



# Nouvelles de l'ICBN

Nouvelles de l'ICBN No 9    Septembre 2003  
ISSN 1488-2000

INITIATIVE CANADIENNE  
sur les  
bibliothèques numériques



## À l'ordre du jour



### La réunion ouverte de novembre 2002 : à l'origine du Rapport Godolphin

Les points de vue exprimés à la réunion ouverte de novembre 2002 de l'ICBN ont donné lieu à une étude de l'organisation. Le Comité de direction a décidé, à sa réunion du 20 janvier 2003 (Simon Fraser University), d'embaucher un consultant afin de réexaminer et de réévaluer les directions données à l'ICBN. Il ressort que l'environnement a changé considérablement depuis la fondation de l'organisation en 1997. Les procès-verbaux des réunions du Comité de direction sont affichés sur le site Web de l'ICBN.

Le 10 mars, le Comité de direction a tenu une réunion extraordinaire (Ottawa) et a embauché Jocelyn Godolphin et associés pour préparer un rapport sur les options organisationnelles qui s'offrent à l'ICBN. Le rapport final fut remis au Comité de direction à la fin d'avril 2003.

Le Comité de direction de l'ICBN examinera en détail les options suggérées dans le rapport lors de sa réunion du 18 septembre 2003 à Ottawa. Le rapport est affiché sur le site Web de l'ICBN à <http://www.nlc-bnc.ca/cidl/newe.html>

### NOUVELLE CATÉGORIE DE COMMANDITE À L'ICBN

L'année de cotisation 2003-2004 à l'ICBN offrira un troisième genre de participation, la *commandite*. Moyennant une cotisation annuelle de 1 500 \$, les commanditaires ont droit à des avantages de promotion et ont accès à la ListServ de l'ICBN. Ils n'ont cependant pas droit de vote. Encouragez vos fournisseurs de connexion numérique à participer. Communiquez avec l'ICBN à : [cidl-icbn@nlc-bnc.ca](mailto:cidl-icbn@nlc-bnc.ca)

### OUR ROOTS / NOS RACINES

Le projet « Les histoires locales du Canada en ligne » a mis en ligne le texte intégral d'environ 4 000 nouveaux titres au cours de l'An deux. Vous trouverez sur le site les nouvelles troupes pédagogiques de la maternelle à la 12<sup>e</sup> année.

Des quelques 50 participants à la réunion ouverte de l'ICBN présidée par Claude Bonnelly en novembre 2002, quatre étaient de l'équipe Nos racines. On aperçoit ici, de l'Université Laval et de l'University of Calgary de gauche à droite : Université Laval : Claude Bonnelly, directeur de la bibliothèque, et Guy Teasdale, bibliothécaire et conseiller en documentation électronique. University of Calgary : Jackie Bell, gestionnaire du projet Nos racine et Frits Pannekoek, directeur des ressources d'information.

[www.nosracines.ca](http://www.nosracines.ca)



### DANS CETTE PUBLICATION:

**Normes de numérisation de quatre projets:**  
« Métadonnées pour « Les histoires locales du Canada en ligne »

Linda Pearce, University of Calgary

« Métadonnées & Notre mémoire en ligne »

William Wueppelmann, ICMH

« Les métadonnées dans le projet des thèses et mémoires de l'Université Laval »

Rida Benjelloun, Pierre Lasou, Université Laval

**Recherche de métadonnées et de ressources au sein du gouvernement du Canada**

Rôle de la Bibliothèque et Archives Canada  
Deane Zeeman, Katherine Miller-Gatenby, BAC

**Forum sur les métadonnées au Canada, Ottawa**  
Le 19 et 20 septembre 2003

**Profil @ l'ICBN:**

Tim Au Yeung, University of Calgary

**Répertoire des projets canadiens de numérisation**

## **OUR ROOTS / NOS RACINES : MÉTADONNÉES POUR LES HISTOIRES LOCALES DU CANADA**

**Linda Pearce**, gestionnaire, Systèmes de bibliothèque,  
Services des technologies de l'information  
University of Calgary

« *Our Roots / Nos racines* » est un projet canadien de numérisation passionnant entrepris par University of Calgary Press et la bibliothèque de l'Université Laval. Il vise à rendre disponible en ligne la collection la plus complète qui soit d'histoires locales canadiennes françaises et anglaises. Ces joyaux littéraires souvent oubliés peuvent servir à une multitude de choses, notamment en histoire, en agriculture, en architecture, en généalogie, en études religieuses et en science économique, pour ne citer que quelques domaines qui y trouveront une riche source d'information et d'inspiration.

Pour rendre ces histoires locales disponibles sur Internet, le Comité national de rédaction et le personnel du projet se sont lancés dans une série d'activités distinctes, notamment pour amasser des fonds, établir des listes de priorités pour la numérisation, coordonner les tâches des établissements participants, mettre au point des logiciels, acquérir du matériel et du soutien informatique, embaucher et former du personnel, se procurer des documents, obtenir l'affranchissement des droits, procéder au balayage et à l'archivage des documents, concevoir des sites Web, télécharger les données et fournir des métadonnées.

Sans ce dernier point, à savoir les précieuses métadonnées, aucun de ces documents ne serait consultable ni utilisable. Les métadonnées sont comme des hameçons qui permettent d'attraper ce dont on a besoin dans n'importe quelle collection d'objets numérisés, qu'il s'agisse de livres, de diapositives, de photographies aériennes, de vidéo-clips ou d'autres types de formats.

Les métadonnées d'un projet fournissent un mécanisme qui permet de gérer des centaines de milliers d'images de pages individuelles et de gérer le déroulement du projet, dans son ensemble. Les métadonnées sont habituellement de trois sortes, à savoir descriptives, administratives et structurelles.

Les **métadonnées descriptives** du projet *Our Roots / Nos racines* sont basées sur la norme du Dublin Core, des AACR2 II et de SGI, ainsi que sur les besoins locaux spéciaux en matière d'étiquetage. La définition des normes pour les métadonnées et des pratiques de catalogage se fait en étroite collaboration entre le responsable de la numérisation, d'autres employés des Services de TI du centre des ressources en information de l'Université de Calgary ainsi que les préposés au catalogage. Dans la mesure du possible, les métadonnées descriptives proviennent des catalogues de bibliothèque, notamment ceux de la Bibliothèque et des Archives du Canada, de l'Université Laval et de l'Université de Calgary. Lorsque l'on ne trouve aucun document, on crée des métadonnées qui sont saisies par des spécialistes en catalogage.

**Voir Pearce page 8.**

### **Profil @ CIDL Tim Au Yeung**

**Ressources de l'information de l'University of Calgary,  
Gestionnaire, Technologie d'archivage des objets  
numériques**

En qualité de gestionnaire de la technologie d'archivage des objets numériques à l'University of Calgary, Tim Au Yeung, joue un rôle central dans les opérations du Centre de numérisation de l'University of Calgary. Le centre a lancé avec succès l'Alberta Heritage Digitization Project ([www.ourfutureourpast.ca](http://www.ourfutureourpast.ca)) et le site Histoires locales du Canada en ligne ([www.nosracines.ca](http://www.nosracines.ca)) en collaboration avec des organismes partenaires partout au Canada. Tim dirige le développement de l'infrastructure technique nécessaire à l'appui de ces vastes projets de numérisation (Our Future, Our Past compte présentement près d'un million d'images en ligne et Nos racines comprend 3000 volumes en ligne constitués de 625 000 pages de texte numérisé) ainsi que le développement des normes et protocoles techniques concernés.

Avant de se joindre à l'University of Calgary, Tim Au Yeung a été surveillant du matricage et de l'assurance de la qualité des produits du service de développement du contenu visuel chez Adobe Systems Incorporated et EyeWire, Incorporated. Au cours de son affectation chez Adobe Systems Incorporated, il a mis sur pied l'équipe d'assurance de la qualité et a conçu les protocoles de développement du contenu visuel.

D'autres établissements d'enseignement se sont rapidement joints au projet. Leur liste est disponible sur le site de **Our Roots / Nos racines**: [www.nosracines.ca](http://www.nosracines.ca) et [www.ourroots.ca](http://www.ourroots.ca) .

Pour obtenir de plus amples renseignements, visitez les adresses suivantes :

**Normes relatives aux métadonnées pour les projets de numérisation**

<http://www.ucalgary.ca/~pearce/metadatastandards.htm>

Lignes directrices pour la création de métadonnées

<http://www.ucalgary.ca/~pearce/metadataguidelines.html>

**Normes techniques du projet *Our Roots* (en anglais) :**

<http://www.ourroots.ca/e/tech1.asp>

# Métadonnées et Notre mémoire en ligne

William Wueppelmann, Spécialiste des systèmes électroniques  
Institut canadien de microreproductions historiques



Récemment, l'Institut canadien de microreproductions historiques (ICMH) a modifié sa façon de considérer et de traiter les métadonnées de Notre mémoire en ligne (NML). NML est hébergée à l'adresse Web suivante : <http://www.canadiana.org/>. Notre démarche se fonde maintenant sur de nouveaux objectifs, de nouvelles normes et méthodes de traitement des métadonnées.

Notre format de métadonnées précédent était basé sur le langage SGML. Toutefois, en raison d'un manque de validation adéquate et d'homogénéité dans l'application des normes, ainsi que d'une variété de pratiques et de méthodes de génération et d'édition des enregistrements, ce format est devenu ce que l'on pourrait appeler un format spécialisé. Seule notre application de base de données pouvait lire ce format de manière fiable et encore, pas toujours si fiable que cela! Nos enregistrements étaient également clairs, contenant peu d'information autre que celle immédiatement pratique pour l'application originale. En concevant un nouveau format et de nouvelles méthodes, nous avons essayé de corriger ces lacunes. Nos nouveaux enregistrements de métadonnées se font dans des applications XML et ils sont validés à l'aide d'une définition de type de document (DTD) conçue à l'interne. En développant notre format et notre marche à suivre, nous avons tenu compte d'un certain nombre de facteurs : objectif, contenu, format, entretien et préservation.

Nous avons d'abord et avant tout conçu nos enregistrements comme des entrepôts de données. L'aptitude à stocker proprement et précisément l'information est plus importante que la compatibilité du format avec n'importe quelle application particulière. En effet, les fichiers XML peuvent facilement être convertis, abrégés et transformés pour créer d'autres documents pouvant être utilisés avec une application particulière.

Nos nouveaux enregistrements ne sont pas directement compatibles avec notre application de base de données existante; mais une simple étape suffit pour en tirer des fichiers de production pouvant être indexés et manipulés par notre base de données. Nous prévoyons passer bientôt à une nouvelle application. Cette application n'utilisera pas non plus les enregistrements de métadonnées originaux tels quels; cependant, il sera facile de les utiliser pour générer des enregistrements pouvant être utilisés plus efficacement.

En ce qui a trait au contenu, notre philosophie globale est la suivante : « Dans le doute, laissons-les en place! »

Les fichiers XML occupent peu d'espace relativement à l'ensemble du projet NML. Notre bibliothèque de métadonnées entière occupe environ 4 Go d'espace disque, comparativement à plus de 120 Go de données d'image. Il est facile de revenir en arrière et de retirer l'information indésirable des fichiers; il est beaucoup plus difficile de revenir en arrière et d'ajouter une nouvelle information dont on a soudainement besoin. Nous incluons donc à nos fichiers de métadonnées tout ce qui pourrait un jour être utile et qui peut être obtenu avec un effort raisonnable. Nous prenons également soin de préserver la granularité; il est facile d'amalgamer des champs, mais difficile de les séparer; alors, nous préférons garder l'information fragmentée et l'assembler au besoin, plutôt que l'amalgamer et découvrir éventuellement que nous aurions besoin d'une partie de ce qui est devenu un champ unique.

Parmi les éléments que nous incluons dans chaque fichier de métadonnées se trouve la notice MARC complète pour l'article (étiquetée selon le champ et le sous-champ); le lecteur de caractères optiques complet, étiqueté selon les images de la page d'où provient chaque bloc de texte et selon les paramètres de langage utilisés pour obtenir ce texte et l'information concernant chaque image, incluant la résolution, le format et les dimensions en pixels. Bon nombre de ces renseignements, ainsi que de nombreuses zones de contrôle de la notice MARC, ne sont d'aucune utilité immédiate; mais il est à la fois plus sécuritaire et plus facile de les inclure maintenant plutôt que de choisir les champs désirés tout en nous demandant si nous faisons le bon choix!


Nos fichiers de métadonnées sont stockés en tant que documents XML, lesquels sont conçus pour être validés par notre DTD. Il est alors facile de nous assurer que chaque fichier est complet et structurellement sain au moment de sa génération ainsi qu'après chaque modification. XML est un format ouvert soutenu par un nombre important et croissant d'applications et d'interfaces de programmes et de scripts (interfaces de programmation - API). Cela rend nos données facilement transférables dans divers systèmes d'exploitation, architectures d'ordinateur et applications; ce qui signifie que nous pourrions facilement les utiliser à de nouvelles fins et dans de nouveaux contextes, si besoin est.

Une autre raison pour laquelle nous avons choisi XML pour nos fichiers est la facilité d'entretien de ces derniers. Nos fichiers sont générés et modifiés par un certain nombre de petits programmes Perl que nous avons créés, mais ils pourraient également être entretenus à l'aide d'autres outils sur mesure, commerciaux ou en libre accès. Après la création ou la modification de chaque document, nous le validons à l'aide de notre DTD pour nous assurer,

**Wueppelmann, suite à la page 4.**

**Wueppelmann  
de la page 3**

qu'il est correct. Bien que ce processus ne puisse éliminer toute donnée incorrecte due à l'erreur humaine, il empêche bel et bien les erreurs de syntaxe, les champs manquants ou déréglés et autres erreurs structurelles de se glisser dans nos métadonnées. Si, à un certain moment, nous décidions de retirer notre format actuel, nous sommes certains de pouvoir réaliser une migration informatisée sans erreur vers n'importe quel nouveau format que nous pourrions choisir ou concevoir.

Nous ne considérons pas nos métadonnées comme une entité statique. Au contraire, nous croyons que le contenu évoluera au fil du temps. Par exemple, les éventuelles améliorations technologiques apportées aux lecteurs optiques pourraient nous inciter à re-numériser plusieurs de nos images, afin d'obtenir une meilleure qualité de reconnaissance de texte ; des erreurs de catalogue seront découvertes et devront être corrigées, et d'autres changements devront être effectués. Notre but est d'assurer la disponibilité de nos métadonnées dans le format le plus à jour possible. Nous disposons d'un entrepôt central, où nous stockons des copies documentées de chaque fichier. Tous les changements aux fichiers sont effectués dans cet entrepôt et, au besoin, copiés sur d'autres exemplaires de nos fichiers. Nous maintenons de plus une copie à jour de nos métadonnées sur le serveur NML, situé dans une autre partie de la ville. Nous effectuons également des copies de secours périodiques sur CD-ROM, afin de disposer d'une copie portable raisonnablement à jour des données. 



## Projets numériques récemment par les membres de l'ICBN

<http://www.nlc-bnc.ca/initiatives/erella.htm>

### Arctic Blue Books Online

#### University of Manitoba, Archives et collections spéciales

<http://www.umanitoba.ca/libraries/units/archives/arcticbb/index.shtml>

Index unique des documents du parlement britannique du 19<sup>e</sup> siècle concernant l'Arctique canadien.

« Au début, le groupe de projet travaillait dans un petit bureau humide au sous-sol de l'University College. Ayant obtenu certains appuis, il a emménagé dans un local plus grand et plus sec dans la toilette des dames au sous-sol. C'est ainsi qu'est né le « Lady Jane's Loo and Lab ».

### Leonard Frank Photograph Collection

#### Vancouver Public Library

<http://collections.ic.gc.ca/vpl/>

La collection de Leonard Frank, la première importante collection acquise par la Vancouver Public Library, constitue les assises de la collection de photographies historiques de la bibliothèque. La carrière de photographe de M. Frank couvre les années 1910-1944.

### Architecture des hôpitaux de Montréal

#### Programme des collections numériques de l'Université McGill

<http://digital.library.mcgill.ca/hospitals/>

Couvrant 358 ans, de 1642 à nos jours, ce site contient plus de 1000 images, une chronologie de la construction d'un hôpital, des dossiers d'architecture et une étude approfondie de l'hôpital Royal Victoria.

### Urbanisme

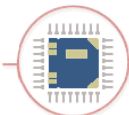
#### Programme des collections numériques de l'Université McGill

<http://digital.library.mcgill.ca/urbanplanning/>

La collection sur l'urbanisme est une base de données de rapports et de plans d'aménagement de régions urbaines et rurales de toutes les provinces du Canada depuis la fin des années 1950 jusqu'à maintenant. La valeur de cette collection, constituée au fil des ans comme outil pédagogique par l'École d'urbanisme de l'Université McGill, vient du fait qu'elle sert de dossier historique du façonnement des paysages urbains et ruraux du Canada et de certaines villes des États-Unis.

Le Répertoire fournit des informations au sujet des projets offrant des ressources accessibles sur le Web ainsi que les liens permettant de s'y rendre. Tous les sujets sont couverts. Les entrées d'organismes publics, privés et sans buts lucratifs sont toutes les bienvenues. Les entrées conduisant à des ressources accessibles sur le Web créées à l'extérieur du Canada mais couvrant des sujets canadiens sont également acceptées.

Inventory of Canadian  
Digital Initiatives



Répertoire des projets  
canadiens de numérisation

# Les métadonnées dans le projet des thèses et mémoires de l'Université Laval

Rida Benjelloun, Conseiller en gestion de l'information numérique  
Pierre Lasou, Spécialiste en ressources documentaires  
Bibliothèque, Université Laval



La collection des mémoires et thèses de l'Université Laval (<http://www.theses.ulaval.ca>) a été lancée en novembre 2002. Le format d'archivage des thèses et mémoires est le XML (eXtensible Markup Language). Les formats de diffusion sont le XHTML et le PDF.

Les choix technologiques en matière de métadonnées pour le projet des thèses de l'Université Laval ont été effectués en fonction de normes et de standards internationaux. L'une des normes de métadonnées les plus reconnues à présent est le Dublin Core (NISO Z39.85 et ISO 15836). Elle prévoit 15 éléments de métadonnées qui peuvent être exprimés dans différents formats dont le HTML, le XML ou encore le RDF (Resource Description Framework).

---

Pour les thèses et mémoires de l'Université Laval, nous utilisons un standard dérivé du Dublin Core : le ETD-ms.

---

## Formats des métadonnées

ETD-ms (Interoperability Metadata Standard for Electronic Theses and Dissertations [version 1.00 <http://www.ndltd.org/standards/metadata/current.html>]) est un standard international pour les métadonnées des thèses et mémoires développé par le NDLTD (Networked Digital Library of Theses and Dissertations). Ce standard se présente sous la forme d'un schéma XML simple. Il est fondé sur certains éléments du Dublin Core et en prévoit d'autres propres aux thèses et mémoires. Ces derniers concernent essentiellement le diplôme : nom, niveau, discipline.

---

En l'absence de standards internationaux pour les modèles de document propre aux thèses et mémoires, nous avons opté pour une solution transitoire en choisissant de nous conformer à la DTD (Document Type Definition) XML DocBook. ([www.oasis-open.org/committees/docbook/](http://www.oasis-open.org/committees/docbook/)).

---

Le ETD-ms est la pierre angulaire des métadonnées dans le projet des thèses et mémoires de l'Université Laval. A partir des notices XML de métadonnées, nous dérivons d'autres formats de métadonnées :

- Dublin Core pour les versions XHTML des thèses et mémoires,
- Éléments de métadonnées inclus dans les versions XML.

## Indexation, recherche et dissémination

Les fichiers de métadonnées ETD-ms sont indexés par le moteur de recherche Inktomi qui permet une indexation de chaque élément XML. La recherche dans la collection des thèses et mémoires s'en trouve considérablement enrichie. Il

---

Pour donner une visibilité plus grande à la collection, nous nous sommes engagés dans l'Open Archives Initiatives ... qui a développé un protocole ...

est en effet possible de réaliser des recherches selon des champs prédéfinis (auteur, titre, résumé français, direction de recherche, grade postulé, etc.).

Pour donner une visibilité plus grande à la collection, nous nous sommes engagés dans l'Open Archives Initiatives (<http://www.openarchives.org>) qui a développé un protocole pour l'échange et la diffusion des métadonnées. Le protocole prévoit deux catégories de participants : les « data provider » (fournisseurs de données) et les « service provider » (fournisseurs de services).

Pour les thèses et mémoires électroniques la Bibliothèque participe en tant que « data provider ». L'application choisie pour la mise en place du protocole OAI est celle développée par l'OCLC : OAICat (<http://www.oclc.org/research/software/oai/cat.shtm>).

Dans le cadre de l'OAI, la collection est « récoltée » par plusieurs « service provider » dont :

« XTCat Experimental Thesis Catalog » de l'OCLC : <http://alcme.oclc.org/ndltd/index.html>

Public Knowledge Project's metadata archive <http://www.pkp.ubc.ca/harvester/>

## La gestion de métadonnées

Les métadonnées peuvent nécessiter des mises à jour ou des corrections. Il est donc important de prévoir un système qui permette de les gérer. Actuellement, les notices XML sont stockées dans une arborescence de répertoire. Nous essayons de nous orienter vers la gestion des métadonnées dans une base de données XML.

# Recherche de métadonnées et de ressources au sein du gouvernement du Canada Rôle de la Bibliothèque et Archives Canada



**Deane Zeeman**, coordinatrice des métadonnées, Accès bibliographique  
**Katherine Miller-Gatenby**, directrice, Groupe de travail du Gouvernement en direct  
Bibliothèque et Archives Canada

## Gouvernement en direct (GED)

Comme plusieurs autres gouvernements (par ex., ceux du R.-U., de la Nouvelle-Zélande, de Singapour, etc.), le gouvernement canadien reconnaît qu'il doit compter sur des citoyens informés pour assurer le bien-être social et économique du pays. C'est pourquoi ceux-ci doivent disposer d'un accès facile et fiable à l'information, aux programmes et aux services du gouvernement fédéral. Dans cette optique, le Canada s'est appuyé sur un effort inter-organisationnel appelé *Un Canada branché*, notamment sur l'initiative Gouvernement en direct (GED), qui vise à faire de l'Internet un outil de prestation de services pour le gouvernement du Canada.

## Le défi de la normalisation

Le gouvernement du Canada compte plus de 200 ministères et organismes qui disposent de sites Internet leur permettant de couvrir une vaste gamme de sujets. Il va de soi que le fait d'assurer une présence cohérente du gouvernement du Canada qui intègre tous ces sites représente une tâche exigeante, mais néanmoins nécessaire. Dès le début de l'élaboration de l'initiative GED, le gouvernement fédéral a reconnu que les citoyens devaient pouvoir disposer d'un accès cohérent et fiable à l'information affichée sur le Web. C'est pourquoi il a mis au point et implanté un processus de Normalisation des sites Internet (NSI) [http://www.cio-dpi.gc.ca/clf-upe/index\\_e.asp](http://www.cio-dpi.gc.ca/clf-upe/index_e.asp).

La NSI répond à plusieurs préoccupations entourant l'accès électronique à l'information gouvernementale, notamment en ce qui concerne l'utilisation de métadonnées afin d'améliorer la capacité des citoyens de trouver les renseignements dont ils ont besoin sur le Web. En octobre 2001, les ministres du Conseil du Trésor ont approuvé la norme Dublin Core (DC) sur les métadonnées comme norme de gestion de l'information gouvernementale et cinq éléments de métadonnées ont été adoptés à titre d'éléments essentiels pour la NSI.

## Éléments de métadonnées

Les cinq éléments de métadonnées DC établis pour être intégrés à la norme du gouvernement du Canada sont les suivants : titre, créateur, langue, date et sujet. La norme prescrit des ensembles de valeurs prédéterminés devant être utilisés comme contenu pour quatre de ces cinq éléments. (Le **titre** est le titre réel de la ressource en information sur le Web et n'est, par conséquent, pas sujet à la normalisation). Cette démarche facilite la recherche d'information semblable et favorise l'interopérabilité pour l'ensemble du gouvernement. Les ensembles de valeurs prescrits sont les suivants :

**Créateur** : Conseil du Trésor du Canada. *Titres des organismes fédéraux* (12 mars 1998)

[http://www.tbs-sct.gc.ca/Pubs\\_pol/sipubs/TB\\_FIP/titlesoffedorg1\\_e.asp](http://www.tbs-sct.gc.ca/Pubs_pol/sipubs/TB_FIP/titlesoffedorg1_e.asp)

**Langue** : ISO639-2 Codes pour la représentation des noms de langue : Partie 2n : code Alpha-3

<http://lcweb.loc.gov/standards/iso639-2/langhome.html>

**Date** : ISO 860 : *Éléments de données et formats d'échange – Échange d'information - Représentation de la date et de l'heure*

<http://www.w3.org/TR/NOTE-datetime>

## Forum sur les métadonnées au Canada, le 19 et 20 septembre 2003 Bibliothèque et Archives Canada

Le Forum vise à mettre en présence des participants des collectivités d'utilisateurs gouvernementaux et non gouvernementaux de métadonnées du Canada, qui partageront leurs préoccupations et élaboreront une démarche commune relativement à l'utilisation des métadonnées en vue d'améliorer la recherche d'information.

Pour plus de renseignements sur le Forum sur les métadonnées, consulter le <http://www.nlc-bnc.ca/metaforum/>

**Sujet :** La démarche établie pour l'application du vocabulaire contrôlé à l'élément de métadonnées sujet est complexe et comporte plusieurs couches. La première couche est un thésaurus large de haut niveau appelé Thésaurus des sujets de base du gouvernement du Canada (TSB) [http://en.thesaurus.gc.ca/these/thes\\_e.html](http://en.thesaurus.gc.ca/these/thes_e.html), qui est prescrit comme thésaurus par défaut pour les organisations fédérales par la Norme du Conseil du Trésor sur les technologies de l'information (NCTTI 39.2) [http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit392\\_e.asp](http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit392_e.asp). En outre, plusieurs ministères ont élaboré leurs propres thésaurus de sujets ou utilisent des thésaurus mis au point à l'échelle internationale pour leur domaine particulier. Ces outils procurent des couches de couverture de sujet plus étroites et plus profondes que le TSB.

### **Vocabulaires contrôlés pour la gestion de l'information**

Le Secrétariat du Conseil du Trésor (SCT) reconnaît que ces ensembles de termes sont utiles à la fois pour la découverte de ressources et pour la gestion du contenu des sites, mais que leur utilisation doit être contrôlée. La Bibliothèque et les Archives nationales du Canada ont le mandat d'enregistrer tous les thésaurus et les autres vocabulaires contrôlés en usage au sein du gouvernement fédéral et de rendre disponibles les listes des ensembles de termes pouvant être utilisés dans l'espace Web du gouvernement du Canada. Ce registre est appelé, à terme, à devenir un outil puissant puisqu'il reliera la terminologie utilisée dans les recherches effectuées au moyen des moteurs de recherche au contexte d'un vocabulaire contrôlé particulier.

### **Boîte à outils relative aux métadonnées**

Les outils en place sont les suivants : un cadre de métadonnées à l'échelle du gouvernement, des normes et une terminologie prédéterminée pour appliquer des valeurs à quatre des cinq éléments obligatoires; des consignes pour déterminer le cinquième, un thésaurus de haut niveau standard et un registre de vocabulaires contrôlés. Autre outil d'accès du public à l'information du gouvernement du Canada : le modèle des *regroupements*. Des ministères oeuvrent ensemble afin de regrouper leur information autour de thèmes et de manière à couvrir toutes les divisions organisationnelles. Ces regroupements dépassent les limites des ministères et comprennent souvent des contributions d'un grand nombre de ministères et d'autres fournisseurs d'information (par ex., les provinces, les OSS, etc.). Par le biais de trois passerelles, les regroupements sont actuellement utilisés sur le site Web du gouvernement du Canada au [www.canada.gc.ca](http://www.canada.gc.ca).

### **Questions en suspens**

Des questions subsistent toutefois : comment les utilisateurs effectuent leurs recherches dans l'espace Web du gouvernement fédéral et peut-on faire en sorte que les outils qui leur sont fournis soient plus utiles ? Est-il possible de rendre la recherche dans cet espace plus facile, plus intuitive et plus fiable ? Il ne fait pas de doute que la structure nécessaire est en place. Elle est robuste et souple, mais les recherches démontrent que, malgré le fait que la recherche est une option populaire de découverte des ressources en place, les gens se sentent plus à l'aise de fureter là où ils peuvent voir le contexte à mesure qu'ils passent d'une couche à l'autre. En outre, ce qui est plus important, ils veulent se servir de leur propre terminologie, et non d'un vocabulaire de spécialistes imposé de l'extérieur.

Pour combler cette lacune, la Bibliothèque et les Archives nationales et le SCT se sont associés à d'autres partenaires gouvernementaux pour élaborer des taxonomies à l'appui de la recherche selon le point de vue du client. Il nous faut manifestement créer une terminologie de premier ordre qui reliera l'univers de notre information organisationnelle et nos sujets collectifs aux modèles conceptuels des publics et de nos utilisateurs, ce qui permettra la navigation dans le contenu des regroupements et nous permettra de déterminer tout le contenu pertinent à un concept particulier. Ainsi, les utilisateurs pourront effectuer des recherches et naviguer de manière cohérente. Les tests actuels touchant les modèles conceptuels s'avèrent positifs, mais il faudra poursuivre le travail.

Par une utilisation efficace des métadonnées et des normes s'y rapportant, le gouvernement du Canada permettra aux citoyens de se servir de l'Internet pour trouver l'information gouvernementale dont ils auront besoin au moment où ils en auront besoin. 

**ICBN: un commanditaire du  
Forum sur les métadonnées au Canada  
le 19 et 20 septembre 2003  
Bibliothèque & Archives Canada**

**Membres d'ICBN parlant au Forum sont:**  
**Alexander Eykelhof**  
*Director, Information technology and  
Ontario Colleges Digital Library  
The Bibliocentre, Centennial College*

**David McKnight**  
*Digital Collections Librarian  
Bibliothèque, Université McGill*

*Le but de ce forum est de rassembler des participants en provenance de différentes communautés canadiennes de métadonnées, que ce soit du secteur public ou du secteur privé, telles que les bibliothèques, les archives, les musées, l'industrie, les institutions d'enseignement et le monde universitaire. Des exposés seront présentés sur tous les aspects des métadonnées.*

**Information sur le programme:**  
<http://www.nlc-bnc.ca/metaforum/n11-201-f.html>



**Pearce**  
**de la page 2**

Le site *Our Roots / Nos racines* permet aux visiteurs d'explorer et de faire de la recherche par auteur, titre ou sujet. Les utilisateurs peuvent cliquer sur une carte du Canada et en extraire tous les éléments pour une province donnée. En outre, on peut consulter toutes les histoires locales dans leur intégralité et visualiser les documents dans trois résolutions différentes.

Choisir un ensemble final de champs pour un projet de grande envergure n'est pas une mince affaire car les exigences locales empêchent toujours que l'on adopte simplement, en bloc, l'un des schémas existants. Les métadonnées que nous avons conçues permettent d'élaborer des mécanismes de recherche et d'extraction beaucoup plus sophistiqués.


Il est extrêmement important de faire preuve de souplesse dans l'élaboration des métadonnées et dans le soin avec lequel on crée les métadonnées pour ce projet. Nous nous sommes assurés, en adhérant à des normes bien établies, de créer des métadonnées qui puissent facilement être transférées vers d'autres bases de données ou converties en d'autres projets de métadonnées. Nous avons également prévu l'inclusion d'une grande variété de formats afin que les métadonnées puissent facilement être élargies pour manipuler des éléments tels que des diapositives d'art. Nous avons dressé des tableaux de correspondance pour définir les liens entre notre ensemble final de métadonnées et les champs de Dublin Core, CanCore, SGI et MARC.

**Les métadonnées administratives et structurelles** sont définies à l'échelle locale et elles permettent de gérer les objets numériques ainsi que le déroulement des opérations intensives pour un projet d'une aussi grande envergure. En outre, l'agent chargé des droits d'auteur à l'Université de Calgary utilise l'application RightsFlow, créée sur une plate-forme de Remedy, afin de traiter les demandes d'affranchissement des droits d'auteur. Puisque les données de cette application Remedy correspondent en grande partie à celles de la base de données de *Our Roots / Nos racines*, nous sommes en train de discuter de la création d'interfaces entre les deux pour éviter d'avoir à entrer les données deux fois. Toutefois, les histoires locales ont été, pour la plupart, publiées il y a longtemps et beaucoup d'entre elles ne nécessitent pas d'affranchissement des droits d'auteur.

Il y a également la question du balisage. Nous avons créé des pages d'index ou des tables des matières pour tous les livres du site. Il est possible d'utiliser TEI ou un autre mécanisme de balisage XML pour insérer des métadonnées dans les documents en vue de créer automatiquement des éléments de contenu cliquables. Pour des raisons de contraintes de temps et de travail, nous ne l'avons pas fait.

Toutefois, nous sommes en train de mettre au point un mécanisme d'implantation de la norme METS (norme relative à l'encodage et à la transmission de métadonnées). Une fois qu'un volume donné est repéré par le biais des métadonnées descriptives, il faut le présenter à l'utilisateur. C'est dans les métadonnées structurelles que se trouvent la plupart des renseignements nécessaires à la présentation de chaque volume, comme s'il s'agissait d'un livre en ligne.

À l'heure actuelle, les données structurelles sont basées sur des travaux réalisés, au départ, pour le projet *Alberta Heritage Digitization Project*. Depuis la sortie officielle de la norme METS, en juillet 2002, nous travaillons à codifier la mise en œuvre d'une démarche axée sur METS en matière de métadonnées structurelles ainsi que l'association objets-métadonnées. Cela se fait par le biais d'un schéma de mise en œuvre prévoyant un ensemble de règles plus rigides pour créer des documents en XML METS, dans le contexte précis de notre projet *Our Roots / Nos racines*.

Ce qui est particulièrement intéressant, c'est qu'un objet METS peut non seulement contenir des métadonnées, mais également préciser les comportements associés à la présentation d'un objet précis. Cette possibilité permettra à *Our Roots / Nos racines* de présenter des documents que l'on pourra transférer sur un plus grand nombre de plates-formes et d'environnements qu'on ne le peut actuellement, dans l'infrastructure existante. 

**Nouvelles de**  
**l'ICBN #9**  
**Septembre**  
**2003-08-26**  
**ISSN 1488 2000**

Révision : Michelle Landriault  
Coordinatrice, ICBN

Cette publication est  
présentement livrée deux fois  
par année par l'Initiative  
canadienne sur les  
bibliothèques numériques  
(ICBN).

Les articles sont publiés sous  
droit d'auteur de leurs auteurs  
respectifs.

Envoyez vos questions ou  
soumissions à :  
cidl-icbn@nlc-bnc.ca.

Pour vous joindre à l'ICBN,  
visitez notre site Web :  
<http://www.nlc-bnc.ca/cidl/>

Initiative canadienne sur les  
bibliothèques numériques  
Salle 215,  
395, rue Wellington  
Ottawa ON K1A 0N4

