

Cataloguer le cyberspace : le défi des ressources électroniques

Cataloguing in cyberspace: the challenges of electronic documents

Catalogar el cyberspacio: el desafío de los recursos electrónicos

Roman S. Panchyshyn and France Bouthillier

Volume 43, Number 3, July–September 1997

URI: <https://id.erudit.org/iderudit/1033029ar>

DOI: <https://doi.org/10.7202/1033029ar>

[See table of contents](#)

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Panchyshyn, R. S. & Bouthillier, F. (1997). Cataloguer le cyberspace : le défi des ressources électroniques. *Documentation et bibliothèques*, 43(3), 137–147. <https://doi.org/10.7202/1033029ar>

Article abstract

Cataloging documents taken from the Internet is truly challenging. Several projects undertaken in the United States, such as OCLC Internet Resources and Interact Project, the Text Encoding Initiative (TEI), and the OCLC Dublin Core Project, helped to identify major problems. Tools such as URNs, URNs and PURLs were developed to assist cataloguers in bringing order to the chaos of Internet. The article describes those projects and tools used to identify the problems encountered by cataloguers in the course of their work. Finally, the levels of competency required by cataloguers in the future and their role in establishing standards for information exchange in the Internet community are briefly discussed.

Cataloguer le cyberspace : le défi des ressources électroniques

Roman S. Panchyshyn*

Diplômé de la Graduate School of Library and Information Studies (GSLIS)
Université McGill

France Bouthillier*

Professeure adjointe
GSLIS, Université McGill

Le catalogage des ressources accessibles dans le réseau Internet est problématique. Plusieurs initiatives américaines ont permis d'identifier des problèmes majeurs, par exemple les projets Ressources Internet et InterCat d'OCLC, le projet d'encodage de texte (TEI) et le projet Dublin Core d'OCLC. Par ailleurs, des outils tels les URC, les URN et les PURL ont été conçus pour aider les bibliothécaires à mettre de l'ordre dans le chaos existant dans Internet. L'article décrit ces projets et outils pour identifier les principaux problèmes auxquels les bibliothécaires doivent faire face dans le traitement de ces ressources. Enfin, les niveaux de compétence dont les bibliothécaires au catalogage auront besoin à l'avenir et leur rôle dans l'établissement de normes pour l'échange d'information dans la communauté Internet sont brièvement discutés.

Cataloguing in cyberspace: the challenges of electronic documents

Cataloging documents taken from the Internet is truly challenging. Several projects undertaken in the United States, such as OCLC Internet Resources and InterCat Project, the Text Encoding Initiative (TEI), and the OCLC Dublin Core Project, helped to identify major problems. Tools such as URCs, URNs and PURLs were developed to assist cataloguers in bringing order to the chaos of Internet. The article describes those projects and tools used to identify the problems encountered by cataloguers in the course of their work. Finally, the levels of competency required by cataloguers in the future and their role in establishing standards for information exchange in the Internet community are briefly discussed.

Le catalogage doit être considéré comme l'activité de base en bibliothéconomie. L'établissement et l'entretien de systèmes d'organisation de l'information sont des tâches essentielles pour garantir l'accès à l'information. Cependant, dans le contexte actuel de compressions budgétaires, les activités traditionnelles de catalogage sont remises en question. En effet, les bibliothécaires affectés au catalogage font face à la déprofessionnalisation, à l'impartition et à l'éro-

Catalogar el ciberespacio: el desafío de los recursos electrónicos

La catalogación de los recursos accesibles por Internet es problemática. Muchas iniciativas americanas han permitido de identificarlos problemas más importantes, por ejemplo los proyectos Recursos Internet et InterCat de la OCLC, el proyecto de codificación de texto (TEI) y el proyecto Dublin Core de la OCLC. Por otra parte, instrumentos como los URC, los URN et los PURL han sido concebidos para ayudar a los bibliotecarios a poner orden en el caos actual que existe en el Internet. El artículo describe estos proyectos y instrumentos para identificar los principales problemas encontrados por los bibliotecarios en el tratamiento de estos recursos. Por fin, los niveles de calificación necesarios para los bibliotecarios responsables de la catalogación en el porvenir y sus papeles en el establecimiento de normas para el intercambio de información en la comunidad Internet son brevemente discutidos.

sion des standards de catalogage (Dunkle 1996; Zyroff 1996). Quel rôle devront jouer ces bibliothécaires dans le futur?

Avec l'évolution du réseau Internet et la prolifération de ressources électroniques, l'accès à l'information subit une transformation rapide. Dans les bibliothèques, l'objectif visant à posséder et à organiser une collection locale est en train de céder la place à un autre objectif qui consiste à organiser «l'accès» à de multiples collec-

tions entreposées localement et ailleurs (Demas, McDonald and Lawrence 1995; Lougee 1995). Ce phénomène affecte le catalogage dans la mesure où de nouvelles approches doivent être développées de façon à cataloguer des documents publiés sur des supports non traditionnels

* Les auteurs remercient le professeur John E. Leide pour la supervision du projet de recherche à l'origine de cet article et Ared Misirliyan pour l'aide apportée à la traduction.

qui sont souvent stockés en dehors des limites d'une collection locale. Ces développements récents ainsi que la prolifération des périodiques électroniques ont provoqué une réflexion sur le catalogage des documents numérisés (Morgan 1995, 1996). L'information dans Internet continue de croître et de prendre de l'importance ainsi que d'acquérir une légitimité au sein de la communauté scientifique et académique. Toutefois, les utilisateurs sont souvent frustrés et confus en raison de la croissance rapide du réseau Internet et de leur difficulté à y trouver l'information pertinente rapidement. Ils doivent alors se tourner vers les bibliothécaires pour avoir de l'aide. Le défi à relever pour les responsables du catalogage est évident: comment les ressources électroniques peuvent-elles être organisées, cataloguées, et classifiées de façon à répondre aux attentes et aux besoins des utilisateurs?

Cet article résume plusieurs projets américains importants concernant le catalogage des ressources Internet et autres ressources ordinolingues et permet de voir comment les bibliothécaires américains travaillent à régler cette question. L'article cherche à identifier les facteurs et tendances qui peuvent entraîner des changements réels dans la fonction du catalogage traditionnel et dans les procédures de catalogage des ressources Internet. L'examen de ces projets permet aussi de se rendre compte des nouvelles variables et conditions qui auront un effet sur la pratique du catalogage. Enfin, l'article discute de la façon dont la profession de bibliothécaire au catalogage est affectée par ces nouveaux éléments. Bien que le texte qui suit ne relate pas de projets francophones, il devrait être utile pour dégager des conclusions pertinentes à la pratique du catalogage en milieu francophone.

Le catalogage des ressources dans Internet: termes et notions de base

S'il n'est plus nécessaire de définir le terme Internet, il est essentiel de clarifier ce que sont des «ressources Internet». Ces ressources correspondent à des livres numérisés, des journaux électroniques, des bases de données composées d'images, des encyclopédies hypermédias et multimédias, des sites Web ou des groupes de nouvelles et de discussions.

Le débat sur la définition de ressources Internet a débuté à l'été 1991 quand le Comité MARBI¹ a proposé une série d'éléments qui pouvaient être utiles pour la description des ressources en ligne. Leur caractéristique principale et commune a été définie comme étant la possibilité d'accès à distance. Les ressources en ligne, par opposition aux disques rigides ou disquettes d'ordinateur, ne peuvent être décrites physiquement, posées sur des rayons ou prêtées aux usagers. De plus, on s'entendait pour dire que l'univers des ressources accessibles à distance pouvait être divisé en deux catégories: (1) des ressources de données, par exemple des logiciels ou des bases de données numériques, bibliographiques et textuelles et (2) des systèmes ou des services, par exemple des systèmes de catalogues et des tableaux d'affichage électroniques (Caplan 1993). Ce rapport de discussion a donné lieu, en 1993, à la publication d'un autre document (*MARBI Proposal 93-4*) qui recommandait des changements au format bibliographique USMARC pour accommoder les journaux électroniques et autres documents disponibles dans Internet. Ainsi, les sites FTP (protocoles de transfert de fichiers) par exemple pouvaient être traités de la même façon que les ressources physiques comme les livres. Enfin, il a été recommandé de créer une nouvelle zone du MARC (856) pour y indiquer les adresses électroniques et les façons d'accéder aux documents dans Internet (Caplan 1993).

Par ailleurs, le catalogage des ressources Internet doit s'harmoniser avec la mission générale des bibliothèques. Toutefois, cette mission n'a pas encore vraiment été assumée en ce qui concerne les documents numérisés. Premièrement, les stratégies de développement des collections, les méthodes et les critères de sélection doivent être ajustés pour assimiler les ressources Internet. Les politiques de sélection doivent inciter les bibliothécaires à relever le défi consistant à choisir soigneusement les ressources électroniques et à intégrer des ressources accessibles à distance dans les collections de bibliothèques (Demas, McDonald and Lawrence 1995). Les politiques doivent aussi tenir compte des questions comme le droit d'auteur, l'entreposage optimal, la diffusion des documents, la préservation et les méthodes d'archivage.

Les politiques de bibliothèque, en ce qui concerne l'accès aux ressources électroniques, doivent répondre aux besoins de leurs usagers. En ce moment, les outils disponibles dans Internet qui visent à satisfaire les exigences des usagers en matière d'accès sont déficients. Cette déficience est due à divers facteurs tels que la complexité des commandes des divers moteurs de recherche (ex.: Lycos, Yahoo, Alta Vista), les limites des options de recherche de ces moteurs, la qualité du contenu des bases de données dans lesquelles ces moteurs exécutent la recherche, et la compétence du chercheur à formuler des stratégies de recherche adéquates (Venditto 1996). Jusqu'à tout récemment, les systèmes traditionnellement utilisés dans les bibliothèques, comme les systèmes de classification et les vedettes-matière qui correspondent à un vocabulaire contrôlé, n'étaient pas appliqués dans les outils de navigation Internet. Deux raisons expliquent cette situation: les bibliothécaires n'ont joué aucun rôle dans la construction de ces outils et leur culture professionnelle n'encourageait pas leur participation au développement chaotique d'Internet, surtout dans un contexte où des solutions connues ne paraissent pas être évidentes et immédiatement applicables (Flannery 1995).

Récemment, des sites comme *CyberStacks*, à la Iowa State University², ont commencé à organiser les ressources Internet avec des systèmes de classification standards et non standards ainsi qu'avec un vocabulaire contrôlé. Ces systèmes comprennent la Classification décimale Dewey (CDD), la Classification décimale universelle (CDU), la classification de la Library of Congress (LCCN) et les vedettes-matière de la Library of Congress (LCSH). Le World Wide Web Consortium (W3C), composé de diverses organisations intéressées à définir des spécifications communes pour le Web, a aussi

1. MARBI est le sigle de Machine-Readable Form of Bibliographic Information. Le Comité MARBI est responsable de la préservation et des modifications du format MARC (Machine-Readable Cataloguing) américain ou le USMARC.

2. Voir <http://www.public.iastate.edu/~CYBERSTACKS/homepage.html>. Ce site a été créé par Gerry McKiernan.

amorcé le projet PICS (Platform for Internet Content Selection). Ce projet visait à créer un système de classement permettant de fournir des moyens techniques pour définir le contenu et pour sélectionner des documents dans le Web³. Le PICS se veut un filtre pour classer les systèmes et les services Web et est une réponse à la pression politique et aux menaces de censure qui pourraient en effet menacer le libre échange de l'information dans Internet (Weibel 1996).

Une des notions les plus importantes pour le catalogage des ressources Internet est celle de « métadonnée » (*metadata*). Priscilla Caplan (1995) a défini les métadonnées comme étant des données relatives à d'autres données. Une notice catalographique est ainsi considérée comme une métadonnée. Le terme a été inventé pour désigner la documentation lisible par machine relative aux fichiers de données lisibles par machine. Dans la communauté Internet, le terme « métadonnées » a acquis le sens de données sur des ressources d'information électronique. L'expression « métadonnées bibliographiques » a un sens plus étroit qui signifie des données ou de l'information sur des objets qui contiennent des données telles que des livres (Wool 1996). Ce dernier terme est important pour comprendre l'enjeu de certains projets discutés plus loin (TEI et Dublin Core).

Aux États-Unis, l'entreprise OCLC (Online Computer Library Center) a joué un rôle de leadership dans la recherche sur les questions relatives au catalogage de l'information électronique. Situé à Dublin, dans l'État de l'Ohio, OCLC est le centre de recherche le plus important au monde en matière de réseaux de bibliothèques et de catalogues en ligne. Les chercheurs affiliés à OCLC ont contribué de façon majeure aux projets Intecat et Dublin Core. Le présent article fait le point sur les quatre projets suivants: les projets Ressources Internet et Intecat d'OCLC; le projet d'encodage de textes ou TEI (Text Encoding Initiative) avec le langage SGML et le projet Dublin Core.

Les projets Ressources Internet et Intecat d'OCLC

La première grande initiative visant à définir des règles de catalogage des res-

sources accessibles dans le réseau Internet était le projet Ressources Internet d'OCLC (*OCLC Internet Resources Project*). Ce projet a été entrepris à partir des prémisses suivantes: - Il existe beaucoup de ressources disponibles dans Internet - Ces ressources doivent être organisées pour être accessibles - La méthode la plus efficace pour accéder à ces ressources est d'utiliser des techniques et des procédures validées en bibliothéconomie et de créer des notices bibliographiques dans les catalogues en ligne actuels (Olson 1995).

Le projet a débuté en 1992 après la parution du document du Comité MARBI (*Discussion Paper #49*) détaillant les changements à apporter au USMARC pour accommoder les ressources électroniques. Le Bureau de recherche d'OCLC, financé en partie par le Département d'éducation américain, a commencé à étudier la nature de l'information disponible dans Internet (Caplan 1994; Dillon and Jul 1994; Dillon et al. 1994). Le projet comportait deux phases: la première portait sur l'étude de la quantité et du type d'information disponible dans Internet et la seconde avait pour but d'examiner les problèmes de catalogage de cette information reliés à l'utilisation des règles AACR2 (Anglo American Cataloguing Rules 2) et plus particulièrement des règles relatives aux fichiers de données lisibles par machine. Pour la deuxième phase, les chercheurs ont choisi 300 « objets » Internet. Les documents étaient groupés au hasard en 30 groupes de 10 notices chacun, puis multipliés par trois pour un total de 900 notices bibliographiques. Trente bibliothécaires bénévoles provenant de diverses bibliothèques ont été choisis. Ces bénévoles étaient priés de cataloguer le mieux possible ces notices en utilisant leurs connaissances des AACR2 et des règles relatives aux fichiers de données lisibles par machine et de noter les problèmes qu'ils rencontraient. Une fois le catalogage achevé, les notices ont été analysées avec des méthodes automatisées et manuelles (Caplan 1994).

La principale conclusion fut qu'avec certaines modifications, les règles relatives aux fichiers de données lisibles par machine, le chapitre 9 des AACR2, pouvaient être adéquates pour le traitement des ressources Internet. L'expérience a

donné lieu à trois séries de recommandations suggérant des modifications aux règles de catalogage, des modifications au format USMARC et la rédaction de directives pour le catalogage. La première série de recommandations incluait une proposition pour l'expansion de la règle 9.3B1 des AACR2 de manière à inclure des termes comme journal électronique, document électronique et base de données bibliographiques.

La deuxième recommandation proposée et acceptée a mené à l'établissement de la zone 856 du MARC. Cette zone est équivalente sur le plan conceptuel à la zone 852 utilisée pour indiquer le lieu où un document est physiquement entreposé. La zone 856 sert à indiquer le lieu des documents entreposés à distance. Par ailleurs, la zone 538, utilisée pour fournir des détails relatifs au système, est toujours valable pour indiquer le mode d'accès (Olson 1995).

La troisième proposition consistait à rédiger une série de directives pour le catalogage des ressources Internet ce qui a donné lieu au guide de Nancy Olson (1995) et à un guide du Bureau de la Library of Congress sur les normes MARC (*Guidelines for the Use of Field 856*, March 1996)⁴. La zone 856, approuvée par le Comité MARBI, est utile pour définir des éléments de données afin de permettre aux usagers d'extraire des fichiers au moyen des protocoles utilisés dans Internet et de s'abonner aux journaux et aux conférences électroniques. Malgré ce potentiel, l'utilisation de cette zone n'est pas sans problème. Un des problèmes principaux est de s'assurer que les données soient inscrites correctement et qu'elles soient exactes. Ainsi dans le cadre de l'expérience citée plus haut, les bibliothécaires n'étaient pas autant à l'aise avec les concepts ou la terminologie d'Internet qu'ils le sont maintenant et cela nuisait à la précision des données lors de leur saisie. Un autre problème réside dans la question de la permanence et de l'entretien des fichiers à distance et de leurs adresses qui

3. Voir <http://18.23.0.22/pub/WWW/PICS/Overview.html>.

4. Voir <http://lcweb.loc.gov/marc/856guide.html>.

peuvent constamment changer. De plus, dans l'expérience citée, les bénévoles devaient faire du catalogage original, et le laps de temps nécessaire pour la création de notices électroniques originales était long entraînant un problème potentiel de temps et de ressources pour une telle activité. Enfin, une autre question, qui n'est pas reliée spécifiquement à l'activité de catalogage mais peut être problématique, a été soulevée: il s'agit de l'entreposage et de l'intégrité des données. En effet, qui devrait être responsable de l'exactitude des données?

L'usage du format MARC pour le catalogage des ressources Internet comporte plusieurs avantages (Sha 1995). Ce format est en fait une norme (ANSI/NISO Z39.2) qui permet de transférer des ressources d'information d'un système à un autre. Ainsi, lors d'un transfert de données ou d'un changement de système, cette norme assure la compatibilité des données. Une notice composée en MARC fournit une structure simple et flexible permettant une description de différents niveaux de renseignements bibliographiques. Le format MARC implique une multitude de points d'accès qui sont utiles à la recherche par mots clés. Ces éléments peuvent ainsi être utilisés pour limiter la recherche. La description bibliographique complète permet aux usagers de décider si la ressource en question correspond à leurs besoins. Le format MARC rend aussi possible la modification des formats de présentation. Par exemple, une fois introduites dans un système local, les données peuvent être disposées de différentes façons pour divers usages ou selon les particularités des formats d'affichage de catalogues en ligne. Le format MARC rend ainsi possible l'intégration de notices catalogographiques de divers types de documents dans le catalogue électronique d'une bibliothèque. Enfin, l'usage du format MARC serait un moyen pour garantir un libre accès aux ressources Internet car, d'une part, les bibliothèques peuvent cataloguer ces ressources et, d'autre part, les usagers dépourvus d'accès à Internet peuvent consulter de tels catalogues.

Ayant conclu que les ressources Internet pouvaient être cataloguées en utilisant les AACR2 et le format MARC, OCLC a amorcé le projet Intecat. Ce deuxième projet visait à créer une base de données en ligne regroupant des ressource

Internet cataloguées avec les AACR2, USMARC, LCSH et des indices numériques de classification. En plus de la création d'un tel catalogue, l'objectif global était de tester et d'évaluer l'efficacité de l'utilisation de l'USMARC pour fournir la description, le lieu et les renseignements d'accès pour des documents accessibles à distance (Sha 1995).

Le projet Intecat a été subventionné en partie par le Bureau des programmes de bibliothèques du Département d'éducation américain. Au départ, la construction de la base de données devait se réaliser entre le 1^{er} octobre 1994 et le 31 mars 1996 grâce à la participation de plusieurs bibliothèques. Le catalogue Intecat est opérationnel et son développement se poursuit. Il a été intégré à d'autres services d'OCLC comme NetFirst et FirstSearch⁵. En 1996, il y avait environ 200 bibliothèques qui participaient au projet. Avec une structure semblable à celles des catalogues de la Library of Congress, d'OCLC, du Research Libraries Network et du Western Libraries Network, la base de données Intecat reçoit, sauvegarde et distribue les notices bibliographiques fournies par les bibliothèques membres. Le catalogue Intecat représente un pas important vers l'intégration de diverses ressources d'information dans les catalogues électroniques des bibliothèques. Un autre pas majeur sera l'amélioration des catalogues électroniques actuels et le développement des catalogues intégrés au WWW utilisant la norme Z39.50, ce qui facilitera l'accès à l'information pour l'utilisateur. Celui-ci pourra alors accéder directement aux autres documents catalogués par des hyperliens qui seront créés dans la zone 856. Des catalogues électroniques impliquant de telles options sont actuellement développés par plusieurs compagnies comme DRA (Data Research Associates), SIRS, WebCat, INNOPAC et autres (Sha 1995). Toutefois, le format USMARC n'inclut pas actuellement les caractères d'espacement comme le symbole tilde (~) et celui de soulignement (), qui sont fréquemment utilisés dans les adresses URL. Par conséquent, il est parfois difficile d'établir des hyperliens à partir de la zone 856 lorsque les adresses comprennent de tels symboles.

Selon Vianne Sha (1995), le projet Intecat démontre que le catalogage des ressources Internet est une tâche énorme

et complexe qui nécessite un travail d'équipe et la collaboration de différents acteurs. Ces acteurs sont les gestionnaires de bibliothèques, les spécialistes en développement de collections, les équipes de catalogage, les bibliothécaires de référence, les techniciens en réseautique, les auteurs et les rédacteurs de ressources Internet et les vendeurs de systèmes pour bibliothèques. Pour réussir à donner accès à l'information disponible sur divers supports au moyen d'un seul système d'information, soit le catalogue d'une bibliothèque, la coopération entre toutes ces personnes est essentielle. Réaliser ce catalogue électronique intégré serait le meilleur moyen pour répondre aux besoins d'information du public.

Le projet Intecat a soulevé d'autres questions, notamment celle relative à la définition de la nature des documents. Ainsi, en ce qui concerne les journaux électroniques, la question est de déterminer s'il s'agit de bases de données ou de périodiques. Des efforts sont faits pour définir la nature des journaux électroniques, leurs caractéristiques et leur catalogage⁶. Un projet important sur le catalogage et la classification de ces journaux, *Mr. Serials*, a été mené à la North Carolina State University sous la direction d'Eric Lease Morgan. Des procédures de catalogage développées dans le cadre de ce projet permettront d'automatiser la collection, l'organisation, l'archivage, l'indexation et la diffusion des périodiques électroniques. Ces procédures de catalogage facilitent la création de notices MARC à partir de documents HTML en utilisant des modèles ou *templates* (Morgan 1995; 1996). Ce mode de création de notices MARC à partir de modèles est une variation de l'approche qui consiste à élaborer des notices MARC tirées de métadonnées ou données rattachées aux documents électroniques.

5. Intecat est disponible à l'adresse suivante: <http://orc.rsch.oclc.org:6990/>. La base de données contient plus de 40 000 notices et croît à un rythme de 4 000 notices par mois environ.

6. Le manuel de CONSER (pour Cooperative Online Serials Program) consacre tout un chapitre à ce sujet; voir <http://lcweb.loc.gov/acq/conser/aboutcon.html>.

SGML et TEI (Text Encoding Initiative)

Le projet InterCat a prouvé que des normes établies comme les AACR2 et le format MARC sont adéquates et adaptables pour le catalogage de textes électroniques. Un des problèmes majeurs pour les spécialistes en catalogage qui travaillent avec des documents électroniques consiste à extraire les renseignements nécessaires pour la création de notices de bonne qualité. Contrairement à la nature statique des pages imprimées d'un livre ou d'un article de périodique, l'aspect physique des documents électroniques peut varier selon le système de traitement. Cet aspect physique comporte des caractéristiques (caractères gras, taille et type de caractère) qui sont propres à un certain système d'exploitation et à un certain logiciel. Si le contenu d'un document électronique doit être réutilisé dans un autre système ou logiciel ou style, les codes spécifiques devront alors être enlevés et de nouveaux codes devront être ajoutés pour réaliser la nouvelle disposition du texte. L'échange de documents électroniques impliquant des mécanismes de balisage fonctionne bien seulement si l'expéditeur et le receveur utilisent le même système (Gaynor 1996).

Afin d'assurer l'uniformité des documents, des efforts sont faits dans le domaine de l'édition électronique pour développer des langages visant à transformer des pages imprimées en un format électronique normalisé et transférable. Ces langages précisent comment l'encre (ou son équivalent) sera déposée sur le papier (ou autre support). Par exemple, le langage de balisage généralisé (SGML pour *Standard Generalized Markup Language*) est un langage descriptif. Il décrit la structure ou le contenu d'un document plutôt que son apparence physique sur une page ou sur un écran. Le contenu est dissocié du style de présentation et la nature de ce contenu ne dépend plus du système d'exploitation ou logiciel (Gaynor 1996).

Le langage SGML, développé d'abord en 1970 sous le nom de GML (Generalized Markup Language) est une norme internationale depuis 1986 (norme ISO, no 8879). Le SGML est un métalangage, c'est-à-dire qu'il décrit et définit d'autres langages de balisage. Il possède

deux caractéristiques: il catégorise des parties du texte au lieu de préciser les processus de traitement de texte; et il permet la définition du type de document (DTD), ce qui fournit une syntaxe normalisée pour la définition de classes de documents (Gaynor 1994; Sperberg-McQueen and Burnard 1994).

Le SGML est en train de devenir le format préféré des rédacteurs à cause de ses capacités d'analyse du contenu et de ses propriétés qui ne dépendent pas d'un système particulier (Cole and Kazmer 1995). Des documents codés en SGML peuvent être consultés par des outils de navigation Web. Un des meilleurs outils de navigation SGML mis gratuitement à la disposition des utilisateurs d'Internet est Panorama conçu par SoftQuad⁷.

Pourquoi le langage SGML est-il important pour les bibliothèques et pour le travail de catalogage? Principalement parce que le format MARC n'est pas tout-à-fait adéquat lorsqu'on essaie d'établir l'accès à plusieurs versions de documents numérisés. Des projets comme le Berkeley Finding Aids Project et le Columbia University Digital Image Access Program ont démontré les déficiences suivantes du MARC: incapacité à structurer l'information analytique non bibliographique essentielle pour repérer des documents électroniques et impossibilité d'accéder à différents niveaux d'analyse (Gaynor 1996).

La complexité du travail de catalogage est particulièrement évidente lorsqu'on considère qu'un document peut en fait engendrer plusieurs documents différents. Par exemple, un livre imprimé qui comprend des cartes géographiques en couleur, des dessins en noir et blanc et 300 pages de texte peut être transformé en: un microfilm du livre complet incluant les illustrations; un ensemble de 300 images numérisées du livre complet, sans les illustrations; des images électroniques à haute résolution des cartes couleurs faites à partir de cartes plus grandes que celles dans le livre et des reproductions numérisées des illustrations à résolution multiple, faites à partir de diapositives des originaux.

Le catalogage de toutes ces ressources devrait permettre d'établir des liens hiérarchiques indiquant la source du do-

document et les particularités de sa transformation. Alors que le format MARC exigerait la fragmentation de composantes reliées de façon hiérarchique en vue de créer des notices séparées et uniques, le langage SGML permet d'enregistrer des notices structurées hiérarchiquement, ce qui est fort utile pour les ressources électroniques. Un catalogue en ligne ne serait ainsi plus horizontal mais pourrait être hiérarchisé avec des notices reliées entre elles à plusieurs niveaux (Gaynor 1996). Pour réaliser cela, le langage SGML doit être utilisé avec des en-têtes TEI (pour Text Encoding Initiative) mises au point dans le cadre d'un projet de coopération internationale du même nom. Ce projet visait essentiellement à développer des règles pour faciliter la préparation et l'échange de textes électroniques. Une première version complète des directives (*Guidelines for Electronic Text Encoding and Interchange*) a été publiée en 1994⁸. L'objectif recherché était de mettre au point une méthode de classement compatible avec les conventions de catalogage utilisées dans les bibliothèques permettant de retracer l'histoire de la production ou de la transformation d'un texte électronique. Cette méthode devait donner les moyens d'authentifier la provenance d'un texte et les modifications faites à celui-ci (Ide and Sperberg-McQueen 1995).

L'en-tête TEI, attaché aux documents SGML, est en fait une métadonnée. Il correspond à un ensemble de détails descriptifs qui est l'équivalent de la page de titre d'un livre imprimé, laquelle comprend l'information nécessaire aux encodeurs pour documenter la description du document. Cet en-tête est pertinent pour trois groupes d'acteurs: les concepteurs de textes numériques qui doivent fournir une certaine documentation relative à leurs textes; les utilisateurs de ces textes tels que les bibliothécaires qui doivent structurer l'information pour les fins de catalogage et les chercheurs. L'en-tête TEI est composé de quatre parties fonctionnelles: la description du fichier qui documente la description bibliographique du document

7. Le logiciel de Panorama est disponible pour téléchargement à <http://www.oclc.org:5047/oclc.research/panorama>.

8. Voir <http://etext.virginia.edu/TEI.html>.

et de sa source; celle de l'encodage du texte; celle de ses caractéristiques non bibliographiques et celle des révisions du texte, c'est-à-dire les changements ou les mises à jour apportés au document. Les bibliothécaires sont particulièrement préoccupés par la section incluant la description du fichier parce qu'elle contient la description bibliographique du document et de sa source qui est essentielle pour repérer un texte (Giordano 1995).

Selon les directives publiées, la fonction des en-têtes TEI est d'assurer que l'information nécessaire pour créer une notice catalographique soit facilement repérable et que le lien entre cette notice et l'en-tête soit maintenu d'une façon aussi simple que possible (Sperberg-McQueen and Burnard 1994). Les directives expliquent aussi que les notices MARC et les en-têtes TEI n'ont pas le même rôle. 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 toute l'information bibliographique, tout comme une fiche de catalogue, mais également toute l'information non bibliographique qui est déterminante dans le traitement, réalisé par des humains ou des machines, du texte électronique. Cette information, si elle est retenue dans un enregistrement MARC, doit être mise dans des notes qui n'ont pas la structure requise pour le repérage et l'analyse par ordinateur. L'en-tête TEI, avec ses zones descriptives, peut être facilement repéré et analysé par machine et assure un lien direct avec le texte, lui-même relié aux éléments pertinents de l'en-tête (Sperberg-McQueen and Burnard 1994)

Malgré ces avantages des en-têtes TEI, le format MARC assure pour l'instant l'échange d'information entre les catalogues en ligne au moyen de la norme Z39.50. Tant que les catalogues n'auront pas d'outils de navigation pour le Web et ne seront pas compatibles avec le langage SGML, ce qui élargirait leurs potentiels, MARC restera la norme de choix. Les directives relatives aux en-têtes TEI n'ont pas le statut de normes. Cependant, la Library of Congress a entrepris, en 1996, le projet SGML-MARC qui a pour but de développer une version SGML du format MARC. En créant des notices catalographiques à la fois compatibles avec le for-

mat MARC et le langage SGML, ces notices pourraient être accessibles par des logiciels qui fonctionnent avec un des deux protocoles (Davis 1996). Ce projet vise l'établissement d'une norme MARC-DTD qui permettrait la conversion des notices MARC en notices SGML sans perte de données, ainsi que la conversion inverse (SGML vers MARC) ou, en d'autres mots, une convertibilité complète. Un test Alpha de la version MARC-DTD a été entrepris en février 1997 et les conclusions se sont avérées positives, laissant présager l'utilisation à plus grande échelle de cette version⁹.

Parallèlement à ces initiatives, plusieurs projets sont en cours en vue de cataloguer des textes électroniques encodés en SGML. Deux méritent d'être mentionnés: Electronic Texts Center (ETC) à l'University of Virginia¹⁰, auquel il est possible d'accéder grâce à un logiciel spécial permettant de visionner les documents au moyen d'outils de navigation et le Center for Electronic Texts in the Humanities (CETH) à la Rutgers University¹¹. Ces deux sites ont publié des manuels pour cataloguer des textes électroniques qui énoncent les procédures utilisées par les bibliothécaires de ces universités. À l'University of Virginia, on utilise aussi des modèles similaires à ceux utilisés dans le projet Mr. Serials, qui facilitent la création d'en-têtes TEI et leur conversion dans un format MARC.

L'objectif ultime de SGML-MARC est de permettre l'utilisation universelle de ressources électroniques à travers divers réseaux, de façon hiérarchique. Cependant, cet objectif soulève plusieurs problèmes. Premièrement, bien qu'il y ait une norme pour la définition du type de document, il n'y a pas encore de norme pour les en-têtes TEI. Ainsi, la qualité des enregistrements MARC extraits des en-têtes TEI peut varier considérablement. Si les en-têtes étaient normalisés, des programmes experts pourraient être mis au point pour créer des notices catalographiques sans intervention humaine. Mais ce processus pourrait engendrer un déclin dans la qualité d'application des normes de catalogage et rendre des outils comme les AACR2 et MARC superflus. Il serait inapproprié de suggérer la disparition du format MARC car les coûts de conversion au SGML

seraient trop élevés. Cependant, le catalogage avec le format MARC pourrait être transformé pour inclure les mêmes structures hiérarchiques qui sont possibles avec le langage SGML (Heaney 1995). La recherche de normalisation d'une métadonnée conçue comme un en-tête représente le cœur d'un autre projet d'OCLC, le Dublin Core.

Le projet Dublin Core d'OCLC

En mars 1995, OCLC et le National Centre for Supercomputing Applications (NCSA) a parrainé la première conférence sur les métadonnées (Metadata Workshop). Cette conférence devait contribuer à la compréhension des besoins, des forces, des faiblesses et des solutions préconisées par des utilisateurs tels que les bibliothécaires et les archivistes. Cette compréhension devait permettre d'établir un consensus sur les éléments essentiels des métadonnées pour une description adéquate des ressources accessibles dans Internet (Weibel, Godby and Miller 1995). Ces ressources ont été désignées comme les DLO (document-like objects), qui peuvent être assimilés aux versions électroniques de journaux, de dictionnaires, de cartes géographiques et autres, mais qui n'incluent pas des documents comme des diapositives ou des affichages de groupes de nouvelles Usenet. Les DLO, qui sont en fait les équivalents numériques des produits imprimés sur papier, étaient considérés comme les ressources les plus recherchées par les utilisateurs d'Internet.

Le projet voulait relever un défi important: offrir une aide améliorée pour le repérage des ressources DLO. Puisqu'Internet contient plus d'information que tout ce que les indexeurs et les bibliothécaires professionnels peuvent gérer en utilisant les méthodes et systèmes existants, il semblait raisonnable de donner des outils aux auteurs et aux fournisseurs d'information électronique pour leur permettre de décrire eux-mêmes leurs

9. Voir <http://www.loc.gov/marc/marcdtd/marcdtdalpha.html>.

10. Voir <http://www.lib.virginia.edu/etext/ETC.htm>.

11. Voir <http://www.ceth.rutgers.edu>.

documents et ce sans avoir une formation exhaustive sur les normes relatives à la création de notices bibliographiques. Il s'agissait donc d'identifier un ensemble d'éléments pertinents pour créer des métadonnées qui faciliteraient le repérage des documents, peu importe le sujet ou la complexité du texte (Weibel, Godby and Miller 1995).

Suite à la conférence, un ensemble de 13 éléments de base a été déterminé pour l'enregistrement de métadonnées, ensemble maintenant appelé Dublin Core. Depuis, deux éléments ont été ajoutés et d'autres ont été modifiés, de sorte que depuis janvier 1997, le Dublin Core comprend 15 éléments: Auteur ou créateur - Titre - Sujet et mots clés - Description: incluant les résumés - Éditeur - Date - Autres contributeurs: les personnes comme l'éditeur et le transcritteur qui ont contribué de façon significative au contenu intellectuel du document - Identificateur de ressources: chaîne ou nombre utilisé uniquement pour identifier l'objet - Type de ressource: le type d'objet (roman, poème ou dictionnaire) - Support: la présentation des données constituant l'objet (ex.: des fichiers PostScript) - Relation: le ou les liens avec d'autres documents - Langue: langue du contenu intellectuel - Source: les documents imprimés ou électroniques desquels l'objet est dérivé s'il y a lieu - Couverture: les caractéristiques de la localisation spatiale et temporelle de l'objet - Droits d'auteur: pour le contrôle, l'accès et la reproduction¹².

Chaque élément est optionnel, peut être répété et peut avoir des sous-éléments afin de qualifier ou d'expliquer davantage la nature des données. Par exemple, l'élément sujet pourrait avoir un sous-élément appelé zone=LCSH pour indiquer une vedette-matière de la Library of Congress ou une zone=résumé si un résumé était fourni au lieu d'une rubrique sujet. Il faut noter que le type de métadonnée normalisée proposé ne prescrit aucunement la façon d'enregistrer l'information car il se veut indépendant de la syntaxe utilisée. Les données peuvent être enregistrées en USMARC, HTML, SGML ou autrement. Tout ce qui peut être utilisé ou échangé avec d'autres systèmes est accepté (Caplan 1995).

Depuis 1995, il y eu d'autres développements relatifs au Dublin Core. Une

autre conférence s'est tenue en avril 1996 au Royaume-Uni, à la Warwick University, à la suite de laquelle une syntaxe particulière a été mise au point pour le Dublin Core et qui correspond au SGML DTD. Cette syntaxe a été reliée à des étiquettes HTML existantes de façon à incorporer des métadonnées dans les documents Web. Elle a été approuvée par le W3C à l'été de 1996. Cette conférence a aussi permis d'établir le *Warwick Framework* soit une architecture ayant le potentiel de mettre ensemble des métadonnées à structures syntaxiques diverses qui sont accessibles et maintenues séparément (Dempsey and Weibel 1996). Une autre conférence d'OCLC sur les métadonnées relatives aux images accessibles en réseau (CNI/OCLC Metadata for Networked Images), tenue en septembre 1996, a rendu possible l'utilisation du Dublin Core dans le domaine des images (Weibel and Miller 1997). Enfin, la conférence la plus récente (DC-4 Down Under) qui a eu lieu à Canberra en Australie en mars 1997, a engendré une proposition de demande formelle pour faire du Dublin Core une norme internationale.

Le Dublin Core est semblable à l'en-tête TEI parce qu'il n'est qu'un en-tête pour texte numérique, mais il ne se limite pas au langage de balisage généralisé (SGML). Cependant, il comporte quelques problèmes. Les données introduites à la source par les auteurs et non par des bibliothécaires professionnels peuvent échapper au contrôle et compromettre ainsi leur fiabilité pour les fins de repérage. Par définition, les formes d'encodage et les spécifications du Dublin Core ne seront jamais autant normalisées que ne l'est le format USMARC. Toutefois, il est possible d'imaginer que les réseaux de bibliothèque puissent importer certains éléments du Dublin Core et les transformer en notices USMARC (Xu 1996). Ce scénario a déjà été discuté dans deux rapports du comité MARBI (*Discussion Paper #86*, June 1995; *Discussion Paper #99*, January 1997)¹³.

En lien avec ce projet, le concept de *semantic interoperability* ou la capacité d'opération sémantique entre divers systèmes est en train de prendre de l'importance. Pour développer cette idée, le Dublin Core pourrait fournir un modèle suffisamment simple pour être implanté dans plusieurs systèmes descriptifs et

assez complet pour faciliter la recherche documentaire. Bien que chaque élément descriptif soit limité à un champ précis, un chevauchement sémantique crée de ce fait un noyau de champs descriptifs. Ce noyau pourrait servir aux usagers de point de départ pour commencer une recherche impliquant une variété de types de ressources accessibles sur de multiples supports (Weibel 1996). Un projet conjoint auquel participent les bibliothèques nationales d'Australie et de Nouvelle-Zélande appelé le National Document and Information Service (NDIS) vise précisément à appliquer le Dublin Core de cette façon¹⁴.

Les URC, les URN et les PURL

Plusieurs autres projets de recherche sur le catalogage des ressources Internet progressent actuellement et sont connus sous le terme général URI (Uniform Resource Identifiers) ou identificateurs uniformes de ressources puisqu'ils sont produits par le Uniform Resource Identifier Working Group of the Internet Engineering Task Force (Caplan 1996). Ce groupe travaille à développer des outils en vue d'améliorer le repérage de documents. On connaît le terme URL signifiant localisateur uniforme de ressources (Uniform Resource Locator), qui est en fait l'adresse d'un document Web. URC correspond aux caractéristiques uniformes d'une ressource (Uniform Resource Characteristic) et URN est le titre uniforme d'une ressource (Uniform Resource Name) qui a été proposé pour assigner des identificateurs uniques et indépendants des lieux des documents. Ce dernier outil est semblable aux ISBN utilisés dans l'édition (Xu 1996).

Les URC garantissent le repérage par machine des ressources et ont été proposés comme un moyen de connexion entre les URN et les URL. Si l'adresse URL change, les usagers autorisés peuvent entrer dans le service d'URC pour modifier l'adresse URL qui est reliée au URN. Cela signifie que l'URN ne change pas même si

12. Voir http://purl.org/metadata/dublin_core_elements.

13. Les deux documents sont disponibles à gopher://marvel.loc.gov/7waissrc%3a/waissrc/usmarc-d?marbi.

14. Voir <http://www.nla.gov.au/2/NDIS>.

l'adresse URL est modifiée. L'URC peut contenir des métadonnées relatives à l'auteur, au titre, à l'éditeur et au sujet, lesquelles peuvent faciliter la recherche documentaire.

De plus, l'URC peut inclure d'autres renseignements comme des signatures électroniques, de l'information sur la révision du document qui assure l'authenticité de la ressource, de l'information sur les restrictions relatives à l'utilisation du texte ou des éléments permettant de retracer l'histoire de la production ou de l'édition du document (Xu 1996). Toutes ces caractéristiques rendent l'URC très intéressant pour les éditeurs de journaux électroniques.

La dernière version des spécifications d'URC est basée sur les 15 éléments du Dublin Core, ce qui devrait encourager les auteurs et les éditeurs à fournir des renseignements sur leurs propres ressources. Les URN seraient connectés aux serveurs de recherche qui stockeraient les métadonnées. Les bibliothécaires pourraient se servir des métadonnées pour créer des descriptions plus complexes et complètes telles que des notices catalographiques et des en-têtes TEI. L'aspect important à noter est qu'il ne serait plus nécessaire de stocker les URL qui changent souvent dans la zone 856 d'une notice MARC, mais seulement les URN lesquels ne sont pas variables (Caplan 1996). Il s'agit d'une solution pertinente au problème des documents Internet qui sont déplacés, qui changent de titre ou d'adresse, ou qui disparaissent entièrement, soit le problème des cibles qui bougent (moving target problem).

Une solution intérimaire pour le problème d'URN est développée par OCLC et elle est connue sous le nom de PURL (Persistent Uniform Resource Locator)¹⁵. Du point de vue de son fonctionnement, un PURL est un URL, mais au lieu d'indiquer l'adresse d'une ressource Internet directement comme un URL, le PURL indique un service intérimaire de recherche. Le service de recherche associe le PURL avec l'URL et renvoie l'URL au client qui peut donc compléter la consultation d'une façon normale. Les logiciels utilisés par les administrateurs de PURL surveillent et mettent à jour les URL. Grâce à ce service, l'adresse fournie dans le champ 856 d'une

notice MARC indiquerait toujours le PURL et en assurerait la stabilité.

La classification des ressources Internet et l'accès thématique

La possibilité d'utiliser des systèmes de classification et des vedettes-matière pour les ressources Internet existe aujourd'hui grâce à la norme MARC, au développement des MARC-DTD et des en-têtes TEI ainsi que du Dublin Core. La classification sert d'abord à déterminer où se trouvent les documents sur les rayons, à regrouper des documents autour d'un même sujet et à fournir des balises pour leur exclusion. En ce qui concerne la recherche documentaire, la classification joue un rôle important car elle permet à un utilisateur de fureter sur les rayons, de naviguer, d'identifier, d'évaluer et de filtrer les documents. Dans le domaine de la gestion des documents, la classification est essentielle pour la gestion des fonds, la gestion des bases de données et la construction et l'entretien des thésaurus. Un environnement électronique permet l'utilisation simultanée de plusieurs systèmes de classification et, par conséquent, offre une meilleure précision et un meilleur rappel. Il existe déjà des zones dans le format MARC pour les indices de classification. Ces zones pourraient facilement être ajoutées aux MARC-DTD, aux en-têtes TEI, au Dublin Core et aux notices URC. Actuellement, il y a très peu de sites sur Internet comme celui de CyberStacks susmentionné qui se servent de systèmes de classification. Il a été proposé d'introduire des indices de classification dans les adresses URL pour créer une certaine structure dans les ressources Internet (Lincicum 1996).

OCLC a déjà amorcé des recherches sur la classification automatique des ressources accessibles en réseau. Un de leurs projets consistait à assigner automatiquement des numéros de classification Dewey, à l'aide de 1 800 vedettes-matière de la Library of Congress, à des publications pour les jeunes. OCLC a aussi entrepris des expériences visant à utiliser l'encodage SGML avec l'index de la Classification décimale Dewey. La création d'un DTD pour cet index pourrait être le premier pas vers la création ou la restructuration de systèmes de classification nouveaux ou en usage. Une autre solution

envisagée consistait à rendre disponible les outils de classification et de catalogage en formant les auteurs à classer et à regrouper leurs oeuvres dans Internet (Wyly 1996).

Les systèmes de classification décimale Dewey (CDD) et CDU sont les deux systèmes les plus faciles à adapter pour des manipulations par machine. Une version lisible par machine de CDD existe déjà car elle possède des caractéristiques importantes pour le classement de ressources Internet telles que: 1. un système de numération hiérarchique; 2. la particularité de permettre la recherche en passant du général au particulier puisqu'on peut identifier un indice numérique général qui inclut des indices spécifiques; 3. des niveaux de spécificité; 4. un filtrage efficace; 5. un contrôle des synonymes; 6. des traits mnémoniques; 7. des sous-divisions normalisées; 8. une manipulation efficace par ordinateur et 9. une possibilité d'indexation semi-automatique (Micco 1995, Part 2)

Un bon système de classification des ressources Internet doit être relié à un ensemble de vedettes-matière construit à partir d'un vocabulaire contrôlé. Bien que les bibliothécaires aient développé de multiples outils pour le contrôle de vocabulaire, entre autres LCSH et plusieurs thesauri, il n'y a pas eu jusqu'ici suffisamment d'intérêt ou de ressources financières pour automatiser ces outils. Des liens entre les indices de classification, les termes contrôlés et non contrôlés dans les résumés et les textes des documents aideraient grandement les utilisateurs à formuler des stratégies de recherche adéquates et à filtrer l'information (Micco 1995). Si un système d'autorité de contrôle des noms était en place dans Internet, les utilisateurs pourraient vérifier la forme appropriée des noms. En l'absence de système expert à cet effet, l'intervention des professionnels en catalogage demeure nécessaire.

15. Voir OCLC, PURL Frequently Asked Questions (http://purl.oclc.org/docs/purl_faq.html) pour n'importe quelle information sur PURLs.

Conclusion

L'examen de quatre projets majeurs et de plusieurs initiatives mineures a fait ressortir les défis entourant le catalogage des ressources Internet. Le projet Intecat a été un succès, compte tenu de la création d'une base de données de ressources électroniques qui contient maintenant plus de 4 000 documents. Le projet SGML-TEI a connu un certain succès malgré quelques limites. Un problème de dédoublement a été soulevé puisque les bibliothécaires doivent créer des notices MARC ainsi que des en-têtes TEI. De plus, ces en-têtes sont souvent utilisés de pair avec la numérisation de textes réalisée avec des logiciels de reconnaissance optique de caractères qui ne sont pas toujours fiables et demandent une relecture des textes. Enfin, la question des droits d'auteurs n'est pas résolue. Par conséquent, seulement les textes sans droits d'auteurs sont pour l'instant numérisés, ce qui veut dire que les meilleurs textes ainsi que les plus récents ne sont pas disponibles sous forme numérique. Le projet Dublin Core est encore en développement mais récemment, à une conférence sur le concept de métadonnée tenue à la Bibliothèque nationale du Canada (mars 1997), Stuart Weibel, d'OCLC, a annoncé que le gouvernement du Danemark implanterait bientôt le système Dublin Core pour ses publications gouvernementales ainsi que pour sa bibliographie nationale.

L'ampleur de ces projets de recherche donne une idée de la complexité du travail nécessaire à l'organisation des ressources Internet et de l'immense défi que les bibliothécaires doivent relever pour mettre de l'ordre dans le chaos actuel. Des problèmes de différents ordres existent :

- les ressources Internet étant toujours en évolution, le maintien de notices bibliographiques exige beaucoup de temps, mais cet effort est nécessaire pour que les notices continuent d'être utiles - il n'y a pas de document physique à examiner, la documentation est enfouie dans les fichiers
- les documents doivent donc être lus pour qu'on leur assigne des vedettes-matière et des indices de classification - il n'existe pas encore beaucoup de notices bibliographiques de ressources Internet regroupées dans des bases de données qui pourraient servir de modèle pour le catalogage dérivé - la question de droits

d'auteur reste problématique - l'intégrité des documents est difficile à garantir - enfin, les problèmes d'archivage et d'entreposage de ressources électroniques demeurent entiers (Weihs 1996).

Les utilisateurs d'Internet demandent déjà et revendiqueront davantage un meilleur accès aux ressources électroniques. Du point de vue des bibliothécaires, l'activité traditionnelle de catalogage est en train d'être redéfinie. Il est clair que de plus en plus de postes de bibliothécaires au catalogage sont réorientés vers le catalogage et la classification de ressources numériques. Cette réorientation est accompagnée d'un plus grand besoin de compétences chez les professionnels affectés au catalogage. Les besoins de formation initiale et d'éducation permanente sont nombreux. Pour cataloguer les ressources Internet, les bibliothécaires doivent posséder les connaissances traditionnelles: les AACR2, le format MARC, l'analyse documentaire, le contrôle d'autorité, les systèmes de classification et souvent des langues étrangères. En plus de détenir les habiletés relatives à l'utilisation de ces connaissances, ils doivent posséder aussi: des connaissances pour concevoir et développer des bases de données; une connaissance des langages de balisage tels le SGML, le HTML et autres langages de programmation et de description; une connaissance des protocoles et de la terminologie Internet; une connaissance des logiciels et des normes utilisés pour la présentation des divers documents numérisés incluant les outils de navigation, les navigateurs en mode graphique, les catalogues électroniques utilisés dans le Web, la norme Z39.50, etc.; une connaissance des enjeux légaux et commerciaux concernant l'accès, l'utilisation et l'édition des ressources électroniques; une connaissance approfondie des besoins et des exigences des usagers quant aux documents numérisés.

Les résultats d'une enquête récente aux États-Unis indiquent que seulement 6% des étudiants inscrits dans les programmes de maîtrise en bibliothéconomie choisissent le catalogage comme spécialisation (Zyoff 1996). Est-ce que ce phénomène suggère que les étudiants n'entrevoient plus les possibilités d'emploi dans ce domaine ou qu'ils ont l'impression d'être insuffisamment préparés à cette spéciali-

sation ou encore qu'ils doutent de la qualité de la formation offerte dans les programmes? Il faut effectivement se demander si les écoles de bibliothéconomie et des sciences de l'information couvrent bien tous les volets d'une formation en catalogage qui permettront aux futurs professionnels d'exceller dans ce domaine.

Un des aspects les plus importants du catalogage des ressources électroniques est la coopération nécessaire à la création et au maintien de normes internationales de catalogage. Cette coopération est la clé qui permettra l'échange efficace de l'information à travers la communauté Internet. Les bibliothécaires ont toujours été engagés dans l'établissement de normes. Avec la prolifération des ressources Internet, de nouvelles normes devront être créées et celles qui existent devront être adaptées et continuellement mises à jour. La Library of Congress et OCLC prennent les devants pour développer ces normes, mais l'engagement des bibliothécaires dans l'application de celles-ci sera vraiment l'élément déterminant pour l'organisation de l'information électronique. Les bibliothécaires ont donc la responsabilité professionnelle d'apporter leur expertise et leurs connaissances dans le réseau Internet. Toutefois, tout indique que leur rôle dans cet univers en matière de catalogage sera transformé. Étant donné la prolifération des ressources Internet et l'immense travail de catalogage original nécessaire à l'organisation de ces ressources, la participation des usagers d'Internet au processus de catalogage et de classification de ces documents semble être une tendance inévitable. En fait, si le phénomène Internet est en train d'éliminer la distinction entre producteur et utilisateur d'information, car l'un peut être l'autre, nous pouvons nous demander si nous ne sommes pas en train d'assister également à l'élimination de la distinction entre producteur/utilisateur et organisateur d'information. Il semble d'ores et déjà que les bibliothécaires en catalogage auront donc un rôle de formation important auprès de la communauté des internautes qui consistera à les former à l'utilisation de systèmes de classification et de normes de catalogage. On peut ainsi imaginer une situation où les bibliothécaires devront non seulement former le public à utiliser les ressources Internet mais devront aussi former un autre public à organiser ces

ressources, condition essentielle pour permettre leur exploitation maximale. Cette fonction peut impliquer la création de bibliothécaires virtuels assurant une telle formation dans le cyberspace. Les projets décrits démontrent que l'organisation de l'information dans le réseau Internet est possible. Déterminer le rôle des bibliothécaires et celui des autres acteurs impliqués de même que les connaissances et habiletés dont ils auront besoin pour relever ce défi est donc de toute première importance.

Sources consultées

- Basili, Carla. 1995. Subject searching for information: What does it mean in today's Internet environment? *Electronic Library* 13(5): 459-466.
- Beck, Melissa et al. 1996. *Module 31: Remote access computer file serials*. Washington, DC: CONSER, Library of Congress. (URL = <http://lcweb.loc.gov/acq/conser/module31.html>)
- Beckett, David. 1995. *IAPA templates in use as Internet metadata*. Canterbury: University of Kent, Computing Laboratory. (URL = <http://www.hensa.ac.uk/tools/www/iapafatools/paper/paper.html>)
- Brugger, Judith M. 1996. Cataloging for digital libraries. *Cataloging & Classification Quarterly* 22 (3/4): 59-74.
- Burnard, Lou. 1995. What is SGML and how does it help? *Computers and the Humanities* 29 (1): 41-50.
- Caplan, Priscilla. 1993. Cataloging Internet resources. *Public-Access Computer Systems Review* 4 (2): 61-66. (URL = gopher://info.lib.uh.edu/00/articles/e-journals/uhlbrary/pacsreview/v4/n2/caplan.4n2)
- . 1994. Controlling E-Journals: The Internet resources project, cataloging guidelines, and USMARC. *Serials Librarian* 24 (3-4): 103-111.
- . 1995. You call it corn, we call it syntax-independent metadata for document-like objects. *Public Access Computer Systems Review* 6 (4). (URL = <http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>)
- . 1996. U-R-stars: Standards for controlling Internet resources. *Serials Librarian* 28 (3/4): 239-246.
- Caplan, Priscilla and Rebecca Guenther. 1996. Metadata for Internet resources: The Dublin Core metadata elements set and its mapping to USMARC. *Cataloging & Classification Quarterly* 22 (3/4): 43-58.
- Chan, Lois Mai. 1996. Classification, present and future. *Cataloging & Classification Quarterly* 21 (2): 5-17.
- Clark, Mae M., Michael D. Esman and Claudia V. Weston. 1994. Cataloging challenges: providing bibliographic access to Florida's full-text electronic state documents. *Cataloging & Classification Quarterly* 18 (3-4): 97-119.
- Cochenour, Donnice. 1994. Linking remote users and information: Cataloging Internet publications. *Colorado Libraries* 20 (Fall): 20-23.
- Cole, Timothy W. and Michelle M. Kazmer. 1995. SGML as a component of the digital library. *Library Hi Tech* 13 (4): 75-90.
- Davis, Stephen Paul. 1996. *SGML-MARC: Incorporating library cataloging into the TEI environment*. Presentation originally delivered on March 23, 1996 at the workshop on «The Text Encoding Initiative and Guidelines and their Application to Building Digital Libraries», held in conjunction with the *First ACM International Conference on Digital Libraries*, 1996, Bethesda Md. New York: Columbia University. (URL = <http://www.columbia.edu/cu/libraries/inside/projects/sgml/sgmlmarc/davis.9603.html>)
- Demas, Samuel, Peter McDonald and Gregory Lawrence. 1995. The Internet and collection development: Mainstreaming selection of Internet resources. *Library Resources & Technical Services* 39 (3): 275-290.
- Dempsey, Lorcan and Stuart L. Weibel. 1996. The Warwick metadata workshop: A framework for the deployment of resource description. *D-Lib Magazine*. (URL = <http://www.dlib.org/dlib/july96/07weibel.html>)
- Dillon, Martin and Eric Jul. 1994. Assessing information on the Internet: Toward providing library services for computer-mediated communication. *OCLC Systems and Services* 10: 86-92.
- . 1996. Cataloging Internet resources: The convergence of libraries and Internet resources. *Cataloging & Classification Quarterly* 22(3/4): 197-206.
- Dillon, Martin et al. 1994. The OCLC Internet resources project: Toward providing library services for computer-mediated communication. *Clinic on Library Applications of Data Processing (30th: 1993: University of Illinois at Urbana-Champaign)*. Urbana: University of Illinois at Urbana-Champaign, Graduate School of Library & Information Science: 54-69.
- Drabenstott, Karen Markey and Pauline Atherton Cochrane. 1994. Improvements needed for better subject access to library catalogs via the Internet. *Clinic on Library Applications of Data Processing (30th: 1993: University of Illinois at Urbana-Champaign)*. Urbana: University of Illinois at Urbana-Champaign, Graduate School of Library & Information Science: 70-83.
- Dunkle, Clare B. 1996. Outsourcing the catalog department: A meditation inspired by the business and library literature. *Journal of Academic Librarianship* 22 (1): 33-44.
- Fattig, Karl and Rick Block. 1997. Cataloging Internet resources: Policy, procedure, and practice. *Technicalities* 17(1): 1, 3-6.
- Flannery, Melinda Reagor. 1995. Cataloging Internet resources. *Bulletin of the Medical Library Association* 83(2): 211-215.
- Florida, Luciano. 1996. The Internet: Which future for organised knowledge, Frankenstein or Pygmalion? Part 1. *Electronic Library* 14 (1): 43-48.
- . 1996. The Internet: Which future for organised knowledge, Frankenstein or Pygmalion? Part 2. *Electronic Library* 14 (1): 49-52.
- Gaynor, Edward. 1994. Cataloging electronic texts: The University of Virginia library experience. *Library Resources & Technical Services* 38(4): 403-413.
- . 1996. From MARC to markup: SGML and online library systems. *ALCTS Newsletter* 7 (2): Supplement, A-D.
- Giordano, Richard. 1994. The Documentation of electronic texts using text encoding initiative Headers: An Introduction. *Library Resources & Technical Services* 38 (4): 389-401.
- . 1995. The TEI header and the documentation of electronic texts. *Computers and the Humanities* 29 (1): 75-84.
- Heaney, Michael. 1995. Object-oriented cataloging. *Information Technology and Libraries* 14 (3): 135-153.
- Heery, Rachel. 1996. Review of metadata formats. *Program* 30 (4): 345-373.
- Hockey, Susan. 1993. Developing access to electronic texts in the humanities. *Computers in Libraries* 13 (3): 41-43.
- Hoogcarpsel, Annelies. 1994. *Guidelines for cataloging monographic electronic texts at the Center for electronic texts in the humanities*. New Brunswick, NJ: CETH. (Postscript: URL = <http://cethmac.princeton.edu/docs/catguid.ps>)
- Hruska, Martha. 1995. Remote Internet serials in the OPAC? *Serials Review* 21 (4): 68-70.
- Hsieh-Yee, Ingrid. 1996. Modifying cataloging practice and OCLC infrastructure for effective organization of Internet resources. *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, Jan. 19, 1996*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/collq/hsieh.htm>)
- Ide, Nancy M. and C.M. Sperberg-McQueen. 1995. The TEI: History, goals, and future. *Computers and the Humanities* 29 (1): 5-15.
- Leahy, Sheila and Richard J. Smith. 1992. A suggested guide and comments for cataloging electronic files. *Technicalities* 12: 8-11.

- Lincicum, Shirley. 1996. Critical appraisal of the use of classification in the future: Non-traditional uses of classification: Report of a panel discussion. *Cataloging & Classification Quarterly* 21 (2): 87-89.
- Lougee, Wendy P. 1995. Beyond access: New concepts, new tensions for collection development in a digital environment. *Collection Building* 14 (3): 19-25.
- Mandel, Carol A. and Robert Wolven. 1996. Intellectual access to digital documents: Joining proven principles with new technologies. *Cataloging & Classification Quarterly* 22 (3/4): 25-42.
- Micco, Mary. 1995. Subject authority control in the world of the Internet (Part 1). *ALCTS Newsletter* 6 (5): Supplement, A-D.
- . 1995. Subject authority control in the world of the Internet (Part 2). *ALCTS Newsletter* 6 (6): Supplement, A-D.
- Molholt, Pat. 1996. Qualities of classification schemes for the information superhighway. *Cataloging & Classification Quarterly* 21 (2): 19-22.
- Morgan, Eric Lease. 1995. Adding Internet resources to Our OPACs. *Serials Review* 21 (4): 70-72.
- . 1995. Description and evaluation of the 'Mr. Serials' process: Automatically collecting, organizing, archiving, indexing, and disseminating electronic serials. *Serials Review* 21 (4): 1-12.
- . 1996. Mr. Serials revisits cataloging: Cataloging electronic serials and Internet resources. *Serials Librarian* 28 (3/4): 229-238.
- Olson, Nancy B. 1995. *Cataloging Internet resources: A manual and practical guide*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/9256cat/toc.htm>)
- Price-Wilkin, John. 1994. Using the World-Wide Web to deliver complex electronic documents: implications for libraries. *The Public-Access Computer Systems Review* 5 (3): 5-21. (URL = gopher://info.lib.uh.edu/00/articles/e-journals/uhlibrary/pacsreview/v5/n3/pricewil.5n3)
- Reynolds, Regina. 1995. Inventory list or information gateway? The role of the catalog in the digital age. *Serials Review* 21 (4): 75-77.
- Rosenfeld, Louis B. 1994. Guides, clearinghouses, and value-added repackaging: Some thoughts on how librarians can improve the Internet. *Reference Services Review* 22 (4): 11-16.
- Seaman, David M. 1996. Selection, access, and control in a library of electronic texts. *Cataloging & Classification Quarterly* 22 (3/4): 75-84.
- Sha, Vianne T. 1995. Cataloging Internet resources: The Library approach. *Electronic Library* 13 (5): 467-476.
- Sha, Vianne T., Timothy B. Patrick and Thomas R. Kochtanek. 1996. The traditional library and the national information infrastructure. *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, Jan. 19, 1996*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/colloq/sha.htm>)
- Shieh, Jackie. 1996. Does it really matter?: The cataloging format, the sequential order of note fields, and the specifics of field 856. *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, Jan. 19, 1996*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/colloq/shieh.htm>)
- Sperberg-McQueen, C.M. and Lou Burnard. 1995. The design of the TEI encoding scheme. *Computers and the Humanities* 29 (1): 17-39.
- Sperberg-McQueen, C.M. and Lou Burnard, éditeurs. 1994. *Guidelines for electronic text encoding and interchange*. Chicago: Text Encoding Initiative. (URL = <http://etext.virginia.edu/TEI.html>)
- Thorburn, Colleen. 1992. Cataloging remote electronic journals and databases. *The Serials Librarian* 23 (1-2): 11-23.
- Venditto, Gus. 1996. Search engine showdown. *Internet World* 7 (5): 79-86.
- Vizine-Goetz, Diane. 1996. Using library classification schemes for Internet resources. *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, Jan. 19, 1996*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/colloq/v-g.htm>)
- Weibel, Stuart. 1996. The changing landscape of networked resource description. *Library Hi Tech* 14 (1): 7-10.
- Weibel, Stuart and Eric Miller. 1997. Image description on the Internet: A summary of the CNI/OCLC Image Metadata Workshop. *D-Lib Magazine*. (URL = <http://www.dlib.org/dlib/january97/oclc/01weibel.html>)
- Weibel, Stuart, Jean Godby and Eric Miller. 1995. *OCLC/NCSA Metadata Workshop Report*. Dublin, Ohio: OCLC. (URL = http://www.oclc.org:5047/oclc/research/publications/weibel/metadata/dublin_core_report.html)
- Weih, Jean. 1996. Solving the Internet cataloging nightmare. *Technicalities* 16 (4): 4-6.
- Wool, Gregory James. 1996. Bibliographical metadata; or, we need a client-server cataloging code! *Proceedings of the conference Finding Common Ground: Creating a Library of the Future Without Diminishing the Library of the Past (March 30-31, 1996, Cambridge, MA)*. Ed. by Cheryl LaGuardia and Barbara A. Mitchell. New York: Neal Schuman. Available via e-mail at listserv@ubvm.cc.buffalo.edu with command «get metadata report».
- Wyly, Brendan. 1996. What lies ahead for classification in information networks? Report of a panel discussion. *Cataloging & Classification Quarterly* 21 (2): 75-82.
- Xu, Amanda. 1996. Accessing information on the Internet: feasibility study of USMARC format and AACR2. *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, Jan. 19, 1996*. Dublin, Ohio: OCLC. (URL = <http://www.oclc.org/oclc/man/colloq/xu.htm>)
- Zyoff, Ellen. 1996. Cataloging is a prime number. *American Libraries* 27 (5): 47-50.