

Le catalogage des sites web : vers une organisation structurée du web

A Catalogue of Web Sites: Towards a Structured Organisation of the Web

La catalogación de los sitios web: hacia una organización de la red

Laurence Dupourqué

Volume 49, numéro 2, avril-juin 2003

Culture et technologie

URI : <https://id.erudit.org/iderudit/1030239ar>

DOI : <https://doi.org/10.7202/1030239ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

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

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Dupourqué, L. (2003). Le catalogage des sites web : vers une organisation structurée du web. *Documentation et bibliothèques*, 49(2), 63-68.
<https://doi.org/10.7202/1030239ar>

Résumé de l'article

Les bibliothèques ont le privilège d'offrir des catalogues qui indiquent d'une manière structurée aux usagers les titres possédés alors que sur Internet, lieu virtuel et infini de la recherche documentaire, règne l'anarchie. L'élaboration de catalogues sur le Net est laborieuse étant donné la quantité et la mouvance des ressources. Les bibliothécaires, qui appliquent de plus en plus leur expertise dans le monde de l'information électronique, se lancent dans un nouveau concept de bibliothèque sur Internet : la webthèque, un répertoire structuré de ressources électroniques accessibles. Contrairement à la majorité des outils de recherche sur Internet, la webthèque tend à la qualité plutôt qu'à l'exhaustivité. Elle est comme un filtre retenant les ressources pertinentes sur un sujet précis ou à l'attention d'un public particulier. Chaque ressource est sélectionnée, cataloguée et diffusée selon les principes de la bibliothéconomie, au même titre qu'un document traditionnel. Quelles sont les caractéristiques propres aux ressources électroniques en ligne ? Quels sont les avantages et les inconvénients reliés à leur catalogage ? Comment les traiter selon des normes ? Et enfin, quel est l'avenir du catalogage de ce genre de documents ?

Le catalogage des sites web : vers une organisation structurée du web

Laurence Dupourqué

Bibliothécaire au traitement documentaire
Services documentaires multimédias (SDM)
laurence.dupourque@sdm.qc.ca

Les bibliothèques ont le privilège d'offrir des catalogues qui indiquent d'une manière structurée aux usagers les titres possédés alors que sur Internet, lieu virtuel et infini de la recherche documentaire, règne l'anarchie. L'élaboration de catalogues sur le Net est laborieuse étant donné la quantité et la mouvance des ressources. Les bibliothécaires, qui appliquent de plus en plus leur expertise dans le monde de l'information électronique, se lancent dans un nouveau concept de bibliothèque sur Internet : la webthèque, un répertoire structuré de ressources électroniques accessibles. Contrairement à la majorité des outils de recherche sur Internet, la webthèque tend à la qualité plutôt qu'à l'exhaustivité. Elle est comme un filtre retenant les ressources pertinentes sur un sujet précis ou à l'attention d'un public particulier. Chaque ressource est sélectionnée, cataloguée et diffusée selon les principes de la bibliothéconomie, au même titre qu'un document traditionnel. Quelles sont les caractéristiques propres aux ressources électroniques en ligne ? Quels sont les avantages et les inconvénients reliés à leur catalogage ? Comment les traiter selon des normes ? Et enfin, quel est l'avenir du catalogue de ce genre de documents ?

A Catalogue of Web Sites: Towards a Structured Organisation of the Web

Libraries build catalogues that help users find, in a structured fashion, the titles on hand, whereas on the Internet, a virtual and infinite resource for research, anarchy reigns. The creation of catalogues on the Net is a labour-intensive undertaking given the quantity of information and the fact that it is constantly changing. Librarians are increasingly applying their expertise to electronic information and are they becoming involved in a new concept of the library on the Internet. The Web library is a structured directory of electronic resources accessible on the Web. Unlike the majority of research tools available on the Internet, the Web library strives for quality rather than exhaustivity. It acts as a filter, retaining only those resources relevant to a specific subject or to a particular audience. Just as with the more traditional forms of documents, each resource is selected, catalogued and disseminated according to the principles of library science. What are the specific characteristics of electronic resources available on-line? What are the advantages and the limits of cataloguing them? Should they be catalogued according to the current standards? Lastly, what does the future hold for this type of catalogue?

La catalogación de los sitios web: hacia una organización de la red

Las bibliotecas tienen el privilegio de ofrecer catálogos que indican en forma estructurada a los usuarios los títulos que poseen; en Internet, lugar virtual e infinito de búsqueda documental, reina la anarquía. La elaboración de catálogos en la red es laboriosa, dado la cantidad y la naturaleza cambiante de los recursos. Los bibliotecarios, que aplican cada vez más sus conocimientos y experiencia al mundo de la información electrónica, se lanzan en un nuevo concepto de biblioteca en Internet: la «redteca», un repertorio estructurado de recursos electrónicos accesibles. Contrariamente a la mayoría de los útiles de búsqueda en Internet, la redteca tiende a subrayar la calidad más que la cantidad. Funciona como un filtro que retiene los recursos pertinentes a un tema preciso o para un público en particular. Cada recurso se selecciona, se cataloga y se difunde según los principios de la biblioteconomía, como un documento tradicional. ¿Cuáles son las características propias de los recursos electrónicos en línea? ¿Cuáles son las ventajas y las desventajas ligadas a su catalogación? ¿Cómo tratar los recursos, según las normas? Y, finalmente, ¿qué futuro tiene la catalogación de este tipo de documentos?

Caractéristiques des ressources Internet accessibles par le Web

Accès en ligne

Les ressources Internet sont des documents électroniques accessibles à distance. Elles n'offrent donc pas de support tangible à ceux qui les consultent. Elles se différencient des ressources électroniques à accès local, c'est-à-dire des ressources publiées sur un support matériel, tel un cédérom ou telle une dis-

quette, que l'on doit insérer dans un périphérique relié à un ordinateur. Il est à noter que les disques compacts musicaux, les vidéodisques et tous les supports qui ne nécessitent pas l'utilisation d'un ordinateur ne sont pas considérés ici comme des ressources électroniques.

Les sites Web sont des ressources Internet accessibles par le Web via une adresse URL (*Uniform Resource Locator*). Il existe d'autres sortes de ressources Internet, tels les sites FTP, Telnet ou telles les listes de discussion. Les sites Web se divisent en différents types: pages personnelles, commerciales ou d'organis-

mes, monographies et périodiques électroniques, catalogues bibliographiques, bases de données, etc.

Nature hybride

Les sites Web sont des documents qui s'apparentent à la fois aux ressources électroniques par leur aspect informatique, aux périodiques par leur caractère évolutif et aux monographies par leur contenu textuel.

La ressource Internet peut être traitée comme un fichier informatique si on catalogue le système plutôt que le contenu.

Le catalogage des sites Web est une pratique nouvelle et peu de normes spécifiques ont à ce jour été établies.

Elle peut aussi être traitée comme un document textuel lorsque l'importance est accordée au contenu plutôt qu'au système lui-même.

Caractère évolutif

La notice catalographique d'une monographie est le reflet exact du document qu'elle décrit tel qu'il était le jour où il a été catalogué. Comme il s'agit d'un document statique, la notice ne se périmé pas avec le temps. Les ressources Internet sont des ressources instables. Un site Web peut changer de serveur, d'adresse URL, de titre, d'éditeur, de lieu de publication et aussi son contenu peut être remanié au point de rendre inappropriés les vedettes-matière, l'indice de classification ou le résumé de la notice.

Et finalement, le site peut disparaître entièrement.

Le problème de mise à jour peut être plus important si une section du site seulement ou une page particulière est cataloguée. L'URL de la page d'accueil est relativement constante, car il s'agit de l'adresse principale qui tient lieu parfois de marque de commerce. Par contre, les adresses des pages connexes peuvent changer plus fréquemment à l'occasion de mises à jour.

Pour procéder à la mise à jour des adresses URL, il existe des logiciels qui vérifient automatiquement les liens, mais ils ont leurs limites. En effet, il peut arriver que l'adresse URL soit fonctionnelle, mais que le contenu du site ait changé. Les logiciels de vérification de liens ne peuvent détecter ces changements.

Hétérogénéité

Aucune normalisation n'est en usage actuellement pour la conception des sites Web. Comme tout le monde peut publier sur le Net, la variété des ressources est infinie et le manque d'uniformisation flagrant. La qualité des sites Web est inégale

autant du point de vue technique que documentaire. Les sites sont conçus au gré de l'inspiration plus ou moins professionnelle de leurs concepteurs. Les informations ne sont jamais mentionnées au même endroit, obligeant le catalogueur à fouiller de page en page pour trouver un nom d'auteur, d'éditeur, une date ou parfois même un titre. Ces informations primordiales peuvent même être totalement absentes. Ce genre de problème est très rare dans le traitement des documents imprimés, car ceux-ci suivent un processus de contrôle et de normalisation avant d'être publiés.

Le traitement documentaire des sites Web

Toutes les caractéristiques particulières énoncées ci-dessus augmentent considérablement le temps nécessaire au traitement documentaire des sites Web. Le catalogage de ce type de ressources prend deux à trois fois plus de temps qu'un document traditionnel, sans compter le temps passé à l'entretien et à la mise à jour des notices catalographiques. C'est pourquoi le coût élevé relié au traitement de ces ressources se révèle être un inconvénient majeur pour la plupart des bibliothèques voulant intégrer les ressources Internet à leur collection.

La Bibliothèque nationale du Québec (BNQ) a mis sur pied, depuis 2001, une webthèque de référence appelée *BREF*¹, entièrement conçue par des bibliothécaires professionnels. Il s'agit actuellement de la seule webthèque francophone dont les ressources soient sélectionnées et cataloguées selon les techniques et les règles standardisées de la bibliothéconomie. En plus d'être un répertoire de recherche accessible au grand public, *BREF* (qui dans un proche avenir sera intégrée au catalogue général de la BNQ) a pour objectifs de fournir aux bibliothèques les notices catalographiques en format MARC des sites Web sélectionnés et de

permettre l'intégration de ces notices dans les catalogues d'accès. Il sera donc possible aux bibliothèques qui désirent ajouter les ressources du Web à leur catalogue de dériver les notices à partir de *BREF*. Les usagers pourront ainsi trouver en une seule recherche par sujets ou par indices de classification tous les documents disponibles dans une bibliothèque (livres, périodiques, vidéos, cédéroms, etc.) et ceux accessibles en ligne. Ils n'auront plus besoin de consulter divers catalogues ou moteurs de recherche pour accéder à l'information, ce qui leur procurera une économie de temps et d'énergie appréciable dans un monde où l'accès à l'information doit être rapide et efficace.

Catalogage descriptif

Le catalogage des ressources électroniques, tels les logiciels et les cédéroms, ainsi que des périodiques en ligne est aujourd'hui pratique courante et effectué conformément aux normes élaborées dans le chapitre 9 (et dans le chapitre 12 pour les périodiques) des RCAA2R et celles de ISBD-ER. Mais il s'agit là de ressources statiques, même pour les publications en séries électroniques qui ont un caractère moins instable que les sites Web. Les ressources électroniques à accès local et celles à accès en ligne ont des caractéristiques communes, mais aussi quelques différences dues à la nature dynamique des ressources Internet. Le catalogage des sites Web est une pratique nouvelle et peu de normes spécifiques ont à ce jour été établies. L'expérience et le jugement du catalogueur sont par conséquent primordiaux. Chaque ressource est un défi qui engendre de nouveaux questionnements, ce qui nécessite une attitude dynamique et flexible.

1. *BREF* (Bibliothèque de RÉférence électronique), environ 1 500 titres, nouveautés régulières, accès gratuit sur le site de la BNQ : <http://bref.bnquebec.ca> (notices catalographiques en format MARC non disponibles pour le moment).
2. Le chapitre 12 des RCAA2R pour le catalogage des publications en série est actuellement en cours de révision par le *Joint Steering Committee* (JSC), comité directeur de révision des règles de catalogage anglo-américaines. Le contenu du chapitre va être élargi pour intégrer des règles sur la description des ressources à fonction d'intégration, qui comprennent les publications avec des mises à jour sur feuilles mobiles ainsi que les sites Web. Ces règles tiendront compte des caractéristiques spécifiques de ces ressources et de leur nature évolutive. Stewart, Margaret. Mars-avril 2002. Mettre le savoir à la portée de la main : résumé de la réunion du JSC pour la révision des RCAA2. *Bulletin de la Bibliothèque nationale du Canada* : vol. 34, n° 2, p. 13.

Les normes

Les normes utilisées pour le catalogage des sites Web sont les suivantes :

- règles de catalogage RCAA2 rév. 1998 : chapitre 9 (ressources électroniques) modifications de 2001 ; chapitre 1 (généralités) ; chapitre 2 (monographies) ; chapitre 12 (publications en série)² ;
- *International Standard Bibliographic Description for Electronic Resources* (ISBD-ER) ;
- MARC 21.

Les sources d'information

L'information d'un site Web est beaucoup plus dense et développée que celle d'un document traditionnel. Les liens hypertextuels sont comme des miroirs qui se reflètent à l'infini. Le catalogueur doit aller de clic en clic pour prendre connaissance de la ressource. Il faut donc savoir cadrer l'information, définir ce qui appartient au site et est propice au traitement. Comme les informations changent vite, il est préférable de ne pas mettre des informations trop détaillées dans la notice, par exemple un chiffre précis dans le résumé ou un dépouillement trop exhaustif.

En plus des principes de base de catalogage, il est nécessaire de bien connaître le fonctionnement d'Internet et de ses outils. Certaines informations (date, éditeur, auteur) peuvent être trouvées grâce aux navigateurs Internet Explorer ou Netscape, mais il est toujours important de faire preuve de discernement pour bien les interpréter. Il peut être utile aussi de savoir déchiffrer l'URL. Par exemple, l'URL de la page d'accueil se termine généralement par «index.html», un site commercial par «com», un site d'organisme par «org», etc.

La source principale d'information est la ressource elle-même. Les informations nécessaires au catalogage peuvent être repérées dans les pages suivantes : écran d'ouverture, page d'accueil, pages de présentation du site (Qui sommes-nous, À propos de, Crédits, FAQ, etc.), page de l'historique, page de coordonnées (Contactez-nous, Courrier, Pour nous écrire, etc.). Le fichier source HTML³, qui fournit les métadonnées encodées rattachées au site, est aussi une source importante. Il faut privilégier la source qui fournit l'infor-

mation la plus complète. Les crochets ne sont utilisés que lorsque les données sont prises en dehors de la ressource, sur un autre site ou dans tout autre document.

L'écran d'ouverture ou écran-titre, qui est la page sur laquelle ouvre l'adresse URL, est considéré comme l'équivalent de la page de titre pour les livres. Cela peut être soit la page d'accueil (ou écran principal) soit la page de la section du site à cataloguer le cas échéant (qu'on nommera alors l'écran-titre).

Les zones MARC spécifiques au catalogage de sites Web à contenu textuel

Zones de contrôle Guide/06 – Type de notice a (matériel textuel)

Une ressource est codée en fonction de l'aspect le plus significatif du contenu. Le code *a* est utilisé pour les sites Web considérés comme des documents textuels. Pour les sites dont l'aspect électronique prédomine, par exemple des jeux multimédias, le code utilisé sera *m*.

- **Guide/07 – Niveau bibliographique *m*** (document monographique)
Les sites Web ne sont pas considérés actuellement comme des périodiques, même s'ils en ont l'aspect dynamique.
Les zones suivantes sont utilisées en complément pour rendre compte de l'aspect informatique lorsque le code *a* est utilisé en **G/06** :
- **007 – Description matérielle** : code *c* pour fichier informatique à la position 00 (catégorie de document) et code *r* pour accès à distance à la position 01 (indicateur spécifique du genre de document).
- **006 – Caractéristiques matérielles additionnelles** : code *m* pour fichier informatique à la position 00 (genre de matériel).
- **008 – Zones spécifiques pour les livres** (à utiliser lorsque le G/06 est codé *a*) : le code *s* à la position 23 spécifie que le support matériel du document est un support électronique.

Zones variables obligatoires

- **256 – Indication du type de ressource électronique**
Selon ISBD-ER, la mention « Données électroniques textuelles » doit être utilisée. Cependant, le terme *électroniques* est facultatif s'il est déjà mentionné dans l'indication générale du genre de document à la suite du titre propre dans la zone 245.
- **500 – Source du titre propre**
Si la ressource est cataloguée à partir de la page d'accueil, on peut noter « Titre de la page d'accueil » ; si la ressource est cataloguée à partir d'une section du site, on notera alors « Titre de l'écran-titre ». Cette distinction permet de savoir si le catalogage est fait à partir du site entier ou seulement d'une section. On ajoute à cette note la « mention de date » à laquelle le site a été consulté et catalogué. Étant donné que les informations ne sont pas stockées de façon permanente, cette précision permet à l'utilisateur de situer la pertinence des données de la notice par rapport à la date où il la consulte.
- **538 – Note sur les particularités du système**
Mode d'accès (Web, Telnet, FTP, etc.) pour les ressources électroniques accessibles à distance. Dans le cas des sites Web, la note sera rédigée comme suit : « Mode d'accès : Web »
Cette zone est complémentaire de la suivante.
- **856 – Emplacement et accès électroniques**
Adresse URL (*Uniform Resource Locator*), qui localise la ressource électronique en ligne.

Caractéristiques des autres zones variables utilisées

- **082 – Indice de classification Dewey**
L'indice de classification permet de mettre de l'ordre dans l'anarchie du Web en obtenant une vue d'ensemble hiérarchique par grandes classes et sous-classes. Il permet aussi une meilleure intégration de la notice à un catalogue de bibliothèque. La recherche par indices dans une webthèque

3. Dans Internet Explorer, l'accès au fichier source se fait par le menu Affichage, puis Source.

Le catalogage par sujets est la méthode la plus efficace pour organiser et rendre accessibles aux bibliothèques les ressources sur le Web.

équivalent au furetage sur les rayons d'une bibliothèque.

L'indice Cutter n'est pas ajouté car il n'est utile que pour le classement sur les étagères.

► **245 – Titre et mention de responsabilité**

Le titre d'une monographie ou d'un périodique est en général évident et sans ambiguïté. Le titre d'un site Internet peut être difficile à déterminer. Si la page à cataloguer est une section d'un site, il faut tenir compte du titre de la rubrique (titre de partie) et du titre du site (titre propre).

L'indication générale du genre de document est [ressource électronique] selon ISBD-ER.

Le webmestre technique, qui est responsable de la conception technique d'un site, n'est pas considéré comme l'auteur de la ressource. Par contre, le webmestre éditorial ou édimestre, responsable du contenu du site, sera mis en mention de responsabilité. Un webmestre peut être le concepteur à la fois intellectuel et technique du site. Il sera alors mis en mention de responsabilité et souvent aussi en mention d'éditeur.

Si les noms d'individus sont reliés à un nom de collectivité, il est préférable d'indiquer le nom de la collectivité en mention de responsabilité, car il sera moins susceptible de changer. Les crochets pour les inexactitudes dans les mots (titre ou mention de responsabilité) sont rendus inutiles à cause des fréquentes mises à jour et corrections faites aux sites. Cependant, il est possible de les utiliser conformément aux RCAA2R.

► **246 – Variante de titre**

Il est indispensable de faire des variantes à tous les différents titres trouvés dans la ressource. Ces entrées secondaires permettent une plus grande possibilité de recherche et

sont, pour la plupart, non visibles dans la notice consultée par l'utilisateur. Si le titre de la barre de titre diffère du titre de la page, une variante est établie. Lorsqu'il s'agit d'une section d'un site, une variante au titre de la page d'accueil est établie, suivie du nom de chaque section menant à la page cataloguée afin de retracer l'arborescence et de faciliter la recherche à partir du titre de la page d'accueil.

► **250 – Mention d'édition**

Cette zone n'est pas utilisée pour les sites Web. Il est préférable d'inscrire les informations dans une note si nécessaire.

► **260 – Adresse bibliographique**

Les ressources Internet sont considérées comme étant publiées. Il faut donc établir une adresse bibliographique complète. Le webmestre ou la firme de conception du site ne sont pas considérés comme les éditeurs au même titre que le concepteur graphique ou l'imprimeur d'un livre. Cependant, si aucun nom d'éditeur ne peut être supposé, le nom du webmestre ou du concepteur peut être mentionné dans cette zone.

Quant à la date, elle est laissée ouverte à cause des changements et mises à jour fréquents. Si aucune date ne peut être trouvée dans la ressource, on peut mettre la date de l'année courante entre crochets avec un tiret.

► **300 – Description matérielle**

Cette zone n'est pas utilisée pour les ressources électroniques accessibles à distance étant donné qu'elles ne possèdent pas de support matériel.

► **5XX – Notes**

La zone des notes est davantage utilisée pour le catalogage des ressources Internet que pour tout autre type de documents, afin, encore une fois, de pallier la versatilité des ressource

ces et les difficultés qu'entraîne leur information changeante :

► **500 – Notes générales :**

- date de la dernière mise à jour de la ressource d'après laquelle la description a été faite.
- mention d'édition si nécessaire (ex. : Description basée sur l'édition de 2001 ; Description basée sur la version de 2002).
- fréquence de la mise à jour s'il y a lieu.
- index : un moteur de recherche par mots clés intégré à un site est l'équivalent d'un index dans un livre. Il serait utile de mentionner en note la présence d'un instrument de recherche et même de spécifier le genre de recherche (par mots clés, auteurs, titres, cartes cliquables, etc.). Pour rester conforme aux RCAA2R, la note « Index (via le moteur de recherche) » peut être utilisée lorsqu'un moteur de recherche par mots clés est présent dans la ressource.

► **504** pour la webographie : Lorsque le site possède un répertoire de liens vers d'autres ressources connexes, le terme *webographie* peut être mis en note.

► **505** pour le dépouillement : Cette zone est à éviter car les mises à jour des sites peuvent rendre l'information rapidement périmée.

► **516** pour indiquer les caractéristiques de certains fichiers ou de certaines données du site : format PDF, fichiers d'animation, audio, vidéo, etc.

► **530** pour indiquer les autres formats physiques disponibles (version imprimée, cédérom, etc.).

► **538** pour la configuration requise pour les fichiers spéciaux mentionnés en **516** : logiciels Flash Player, Quick Time, etc.

► **546** pour indiquer si le site est disponible en d'autres langues.

Catalogage analytique

► **6XX – Vedettes-matière**

Le catalogage par sujets est la méthode la plus efficace pour organiser et rendre accessibles aux bibliothèques

ques les ressources sur le Web. Certains sites sont tellement touffus et abordent tellement de sujets qu'il peut être difficile d'y cerner les concepts importants. Le catalogueur doit se concentrer sur les parties majeures du site pour ne pas s'égarer et il doit disposer d'un très bon sens de la synthèse. La justesse des vedettes-matière est primordiale car la recherche d'un site se fait le plus souvent par sujets, rarement par titres ou par auteurs. Un site mal analysé est une ressource perdue tout comme un livre mal classé sur les rayons d'une bibliothèque est perdu.

Les outils de recherche actuels sur Internet donnent des résultats souvent décevants (trop de « bruit ») et chercher une information peut même s'avérer long et décourageant. Même si ces outils sont de plus en plus perfectionnés et rapides, la recherche par sujets reste délicate à cause d'un manque de normes pour faciliter et homogénéiser l'indexation automatique. L'indexation des ressources d'une webthèque comme *BREF* est faite par des bibliothécaires, offrant ainsi l'avantage d'une meilleure qualité par rapport à la majorité des outils et répertoires de recherche disponibles sur le Net. L'indexation étant un processus d'analyse du contenu intellectuel, la qualité d'indexation humaine est supérieure à celle faite par des robots, qui ne peuvent tenir compte du contexte intellectuel.

Les sites Web à caractère textuel peuvent être indexés comme des monographies avec les vedettes-matière de Laval. Il faut noter que la subdivision de forme Bases de données doit être utilisée avec parcimonie, de préférence seulement lorsqu'une autre subdivision de forme, comme Répertoires, Catalogues, Bibliographie — Catalogues, ne peut être employée et que le catalogueur juge nécessaire de spécifier qu'il s'agit d'une base de données (qui comprend en général un moteur de recherche élaboré). La subdivision Ressources Internet — Répertoires peut être utilisée lorsque le site à cataloguer est un répertoire de liens sur un sujet précis. Dans ce cas-là, la vedette-matière Sites Web — Répertoires sera ajoutée.

► 520 – Résumé

Un résumé descriptif d'un site doit privilégier les mots clés pour un meilleur repérage et doit rester assez général.

L'avenir du catalogage des ressources Internet

Les métadonnées

Les métadonnées sont des données qui renseignent sur la nature de certaines autres données et qui permettent leur utilisation pertinente. Le terme *métadonnées* fait partie du jargon actuel d'Internet pour parler de l'information descriptive intégrée aux fichiers sources des sites Web. Ce type de données sont depuis longtemps familières aux bibliothécaires puisque la petite fiche cartonnée de nos catalogues d'autrefois contenait déjà des métadonnées représentées par un ensemble d'éléments spécifiques pour décrire un livre ou tout autre document disponible dans la bibliothèque : l'auteur, le titre, la date de publication, les sujets, la cote, permettant ainsi de se faire une idée du document et de le trouver sur les

rayons. Les catalogues informatisés des bibliothèques contiennent aussi des notices de métadonnées, plus communément appelées « notices bibliographiques » ou « catalographiques ». Alors que ces notices sont séparées physiquement du document qu'elles décrivent, celles des sites Internet sont intégrées à la ressource, à l'instar des données de catalogage avant publication imprimées dans les livres.

Les balises *meta* en HTML représentent la forme la plus connue des métadonnées dans les documents électroniques. Ces balises peuvent être utilisées pour fournir des renseignements précis sur la ressource, les mêmes que l'on trouve dans une notice catalographique : titre, auteur, éditeur, date, résumé, sujets. *L'Initiative de métadonnées du Dublin Core* se consacre depuis 1997 à l'élaboration de normes de métadonnées en ligne.

L'étiquette de métadonnées *keywords* permet d'énumérer les mots clés qui seront indexés par les moteurs de recherche. Des vocabulaires normalisés sont peu à peu développés pour uniformiser les métadonnées. Un vocabulaire contrôlé permet d'éviter la polysémie, la synonymie et toute ambiguïté de langue et donc

Voici un exemple relevé dans le fichier source de *Sitebib* comprenant les éléments de métadonnées du *Dublin Core*.

```
<HTML>
<HEAD>
<META NAME="TITLE" CONTENT="Sitebib : Coopération entre sites Web en bibliothéconomie et sciences de l'information">
<META NAME="keywords" CONTENT="library, libraries, bibliothèque, bibliothèques, bibliotheque, bibliothèques, bibliothéconomie, bibliothéconomie, sciences de l'information, information science, coopération, cooperation">
<META NAME="DC.title" CONTENT="Sitebib : Coopération entre sites Web en bibliothéconomie et sciences de l'information">
<META NAME="DC.creator" CONTENT="Sitebib">
<META NAME="DC.date" CONTENT="2000-09-08">
<META NAME="DC.description" CONTENT="Liens vers des sites relatifs à la bibliothéconomie et aux sciences de l'information">
<META NAME="DC.subject" CONTENT="library, libraries, bibliothèque, bibliothèques, bibliotheque, bibliothèques, bibliothéconomie, bibliothéconomie, sciences de l'information, information science, coopération, cooperation">
<META NAME="DC.type" CONTENT="txt.webpage">
<META NAME="DC.type" CONTENT="txt.html">
<META NAME="DC.identifier" CONTENT="http://www.abf.fr/sitebib/">
<META NAME="DC.language" CONTENT="fr">
<META NAME="DC.coverage" CONTENT="france">
<META NAME="DC.rights" CONTENT="free">
<TITLE> Sitebib : Coopération entre sites Web en bibliothéconomie et sciences de l'information </TITLE>
</HEAD>
```

d'améliorer le repérage. La qualité de l'indexation automatique découlera du respect de ces normes par les concepteurs de sites Web. L'inconvénient est que les moteurs de recherche ne sont pas encore tous capables d'utiliser les métadonnées de *Dublin Core* pour la création des index.

Étant donné le coût élevé du catalogage des ressources Internet après leur publication, il serait intéressant que les webmasters fassent appel à des professionnels de l'information pour établir les métadonnées catalographiques, particulièrement selon le *Dublin Core*, à l'étape de la conception du site, ce qui permettrait d'intégrer à la fois les données descriptives (auteur, titre, éditeur, date) et les données analytiques (mots clés, résumé). Ainsi, la notice catalographique serait adéquatement intégrée à la ressource. Les éléments de métadonnées établies de la sorte offriraient l'avantage d'une méthode d'organisation et d'indexation de l'information sur le Web et amélioreraient la pertinence des résultats de recherche dans Internet tout en facilitant la création de notices dans un autre format, en format MARC par exemple, pour les catalogues des bibliothèques. Il serait alors plus facile et plus rapide d'utiliser un programme informatique tel *CORC*, capable d'interpréter les notices de métadonnées intégrées aux ressources et de les transformer en un autre format.

CORC

En service depuis juillet 2000, *CORC* (*Cooperative Online Resource Catalog*) de OCLC est un catalogue collectif de ressources Internet sélectionnées et traitées par des bibliothèques. *CORC* fournit des outils automatisés pour aider à la sélection, au catalogage et à l'organisation des ressources électroniques en ligne. Il expérimente une nouvelle technologie qui permet d'extraire les mots clés et les autres informations d'auteur, d'éditeur et de titre qui se trouvent dans les métadonnées HTML des sites. *CORC* offre, à l'occasion de la création des notices, le choix de formatage en MARC21 ou en *Dublin Core*. Il permet aussi de créer des bibliographies, de maintenir à jour les adresses URL et de fournir un contrôle d'autorité.

L'avantage de *CORC* est que, grâce au partage des ressources, le travail de catalogage est accéléré et par conséquent le coût lié au traitement et à l'organisation des ressources réduit. De nombreuses bibliothèques (principalement anglophones) à travers le monde utilisent *CORC* pour le traitement de leurs ressources Internet.

Cependant, *CORC* comporte quelques inconvénients majeurs :

- ▶ Le catalogue dispose principalement de ressources anglophones, restreignant les bibliothèques francophones dans la sélection ;
- ▶ Pour assurer un catalogage de qualité, chaque notice créée par la technologie *CORC* doit être vérifiée puisque les métadonnées sur les sites ne sont pas encore normalisées (les étiquettes HTML d'auteur, d'éditeur, de mots clés sont souvent absentes ou incohérentes) ;
- ▶ Réservé exclusivement aux bibliothèques, *CORC* ne peut servir d'outil de recherche au grand public.

Le concept de la webthèque

Depuis l'arrivée d'Internet, le mandat des bibliothèques a ainsi évolué afin de s'adapter aux nouvelles habitudes documentaires des usagers. La webthèque n'est pas qu'un simple répertoire de recherche parmi tant d'autres sur Internet. Il s'agit d'un concept tout à fait innovateur répondant aux normes de la bibliothéconomie et permettant à l'internaute de trouver l'information désirée d'une manière efficace et structurée dans le grand chaos régnant au sein du Net. Le catalogage des sites en est à ses débuts et, même si les métadonnées représentent l'avenir, elles ne peuvent remplacer actuellement le catalogage traditionnel. Les webthécaires ou les professionnels de l'information sont les mieux qualifiés pour s'occuper de la description et de l'indexation d'un site, aussi pour s'intéresser à la manière de le faire, c'est-à-dire s'occuper autant de la structure des données que de leur contenu. Leur rôle est donc primordial pour améliorer l'accès à l'information électronique sur Internet. ■

RÉFÉRENCES

- Caraco, Alain. 2000. *Intégrer les ressources d'Internet dans la collection*, Villeurbanne : Enssib, p. 223
- Charuest, Isabelle. BREF : Bibliothèque de référence électronique pour les ressources sur Internet à l'intention des bibliothèques publiques. *Bibliothèques publiques du Québec*, [En ligne]. <http://www.bpq.org/bref.htm> (Page consultée le 28 novembre 2003)
- Conférence des recteurs et des principaux des universités du Québec, Sous-comité des bibliothèques, Groupe de travail sur le traitement de la documentation. *Politique commune pour le traitement des ressources Internet dans les bibliothèques universitaires du Québec*, [En ligne]. http://crepuq.qc.ca/documents/bibl/pub/Res-ources_Internet.htm (Page consultée le 12 novembre 2003)
- DeCandido, GraceAnne A. *Metadata: Always More Than You Think*. Public Library Association. *Tech notes*, [En ligne]. http://www.ala.org/Content/NavigationMenu/PLA/Publications_and_Reports/Tech_Notes/Metadata.htm (Page consultée le 28 novembre 2003)
- Elrod, J. McRee (Mac). 2000. *Classification of Internet Resources: An AUTOCAT Discussion. Cataloging & Classification Quarterly*, vol. 29 (4) : p. 19-38.
- Fédération internationale des associations de bibliothécaires et des bibliothèques. 2000. *ISBD(ER) : description bibliographique internationale normalisée des ressources électroniques*, Paris : Bibliothèque nationale de France : 101 p.
- Guindon, Alex. Printemps 2003. BREF, une expérience de collaboration. *Argus*, vol. 32, n° 1.
- Hillmann, Diane. *Guide d'utilisation du Dublin Core. Initiative de métadonnées du Dublin Core*, [En ligne]. <http://www.bibl.ulaval.ca/DublinCore/usageguide-sectc-20000716fr.htm#publisher> (Page consultée le 2 février 2002)
- Lam, Vinh. 2000. *Cataloging Internet Resources: Why, What, How. Cataloging & Classification Quarterly*, vol. 29 (3), p. 49-59.
- Lardy, Jean-Pierre. *RISI, recherche d'information sur l'Internet*, [En ligne]. <http://www.adbs.fr/adbs/sitespro/lardy/risi.htm> (Page consultée le 3 septembre 2001)
- OCLC. *Cataloging Electronic Resources: OCLC-MARC Coding Guidelines, revised 2001 December 6*, by Jay Weitz. *OCLC Collections & Technical Services*, [En ligne]. <http://www.oclc.org/ca/en/support/documentation/worldcat/cataloging/electronicresources/> (Page consultée le 23 janvier 2002)
- Olson, Nancy B. (sous la direction de). 2^e édition. *Cataloging Internet Resources: A Manual and Practical Guide, OCLC*, [En ligne]. <http://www.oclc.org/oclc/man/9256cat/toc.htm> (Page consultée le 3 septembre 2001)
- Règles de catalogage anglo-américaines*, 2000. 2^e édition, révision de 1998, Montréal : ASTED, 887 p.
- Règles de catalogage anglo-américaines*, 2001. 2^e édition, révision de 1998, modifications de 2001, Montréal : ASTED, 129 p.
- Ward, Diane. 2001. *Internet Resource Cataloging: the SUNY Buffalo Libraries Response. OCLC Systems & Services*, vol. 17, n° 1, p. 19-25.