

# Catalogage du Web : c'est possible

Les professionnels de l'information n'ont de cesse, avec des objectifs variés, de collecter, collationner, cataloguer les informations et les documents qui intéressent leurs publics afin de les proposer dans les catalogues ou bases de données en ligne, dans un portail ou sous forme de signets. Tous les supports de l'information sont concernés, des règles, normes et standards ont été établis afin de les décrire le plus précisément possible : la description bibliographique se décline ainsi différemment et de manière plus ou moins complexe qu'il s'agisse d'un ouvrage, d'un cédérom ou d'un enregistrement sonore. Qu'en est-il du Web ? Faut-il le cataloguer à l'instar des autres supports ? La réponse est oui, mais l'entreprise apparaît immense, coûteuse et incertaine tant les informations sont versatiles sur le Net. La Bibliothèque nationale de France donne <sup>1</sup> quelques chiffres révélateurs concernant les sites français :

- 250 000 sites publics enregistrés en .fr ;
- 250 000 sites publics enregistrés en .com, .org, .edu, .net... ;
- et 4 millions de sites personnels.

Les bibliothèques nationales ont plusieurs projets de catalogage systématique du Web <sup>2</sup>, mais au vu du développement des ressources électroniques, de plus en plus de bibliothèques, services de documentation et d'archives les intègrent dans leurs catalogues.

Quelques questions préliminaires à se poser

Le catalogage du Web impose de se demander dès l'abord quelle ressource électronique est intéressante à sélectionner : est-ce un site Internet dans son ensemble ? Une page seulement ? Ou un document contenu dans le site (un rapport par exemple) ? Plusieurs critères sont déterminants et dépendent avant tout de la politique de catalogage mise en place : celle-ci est définie par rapport à la politique d'acquisition des documents dans son ensemble, du ou des publics à servir, des fonds à proposer et des moyens alloués. Même si un site Web n'a pas besoin d'être « acheté », son traitement (sélection, description, saisie de la notice) est quasiment le

---

<sup>1</sup> <http://www.bnf.fr>, rubrique Professionnels, page Dépôt légal (chiffres au 31 déc. 2004).

<sup>2</sup> La question du catalogage pose également celle du dépôt légal du Web : en 1996 déjà, la Conférence des directeurs de bibliothèques nationales (CDNL) estimait qu'il était essentiel de répertorier et de conserver à perpétuité les publications électroniques, « faute de quoi le passé ne laisserait plus de traces, et les recherches antérieures ne pourraient plus être retrouvées, comprises ou reproduites comme il se doit, pour entretenir le cycle continu de l'expérimentation qui fait progresser la connaissance ». Cette préoccupation majeure voit l'éclosion de projets nationaux divers : la Bibliothèque nationale de France avec le Consortium international pour la préservation de la Toile (IIPC) lancé en 2003 ; la Bibliothèque nationale d'Australie avec le projet Pandora archive depuis 1996 les publications nationales en ligne ; la Bibliothèque royale de Suède avec le projet Kulturarw, lancé lui aussi en 1996, enregistre régulièrement le web pour sauvegarder automatiquement une copie de tous les sites suédois accessibles sans mot de passe ; la Bibliothèque nationale suisse, avec le projet e-Helveticum lancé en 2001, souhaite conserver à long terme les publications électroniques helvétiques (cédéroms, DVD, sites Internet). Ces initiatives qui montrent une réelle prise de conscience sont des initiatives publiques. Une fondation américaine, The Internet Archive, procède à la captation régulière – mais cependant aléatoire – de sites Internet du monde entier, soit à l'heure actuelle plus de 40 milliards de pages.

même qu'un document traditionnel. La bibliothèque municipale de Fresnes <sup>3</sup> catalogue les sites WEB en format InterMarc comme n'importe quelle monographie : les catalogueurs utilisent essentiellement les zones de titre et de mention de responsabilité, ainsi que la zone de note où un copier/coller de la page de présentation est fait. Les liens avec les autorités Rameau et collectivité auteur répondent aux mêmes règles que pour les autres documents.

Une autre question est de savoir où référencer le site Web : certaines bibliothèques l'intègrent directement dans leur catalogue en ligne au même titre qu'un autre document, d'autres choisissent de le répertorier dans leurs signets ou dans leur portail d'information. Différents formats sont alors possibles.

Différentes manières de cataloguer le Web : le choix d'un format de catalogage

L'adoption d'un format standard apparaît comme une solution raisonnable : quel format de description choisir ? Quels champs utiliser dans le format choisi ? Quelles données convient-il de rendre obligatoires ? Quelles règles de catalogage retenir ? Le catalogage au niveau complet des ressources électroniques implique la création d'une notice relativement longue. Il y a plusieurs champs fixes à remplir, souvent plusieurs titres, et plusieurs notes de format. Les formats Marc ou InterMarc proposent d'ajouter des champs :

Aide pour le catalogage des sites web avec le format Marc

008 Fixed field data	256 File characteristics	600 SU - Personal name
041 Language codes	260 Publication info	610 SU - Corporate name
100 Personal author	500 General public note	651 SU - Geographical
110 Corporate author	505 Contents note	690 SU CERES
245 Title	520 Abstract	700 Added author
246 Variant title	538 Technical details	710 Added corporate author

Certains logiciels proposent d'extraire les métadonnées d'un site Web pour créer une notice MARC en format court : c'est le cas de Connexion d'OCLC <sup>4</sup> ou de MARCit <sup>5</sup> de SagebrushLast. MARCEdit de l'Université de l'Orgeon <sup>6</sup> reprend même les informations fournies par les agences d'abonnement pour les revues électroniques ou par les fournisseurs d'ebooks. Le travail du catalogueur est ainsi grandement facilité.

Mais il existe d'autres formats utiles tels AACR2 (Anglo-American Cataloguing Rules) <sup>7</sup>, LCRI 1.0 (Library of Congress Cataloguing Policy) <sup>8</sup> ou Dublin Core <sup>9</sup> ...

---

<sup>3</sup> <http://www.mairie-fresnes.fr/bibliotheque>

<sup>4</sup> <http://www.oclc.org>

<sup>5</sup> <http://www.sagebrushcorp.com/tech/marcit.cfm>

<sup>6</sup> <http://oregonstate.edu/~reese/marcedit/html/>

<sup>7</sup> <http://www.aacr2.org/>

<sup>8</sup> [http://www.loc.gov/catdir/cpsolc/lcrl1\\_0.html](http://www.loc.gov/catdir/cpsolc/lcrl1_0.html)

<sup>9</sup> <http://dublincore.org/>

Selon les producteurs du site CISMEF (Catalogue et index des sites médicaux francophones) <sup>10</sup>, le format Dublin Core semble être actuellement le plus utilisé, notamment par les sites institutionnels. Dans ce format, quinze « éléments » sont définis : Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifiant, Source, Language, Relation, Coverage, Rights. Tous ne sont pas obligatoires et il est possible d'ajouter des éléments supplémentaires, pour des besoins spécifiques. CISMEF ajoute ainsi les éléments suivant : institution, ville, département, pays, coût, parrainage, public concerné. Ces éléments qui ne sont pas standards doivent être documentés. Des qualificatifs permettent aussi de préciser un élément de base. Dublin Core est interchangeable avec le format USMARC comme dans l'application CORC d'OCLC <sup>11</sup>. Le format Dublin Core présente ainsi l'avantage considérable d'une interopérabilité entre systèmes d'information.

Les éléments-clés de la description d'une ressource électronique : les normes actuelles

L'IFLA <sup>12</sup> a édité les normes concernant les documents électroniques, notamment en 1989 l'ISBD (CF) <sup>13</sup>, puis en 1997 l'ISBD (ER) <sup>14</sup> et en 1997, l'AFNOR <sup>15</sup> a publié un fascicule : AFNOR FD Z 44-082 <sup>16</sup>.

Bien qu'un document électronique puisse ressembler à certains types de publications imprimées (une monographie, une publication en série, un article ou un chapitre), les caractéristiques de celles-ci peuvent ne pas apparaître sous la forme électronique (par exemple, le volume, la pagination). Les références bibliographiques des documents électroniques doivent refléter leur identité, et non celle des succédanés du papier. Un document électronique induit un catalogue particulier :

- indication de la nature du document : [en ligne] ou [online] ;
- date de mise à jour : « mis à jour en... » ou « revu le... » ;
- date de référence : [consulté le ] ou [référence du...] ;
- signalisation d'une partie d'un document électronique, par exemple, le nombre total de lignes, d'écrans, ce qui donne [35 lignes] [environ 12 écrans]. À indiquer : « mise à jour hebdomadaire » ou « chargement annuel » ;
- disponibilité et accès : « Disponible sur Internet » ou « Disponible sur World Wide Web » ou « Available from Internet ». L'ensemble sera suivi de l'adresse de l'ordinateur hôte, du nom du répertoire, du nom du fichier. Exemple : Disponible sur Internet : <<http://www.adbs.fr/>>.

## Catalogage des documents électroniques

### Les monographies en ligne

<sup>10</sup> <http://www.chu-rouen.fr/cismef/>

<sup>11</sup> <http://www.asso.univ-paris5.fr/auroc/corc.htm>

<sup>12</sup> <http://www.ifla.org>

<sup>13</sup> International Standard Bibliographic Description Computer Files (traduite par la BnF)

<sup>14</sup> International Standard Bibliographic Description for Electronic Resources

<sup>15</sup> <http://www.afnor.fr/portail.asp>

<sup>16</sup> « Documentation – Catalogage des ressources électroniques – Rédaction de la description bibliographique »

**RESPONSABILITÉ PRINCIPALE.** Titre de l'ouvrage [en ligne]. Mention d'édition si nécessaire. Lieu de publication : éditeur, date de publication, date de mise à jour ou de révision [date de la référence]. Disponibilité et accès.

Les extraits ou les contributions à des monographies

Il n'y a aucune différence avec les monographies imprimées. On introduira le « In » pour les contributions.

Les articles de périodiques électroniques

Là encore, il n'y a pas de différence notable, si ce n'est qu'il faut suivre les références propres aux documents en ligne :

AUTEUR, Prénom. Titre de l'article. Titre du périodique [en ligne]. Année, vol., numéro, date de mise à jour ou de révision [date de consultation], localisation dans le document hôte. Disponible sur Internet : <...>.

Les messages électroniques

**RESPONSABILITÉ PRINCIPALE.** Titre du message. Titre du système de messagerie [en ligne]. Lieu de publication : éditeur, date de publication [date de la référence]. Localisation à l'intérieur du système de messagerie hôte. Disponibilité et accès.

Le catalogage d'un site Web présente donc un certain nombre d'avantages selon la norme et le format choisis. A l'heure où l'accès à l'information s'accroît, le fait de proposer des ressources supplémentaires organisées et pertinentes aux utilisateurs relève du rôle traditionnel des professionnels de l'information. La description bibliographique du Web permet de connaître à l'avance le type d'information présente et d'évaluer rapidement la qualité de l'information fournie.

Jean-Philippe Accart

[Jean-Philippe.Accart@slb.admin.ch](mailto:Jean-Philippe.Accart@slb.admin.ch)