportés localement, doit être privilégiée. De la même manière, certains choisissent de combiner les éléments d'un standard établi, tel que le Dublin Core, avec les éléments appropriés à la situation locale des ressources et leurs objectifs. Il n'y a pas de recette unique ou de solution « one-size » pour l'utilisation d'un schéma de métadonnées ou d'un standard associé aux répertoires par sujet.

3.0 Défis actuels et tendances futures

En règle générale, les défis actuels sont de bonnes prévisions des questions qui nécessiteront une attention particulière dans l'avenir, que ce soit à court ou à long terme. Comme le nombre, la couverture,

la portée et les attentes de l'utilisateur final des répertoires par sujet conçus à l'aide de métadonnées augmentent, de nombreux domaines, comme ceux qui suivent, seront constamment problématiques et ouverts à la résolution :

- Interopérabilité – La nécessité d'améliorer les protocoles de métadonnées inter-domaines et les passerelles pour supporter l'échange des enregistrements ira grandissante; les standards de métadonnées supportant l'interopérabilité aux niveaux technique, sémantique, organisationnel, inter-communauté et international peuvent nécessiter d'être développés ou améliorés
- Collaboration et coopération – Les répertoires par sujet peuvent être développés en utilisant des économies d'échelle pour promouvoir l'accès aux métadonnées accessibles inter-domaines et aux ressources internationales, aussi bien que pour le partage des possibilités et des coûts de création et de gestion des schémas de métadonnées correspondants mis en commun
- Variabilité dimensionnelle – Alors qu'il est clair que plusieurs répertoires par sujet sont limités, la croissance future des services conçus aussi bien pour la portée ou l'inclusion (e.g. OCLC, CORC, UKOLN) semble inévitable. Alors que la réalisation d'un véritable répertoire par sujet universel est peu probable dans un avenir prochain, il s'agit d'un but devant être poursuivi et les fossés dans les structures de métadonnées ou éléments doivent être comblés
- Ressources multilingues – Jusqu'à maintenant les schémas de métadonnées ont été développés et appliqués dans des environnements monolingues; les demandes de l'usager final pour l'accès aux ressources multilingues dans la langue de son choix exigera de nouveaux et significatifs schémas de métadonnées développés et améliorés ou des applications novatrices des schémas de métadonnées non-verbaux déjà existants (par exemple : les systèmes de classification) pour décrire et repérer les ressources multilingues
- Moteurs de recherche et fonctionnalité des interfaces – Des moteurs de recherche de plus en plus sophistiqués exploiteront de façon plus efficace la richesse des métadonnées qui existent pour supporter le repérage des ressources; des inter-passerelles de recherche et de repérage seront requises pour le développement des schémas nouveaux ou améliorés, telles que les métadonnées pour la description de collection, les métadonnées de produits ou de précédés, les métadonnées pour les nouveaux formats ou les innovations techniques, etc., en plus, de nouveaux éléments peuvent être requis pour décrire ou supporter la fonctionnalité améliorée de l'interface (par exemple : l'accessibilité, l'utilisation, la navigation, les caractéristiques pour supporter les besoins spéciaux, etc.)
- Outils de métadonnées – Jusqu'ici la création d'enregistrements de métadonnées a été étudiée comme un vaste exercice manuel (et humain). Comme le nombre de répertoires par sujet s'accroît, les approches automatiques seront de plus en plus utilisées, au moins comme une première étape d'intervention nécessitant la médiation humaine par la suite. Des outils qui amassent et indexent automatiquement les ressources, des structures d'enregistrements de métadonnées pré-définies seront de plus en plus utilisés pour la gestion des répertoires par sujet.
- Registres et utilisation locale – Des perfectionnements et développements aux schémas de métadonnées existants ont eu comme conséquence la création de registres pour enregistrer et repérer les changements; le nombre de variations locales pour le standard a accéléré le développement des registres; à l'avenir il pourrait y avoir un mouvement pour décourager ou limiter les applications locales et non standards, malgré la disponibilité d'un registre
- Politiques, questions juridiques, authentification – La croissance des répertoires par sujet peut nécessiter le développement ou l'amélioration des métadonnées reliées à la gestion des droits de propriété intellectuelle (PI) et l'originalité/authenticité des ressources.
- Conformité des standards – Assurer ou même renforcer la conformité avec les standards internationaux, les standards de métadonnées inter-domaines ou les protocoles, doivent devenir une priorité lorsque les répertoires par sujet se multiplient et que plus en plus de collaborateurs participent

Comme les éléments précédents peuvent le démontrer, la portée et la diversité des questions soulevées par l'application des schémas de métadonnées dans les répertoires par sujet sont suffisamment nombreuses pour occuper les chercheurs et les praticiens des métadonnées, au moins pour un certain temps encore.

Références choisies :

Campbell, Debbie (2000). *Australian subject gateways : metadata as an agent of change*. Présentée à VALA 2000 Conference, Melbourne, Australie, 18 février, 2000. 7 p.

Disponible à : <http://www.nla.gov.au/nla/staffpaper/dcampbell2.html> Visionnée le 10 mai 2003.

DESIRE Information Gateways Handbook. 2000. 149 p.

Disponible à : <http://www.desire.org/handbook/print4.html> Visionnée le 30 avril 2003.

De Jong, Annamieke (2002). Audiovisual domain. In *SCHEMAS Metadata Watch report #8 and Standards Framework report #4 : Appendix A*. 6p.

Disponible à <http://www.schemas-forum.org/metadata-watch/d29/d29.htm> Visionnée le 24 avril 2003.

Howarth, Lynne C. (2003). *Metadata unplugged*. Présentation faite au 25^e Anniversary Conference of Substance Abuse Librarians and Information Specialists, Totonto, Canada, 25 avril 2003.

Disponible à <http://www.SALIS.org/> Visionnée le 19 mai 2003.

Koch, Traugott (2000). Quality-controlled subject gateways : definitions, typologies, empirical overview. *Online information review*. Vol. 24:1. 17 p.

Disponible à : <http://www.lub.lu.se/tk/publ/OIR-SBIG.html> Visionnée le 7 mai 2003.