

Les catalogues des bibliothèques : du web invisible au web social

Isabelle de Kaenel

Bibliothèque Universitaire de Médecine (BiUM)
Centre de Documentation en Santé Publique (CDSP)

Pablo Iriarte

Bibliothèque Universitaire de Médecine (BiUM)
Centre de Documentation en Santé Publique (CDSP)
Centre Hospitalier Universitaire Vaudois (CHUV) – Lausanne

Les catalogues des bibliothèques sont rentrés dans une phase critique. Les dernières évolutions du web, avec l'entrée en jeu enfin de XML, des nouveaux usages et nouveaux outils, ainsi que le déplacement du centre de gravité qui s'est fortement rapproché des utilisateurs, ouvrent de nouvelles voies et de nouveaux champs d'application pour les catalogues en ligne. Le catalogue n'est plus un outil isolé du monde : dans un mouvement à double sens, il doit s'ouvrir à Internet autant pour tirer parti des services web externes, de plus en plus importants, que pour l'alimenter en contenu et fournir des informations structurées et validées tout en permettant aux utilisateurs d'apporter du contenu et du sens, ainsi que de s'approprier les données du catalogue en lui offrant des nouvelles possibilités de réutilisation à travers les réseaux. Cet article fait un inventaire de ces nouveaux champs d'application possibles et analyse les conditions de base qu'un catalogue devrait remplir pour pouvoir quitter le web invisible et investir pleinement les possibilités actuelles du web social pour devenir enfin un « OpenCatalog ».

Mots-Clés : Catalogue de bibliothèque, OPAC, XML, Web 2.0, services web, Open Access

Introduction

Les catalogues des bibliothèques sont tombés en disgrâce (Markey, 2007). Délaissés, ignorés, critiqués¹ et, dans le pire des cas, ridiculisés², ils nous lancent un dernier cri d'alarme avant de tomber dans le tiroir des outils oubliés. Or, les dernières évolutions du web³, avec l'entrée en jeu enfin de XML⁴, des nouveaux usages et nouveaux outils, ainsi que le déplacement du centre de gravité qui s'est fortement rapproché des utilisateurs, ouvrent de nouvelles voies et de nouveaux champs d'application pour les catalogues en ligne. Le temps est donc venu d'aider ces outils de recherche à sortir de cette image négative liée à une complexité, certes inévitable mais jamais

compensée par un aspect créatif. Peu armés pour affronter la rapidité des changements de l'ère Internet au cours de ces dix dernières années, les bibliothécaires se sont plus ou moins contentés d'un fonctionnement qui semblait avoir fait ses preuves, avec de promesses d'améliorations, sans vouloir se rendre compte que, tout autour, le monde de l'information numérique expérimentait plusieurs révolutions. Il est bien connu que celui qui n'avance pas...

La tendance est pourtant depuis quelque temps vers la flexibilité et l'ouverture : « *Open Source*⁵ » (ouverture du code source des logiciels), « *Open Access*⁶ » (accès ouvert aux publications et données issues de la recherche scientifique et

technique) et « *Open Archives Initiative*⁷ » (ouverture et interopérabilité entre serveurs institutionnels dépositaires de cette production), « *OpenURL*⁸ » (liens ouverts grâce aux métadonnées encodées dans l'URL⁹), « *OpenSearch*¹⁰ » (syntaxe des requêtes et format des résultats ouverts), sont des exemples concrets des réalisations qui ont modifié le monde de la documentation numérique. Le but de cet article est donc de faire un petit inventaire de ces nouveaux champs d'application possibles et d'analyser les conditions de base qu'un catalogue devrait remplir (sans devoir tout remettre en question et sans renoncer aux acquis qui font sa spécificité et sa richesse), afin d'investir pleinement les possibilités actuelles et devenir enfin un « *OpenCatalog* »¹¹.

Plusieurs raisons peuvent expliquer ce retard en dépit d'une communauté de bibliothèques bien structurée et dynamique. Les contraintes physiques et matérielles importantes imposées aux catalogues (catalogues papier reliés en volumes, format des cartes pour les cardex ou limitations de mémoire pour les premiers systèmes informatisés) ont probablement joué un rôle majeur (Calhoun, 2006, p. 36). Ainsi, le catalogue a pris l'option d'être un système autoréférentiel, autarcique, dans le sens où il s'autosuffisait, ne citant pratiquement pas de ressources externes en dehors de son propre univers informationnel. Par conséquent, la bibliothèque de type universitaire ou encyclopédique, utilisant ce modèle de catalogue, qui a parfaitement fonctionné et perduré pendant des siècles, complété par d'imposantes bibliographies, des répertoires des périodiques, des services de commandes et de prêt entre bibliothèques, pouvait alors prétendre à une très large exhaustivité.

La combinaison de ces deux aspects a engendré des méthodes de travail : le catalogage, et un produit : le catalogue, qui ont peu évolué en comparaison avec le reste des outils informatiques dans le domaine de l'édition commerciale qui, eux, ont dû faire face et s'adapter plus

rapidement aux changements radicaux survenus depuis l'arrivée du web et des NTICs¹². Les bibliothèques ne se sont pas mobilisées pour faire évoluer leurs catalogues face à ces bouleversements. Les évolutions restent très lentes, avec beaucoup d'expérimentations et des réalisations partielles¹³.

Si nous remontons dans le temps, nous pouvons constater que les bibliothèques ont été très actives dans la période d'informatisation des catalogues au cours des années 70-80, dont l'un des meilleurs exemples est le système de catalogage et de gestion SIBIL¹⁴ développé à la BCU¹⁵ de Lausanne (Gavin, 1997). Cette étape a apporté le format MARC¹⁶ utilisé actuellement dans la plupart des bibliothèques, malgré une remise en question récurrente. Les bibliothèques ont aussi participé activement au projet du WWW. Par exemple, il est significatif que le premier site web en dehors du CERN¹⁷ fût créé par Louise Addis, bibliothécaire du *Stanford Linear Accelerator Center (SLAC)*¹⁸, qui devenait ainsi la première bibliothécaire-webmaster de l'histoire¹⁹. Aussi, dès les premiers temps du web, les catalogues des bibliothèques sont devenus accessibles sur la Toile de manière libre et gratuite pour la plus grande satisfaction des utilisateurs du monde entier. Ces OPACs (« *Online Public Access Catalog* ») ont été aussi l'une des premières réalisations à grande échelle du principe de la séparation entre contenu et mise en page, principe popularisé plus tard pour les systèmes de gestion de contenu (CMS²⁰). L'aspect collaboratif du travail de catalogage partagé (chaque notice du catalogue peut en principe être corrigée ou améliorée par n'importe quel autre catalogueur du réseau) était aussi en avant par rapport à son temps. Cependant, ce sont les wikis²¹ qui, en donnant cette possibilité d'édition des données à tout un chacun, ont poussé le concept de travail collaboratif à l'extrême et l'ont popularisé dans l'univers du web. De la même façon, le protocole de communication Z39.50²², développé et

maintenu depuis plus de 20 ans par la *Library of Congress*²³, fut l'un des précurseurs d'Internet et, malgré un déclin important dû à son « incompatibilité » avec les technologies web²⁴, il est toujours utilisé par un bon nombre de logiciels (bibliographiques ou de pompage des notices) et de plateformes de méta recherche comme le KVK²⁵.

Lorsque les catalogues informatisés (souvent gérés, comme dans le cas de SIBIL, avec des outils développés localement) ont atteint une taille trop importante, ils ont dû migrer sur des outils devenus propriétaires et développés par des entreprises commerciales internationales. Cette évolution a peut-être tué une bonne partie de l'initiative des bibliothèques qui se sont peu à peu tournées vers d'autres fronts (Open Access et serveurs institutionnels pour lutter contre la crise des prix des périodiques et gérer les publications institutionnelles par exemple) et vers d'autres outils destinés au web et venant compléter le catalogue (portails, CMS, blogs²⁶ et wikis, podcasts²⁷, outils de gestion de liens²⁸ et de recherche fédérée...), laissant un peu pour compte son outil principal de travail dont on annonce régulièrement plus ou moins la fin ou la désintégration.

Ainsi, la publication web des catalogues reflète encore passablement cet ancien esprit autarcique, et les OPACs restent souvent « déconnectés » du reste des ressources en ligne²⁹ et sont, encore aujourd'hui, conditionnés par d'anciennes limitations qui n'ont plus de sens dans l'environnement culturel et technologique actuel.

Le catalogue n'est plus un outil isolé du monde. Dans un mouvement à double sens, il doit s'ouvrir au web autant pour tirer parti des services web externes, de plus en plus importants, que pour alimenter le web en contenu et fournir des informations structurées et validées. L'enjeu est de taille : comment rester fidèle à son sens premier (répertoire de manière cohérente les ressources mises à disposition du public et aider à la recherche, découverte,

localisation, et gestion des collections) tout en permettant aux utilisateurs d'apporter du contenu et du sens, ainsi que de s'approprier les données du catalogue en lui offrant des nouvelles possibilités de réutilisation à travers le web. Sans se pervertir, le catalogue doit évoluer rapidement pour pouvoir rester dans la course où il a déjà pris un retard considérable.

La conclusion d'un rapport commandé par la *Library of Congress* en 2006 ne laisse pas de doutes sur le chemin qui reste à parcourir :

“The future will require the kind of catalog that is one link in a chain of services enabling users to find, select, and obtain the information objects they want. One requirement of this future catalog is thus to ingest and disperse data from and to many systems inside and outside the library. It would be helpful to reconsider what needs to be part of catalog data —and where catalog data needs to be present— to facilitate the user’s process of discovering, requesting, and getting the information they need.” (Calhoun, 2006, p. 38).

Dans les pages suivantes, nous essayerons d'explorer les possibilités de mise en place de ces deux ouvertures souhaitables du catalogue : l'ouverture à l'intégration des nouveaux contenus (internes ou externes) et, d'autre part, l'ouverture à de nouvelles formes d'utilisation de ses propres données par des tiers.

1^{ère} partie : ouverture du catalogue à l'intégration des nouveaux contenus

1. Intégration de l'hypertexte : deux modèles à suivre

Grâce aux logiciels libres et surtout au couple PHP/MySQL³⁰, l'architecture de la Toile a changé et repose désormais sur un vaste ensemble de bases de données. Le Web est devenu une véritable plateforme de travail, autonome, indépendante des contraintes spatiales ou matérielles liées

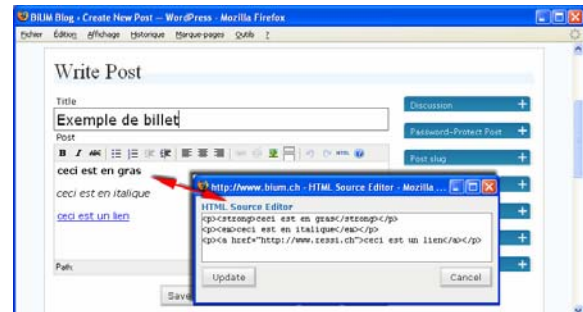
à des systèmes d'exploitation, des versions des logiciels, etc. De la même façon que la messagerie peut être utilisée depuis n'importe quel ordinateur relié à Internet à l'aide d'un simple navigateur, il sera bientôt possible de travailler avec un minimum de logiciels et clients lourds installés sur les postes, car la plupart des outils seront disponibles en version 100% web³¹.

A l'image des développements réalisés pour les autres outils de gestion et de diffusion de l'information, la logique et les mécanismes de fonctionnement du catalogue sont de plus en plus orientés vers le web, au détriment des autres formes de consultation ou de publication (clients professionnels en mode OPAC installés sur les postes de consultation, bulletins de nouvelles acquisitions, bibliographies nationales...). Si cette tendance suit un certain effet de mode, elle correspond aussi à un changement dans le mode de fonctionnement de la société occidentale, de plus en plus relié à Internet à haut débit, et qui dédie de plus en plus de temps à la « consommation » de médias numériques, qui dépassent déjà chacun des autres médias traditionnels (TV, radio, journaux et revues papier, cinéma, etc.)³². Cette évolution qui semble pour l'instant irréversible, nous pousse à repenser le catalogue comme un outil fait par et pour le web, intégrant ainsi de manière véritable ce média dont la caractéristique et l'avantage principal réside dans l'immédiateté et dans la navigation à travers les liens hypertexte. Les deux outils principaux de publication web actuels, les blogs et les wikis, sont des bons exemples de la façon dont cette dimension hypertextuelle peut être ajoutée à l'information de manière simple et rapide.

Les blogs

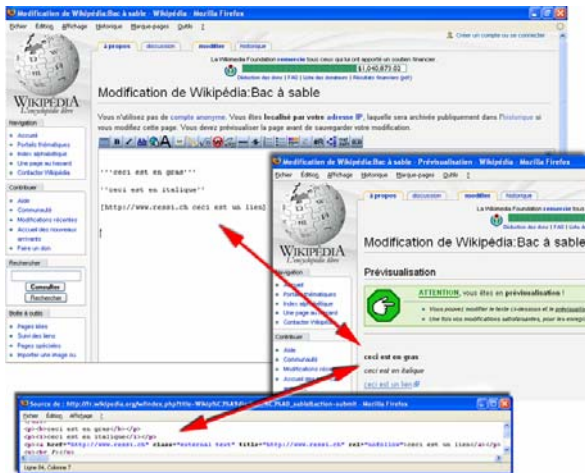
Tout comme la messagerie web, le blog, né directement dans l'univers de l'Internet, a adopté le HTML comme langage principal. Il utilise toutefois un intermédiaire pour aider les non initiés à la saisie : le code HTML est généré et caché automatiquement par un outil d'édition de

type WYSIWYG³³. Pour les courriels, le code HTML est encapsulé dans le corps du message. Dans les blogs, il est enregistré dans la base de données et re-proposé à nouveau tel quel sur le web³⁴ :



Les wikis

Dans les wikis, le code HTML est simplement produit à la volée³⁵ au moment de l'affichage de la page. Les balises HTML sont créées dynamiquement en fonction de la syntaxe propre au wiki. On assiste alors à la transformation d'une syntaxe arbitraire et plus ou moins proche du HTML vers la syntaxe HTML. Par exemple :



Source :

http://opac.rero.ch/get_bib_record.cgi?rero_id=R003636602

La syntaxe wiki étant plus facile à retenir que les balises HTML, les textes peuvent être alors écrits rapidement et on peut utiliser l'aide des boutons de l'éditeur qui effectuent les mêmes fonctions que les éditeurs WYSIWYG, mais sans cacher le code et sans faire la transformation en HTML. Seul le texte avec la syntaxe propre au wiki est alors enregistré dans la base de données et non pas du code HTML.



Source :

<http://www.saphirdoc.ch/permalien.htm?saphirid=41062>

Cette intégration du code au message, en HTML ou selon une syntaxe propre, permet d'apporter les améliorations suivantes avec peu d'efforts :

- Disposer du véritable hypertexte, avec des liens internes ou externes intégrés au message sans devoir à chaque fois afficher l'URL en clair.
- Ajouter une mise en page simple à l'intérieur du texte (courriel, corps d'un billet sur un blog, page entière d'un wiki) : gras, italique, souligné, listes numérotées ou à puces, tableaux, ajout d'images.
- Extraire certaines parties (titres) dans le cas du wiki, pour créer des tables des matières à la volée

Cependant, avant d'appliquer l'un ou l'autre modèle au catalogue, il faut évaluer en profondeur les utilisations pour lesquelles cette introduction des balises ou des syntaxes pourrait être nuisible (impression des catalogues sur papier, maintenance à long terme des tables de conversion entre la syntaxe ou les balises et la mise en page...).

Ces applications peuvent être utiles aussi aux catalogues dans les zones destinées aux notes, commentaires, résumés ou tables des matières, qui souffrent aujourd'hui du manque de mise en page et d'hypertexte :

Malgré l'absence encore du véritable hypertexte dans les catalogues, il faut noter que le format MARC prévoit la zone 856³⁶ pour les URLs en admettant un sous-champ pour ajouter un commentaire qui permet de le qualifier ou de lui donner un contexte sommaire. Ces URLs sont alors transformés à la volée en hyperliens par le système au moment de l'affichage qui, trop souvent encore, se fait uniquement au niveau de la notice complète³⁷.

La présence des URLs dans la notice (champ 856 mais aussi dans le champ du titre, des notes, du résumé...) est un pas important, mais il n'est pas tout. Pour être vraiment utiles à l'internaute, ces URLs doivent impérativement être transformés en lien hypertexte dans la page HTML de l'OPAC, faute de quoi ils seront affichés comme du simple texte qui ne peut pas « être cliqué », et l'utilisateur doit faire recours au copier/coller pour placer cet URL dans l'adresse du navigateur. Pour éviter cela, les logiciels de gestion tentent systématiquement d'effectuer cette transformation URL -> hyperlien en scrutant chaque champ au moment de l'affichage, à la recherche d'un URL. Certains le font mieux que d'autres, car si en principe il est simple pour une machine de trouver le début et la fin d'un URL bien formé au milieu d'une chaîne de caractères (commence par « http:// » et fini par un espace), il est plus complexe de tester toutes les autres variantes possibles (l'URL ne commence pas par « http:// » mais directement par « www. » ; ou ne finit pas par un espace mais par un point, une virgule ou la fin d'une parenthèse...). Ce problème, récurrent aussi dans le cas des courriels qui portent des URLs dans le corps du message, est de difficile solution sans l'utilisation d'un codage à priori comme ceux utilisés par les blogs et les wikis.

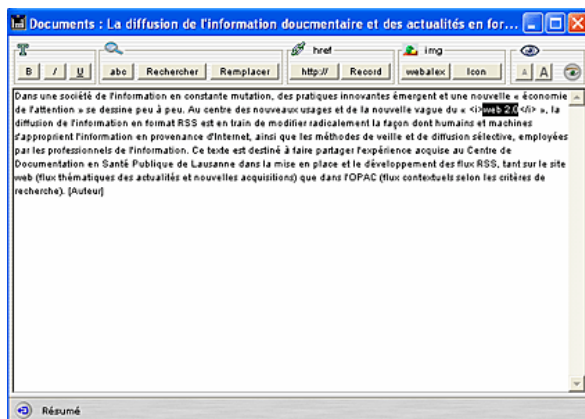
Dans le cas du catalogue de RERO³⁸, le système convertit automatiquement à la volée chaque champ contenant « http:// » (856 mais aussi les champs de titre et des notes) en lien hypertexte, au moment de l'affichage sur l'OPAC. Cependant cette conversion se limite au premier URL trouvé dans le champ³⁹, et elle ne se fait pas pour les URLs qui ne commencent pas par « http:// » mais par exemple « www. »⁴⁰, ce qui limite l'utilisation des liens dans un même champ. Dans le réseau SAPHIR⁴¹, étant donné que le nombre des URLs est plus important à l'intérieur d'un seul champ comme le résumé, la transformation a été étendue à tout URL contenu dans ce champ, et aussi pour ceux qui commencent par « www. ».

Si cette génération à la volée des hyperliens est possible grâce aux caractéristiques reconnaissables des URLs, pour les autres points utiles du codage à priori (mise en page, gras, italique, listes...) il n'y a pas de solution automatique à posteriori. Par conséquent, pour améliorer la mise en page et faciliter la lecture pour les longs résumés correspondant à certains dossiers thématiques, le catalogue SAPHIR a adopté le modèle du wiki de façon encore informelle comme nous le montre la copie d'écran ci-dessous.

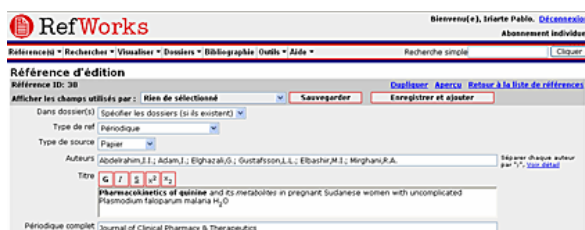


Source : <http://www.saphirdoc.ch/permalien.htm?saphirid=45776>

De son côté, le logiciel Alexandria⁴², SIGB utilisé par ce réseau, a aussi introduit une possibilité de mise en page dans les résumés des documents, en suivant le modèle du blog : trois boutons permettent d'ajouter les balises HTML pour appliquer au texte les styles gras, italique et souligné. Deux autres boutons permettent d'ajouter des liens hypertexte sur des pages externes ou des documents internes à la base, et encore deux pour ajouter des images du serveur :



Dans le champ de logiciels de gestion bibliographiques, le système RefWorks⁴³, qui fonctionne entièrement sur le web, a aussi introduit un éditeur WYSIWYG pour ajouter certains styles (gras, italique, souligné, exposant ou indice) aux champs titre, notes et commentaire :



Le code HTML introduit dans la base est correctement supprimé dans tout format d'export sauf pour les bibliographies de type HTML ou RTF, dans lesquelles l'enrichissement graphique est utile et peut être conservé.

2. L'enrichissement des catalogues

Les systèmes automatisés mis au point pour la recherche d'informations textuelles sur le web, moteurs de recherche, annuaires, méta moteurs, rivalisent en performance et en innovations et montrent le chemin en matière de fonctionnalités et d'enrichissements possibles pour les catalogues. Plusieurs niveaux sont concernés : le graphisme, l'ergonomie de l'affichage, l'intégration d'informations complémentaires (résumés, commentaires, tables des matières), l'aide à la recherche, les possibilités de tri des résultats, etc.

Images de couverture, résumés et tables des matières des livres

Les catalogues de bibliothèques sont des mines d'information librement disponibles. Mais les données sont très dépouillées et certaines informations importantes manquent encore terriblement à l'appel, comme les tables des matières et les résumés.

Depuis plusieurs années, les moteurs de recherche et les sites commerciaux comme Amazon⁴⁴ ont montré une très grande inventivité en enrichissant de plus en plus leur contenu qui, en devenant plus étendu, attire de plus en plus d'utilisateurs qui à leur tour le complètent, dans un phénomène de cercle vertueux qui s'auto-génère une fois dépassée une certaine masse critique. Des fournisseurs de services sont alors apparus permettant aux bibliothèques de s'offrir ce que d'autres avaient réussi à intégrer : tables des matières, résumés et images de couverture sont maintenant vendus ou loués par des sociétés comme *Electre*⁴⁵ ou *Syndetics*⁴⁶. Il est ainsi possible de les ajouter au catalogue à la demande (au moment de la visualisation de la notice complète par exemple) à partir de l'ISBN⁴⁷ du document.

Avec une autre logique commerciale visant plutôt à étendre son rayon d'influence et à promouvoir son site grâce aux liens hypertexte, les services web mis en place par Amazon⁴⁸, permettent depuis l'année 2002 d'utiliser gratuitement le contenu en provenance de sa propre base (images des couvertures, revues des lecteurs, prix, etc.) sur n'importe quel autre site Internet qui respecte les conditions légales d'utilisation. Ainsi plusieurs catalogues utilisent déjà les services web d'Amazon pour enrichir les pages du catalogue : *Dreiländerkatalog*⁴⁹, *Lamson Library*⁵⁰, etc.

Le catalogue SAPHIR, qui a toujours repris dans ses notices des résumés et éléments de tables des matières, utilise

maintenant aussi ces services web pour afficher l'image de couverture sur la page de la notice complète des livres. En utilisant de façon contextuelle l'ISBN et le code de la langue du document introduits au catalogage, la technique AJAX⁵¹ permet d'appeler les services web d'Amazon.fr, Amazon.de ou Amazon.com du côté client (c'est le navigateur qui fait le travail) sans aucune charge supplémentaire du côté serveur⁵² :

SAPHIR (Swiss Automated Public Health Information Resources)

Recherche rapide : OK

Identifiant : Mot de passe : OK

Recherches : Révisions : Coordonnées : Nouvelles 2497 : Publications : Bibliographies : Non espace

Importer | Exporter (Reference Manager, EndNote...) | Liens externes | Permalien

Monographies | Bibliothèque(s) : EUNSP | Classif. IJMSIP : QV-137 Tob-2004 | N° 44787

Tobacco and public health : science and policy / ed. by Peter Boyle... [et al.]

Auteur : Boyle, Peter Ed.
 Lieu/dates éditeur : Oxford University Press, 2004
 Format : KOOL, 798 p., ill., 25 cm
 ISBN/ISSN : 028252879

Résumé : This book comprehensively covers the science and policy issues relevant to one of the major public health disasters of modern times. It pulls together the aetiology and burden of the myriad of tobacco-related diseases with the successes and failures of tobacco control policies. The document looks at lessons learnt to help set health policy for reducing the burden of tobacco-related diseases. It also deals with the international public health policy issues which bear on control of the problem of tobacco use and which vary between continents. The editors are an international group distinguished in the field of tobacco-related diseases, epidemiology, and tobacco control. The contributors are world experts drawn from the cancer clinical fields. This major reference text gives a unique overview of one of the major public health problems in both the developed and developing world. The book is intended for an international public health and epidemiology audience including health economists and those interested in tobacco control. [PubMed]

Contenu : 1. Tobacco : history. 2. Tobacco : composition. 3. Nicotine and addiction. 4. Tobacco use : prevalence, trends, influence. 5. Tobacco and health : global burden. 6. Tobacco and cancer. 7. Tobacco and heart disease. 8. Tobacco and cardiovascular disease. 9. Tobacco and respiratory disease. 10. Tobacco and other diseases. 11. Genetics, nicotine addiction and smoking. 12. Tobacco and alcohol. 13. Tobacco control : successes and failures. 14. Treatment of dependence.

Table des matières complète : <http://www.oup.co.uk/jahn/0-19-521887-3>

Notes : Contient : p. 255-266 : Cancer of the prostate / Fabio Levi and Carlo La Vecchia.
 Titre au dos en italien : Tobacco : science, policy and public health.

Descripteur MeSH : Smoking ; Tobacco Use Disorder
 Qualifiant(s) : Smoking/adverse effects ; Smoking/epidemiology ; Smoking/prevention & control ; Tobacco Use Disorder ; epidemiology ; tobacco use Disorder/complications

Document introduit le : 13.09.2004

Liste d'exemplaires :

No	Bibliothèque	Date	Notes	Disponibilité
1	IJMSIP	SA-2449	No d'inventaire : 193/04. - Arch IJMSIP	emprunté

* Voir aussi : Cancer of the prostate

Source : <http://www.saphirdoc.ch/permalien.htm?saphirid=44787>

Si l'affichage à la demande dans l'OPAC de contenus externes est intéressant et relativement facile à mettre en place, il a le désavantage de ne pas enrichir le contenu de la base de données. Si les tables des matières et les résumés ne sont pas indexés comme les autres champs, ils ne sont pas recherchables en même temps que les données du catalogue. D'autres services permettent de remédier à cet inconvénient en permettant d'importer dans la base de données des résumés et tables des matières qui sont alors vendus, et non plus simplement accessibles sous licence comme dans le cas précédent.

Échanges des notices entre catalogues

Avec l'utilisation des AACR⁵³ comme règles communes de catalogage et, grâce à des outils simples de pompage⁵⁴, intégrés au système de gestion et qui utilisent le protocole Z39.50 mais aussi des techniques d'extraction à partir de pages web, les

bibliothèques ont commencé plus que jamais à réutiliser des notices en provenance d'autres catalogues.

En dépit de la réticence de certains catalogueurs qui voient là un risque de perte de créativité (Gavin, 2006), cet usage est devenu très courant et soutenu⁵⁵ par la plupart des bibliothèques des réseaux suisses (RERO, IDS⁵⁶, etc.)

Importation et syndication des notices d'articles

L'explosion de la production scientifique a poussé depuis quelques années les producteurs de bases de données bibliographiques à développer des partenariats forts avec les éditeurs. Par exemple, la NLM⁵⁷, qui produit une base de la taille et de l'importance de PubMed⁵⁸, a cessé en 2006 d'introduire manuellement les métadonnées dans sa base, grâce au développement à grande échelle de l'importation des notices d'articles en format XML fournis par les éditeurs⁵⁹. Cette technique d'importation des flux XML permet à PubMed de diffuser très rapidement l'information reçue, car les contrôles et les améliorations successives des notices (corrections, indexation...) se font progressivement sans entraver la diffusion des références, qui gardent le même identificateur quel que soit le stade du traitement dans le système. Lorsque les éditeurs ne sont pas à même de fournir les notices dans ce format XML, un système de numérisation puis d'OCR⁶⁰ est utilisé pour importer les références avec les résumés dans la base.

Ces partenariats ont toujours fait défaut dans le domaine des bibliothèques, où la relation avec les éditeurs n'est pas de toute évidence, car l'un des rôles subversifs de la bibliothèque provient du fait qu'elle donne accès, à moindre coût, à des contenus soumis au droit d'auteur (Le Moal, 2004) et surtout aux lois du commerce. Du moment où l'hypertexte entre en jeu, les bibliothèques ont un argument pour la négociation, car elles peuvent rendre service aux éditeurs en ajoutant des liens vers leurs plateformes

de commerce électronique (ou via des librairies en ligne), ce qui pourrait permettre de développer des collaborations plus importantes.

Une autre possibilité à explorer pour les bases qui ont des ressources plus modestes est d'utiliser les flux RSS⁶¹ offerts par les éditeurs comme source pour l'importation des données dans le catalogue. Le projet TOCRRoS⁶² va dans ce sens, en permettant d'ajouter automatiquement et périodiquement les articles publiés par les revues pour lesquelles la bibliothèque dispose d'un abonnement en cours. Le logiciel de gestion Alexandria permet depuis la version 6 d'importer automatiquement des contenus en provenance des flux RSS externes

Sans arriver à l'intégration de ces informations dans le catalogue lui-même, une dernière possibilité consiste à afficher le contenu du dernier numéro d'un périodique proposant un flux RSS, au moment de l'affichage de la notice complète dans l'OPAC (Iriarte, 2006). Cette possibilité nécessite l'enregistrement préalable de l'adresse du flux RSS dans le catalogue, pour ensuite pouvoir utiliser de façon contextuelle des services de conversion *RSS* -> *Javascript* offerts par différents sites⁶³.

Importation des notices en provenance des archives ouvertes : utilisation du protocole *OAI-PMH*⁶⁴

De la même façon qu'aujourd'hui nous pouvons importer dans le catalogue des informations en provenance des flux XML (en format ONIX⁶⁵, RSS ou autre) proposés par les éditeurs, il serait possible d'utiliser le protocole *OAI-PMH* pour ajouter des références en provenance des archives ouvertes ou des dépôts institutionnels comme e-prints⁶⁶, arXiv⁶⁷, HAL⁶⁸, RERO DOC⁶⁹, Infoscience⁷⁰, etc.

Cette possibilité de moissonner des serveurs de documents à partir du catalogue ne semble pas, à notre connaissance, avoir été exploitée dans les bibliothèques. C'est plutôt dans le sens

inverse, c'est-à-dire l'intégration des notices du catalogue dans une base externe, que nous pouvons trouver des réalisations, comme dans les cas des notices de l'OPAC intégrées au serveur institutionnel du CERN (CERN Document Server⁷¹) ou dans la plateforme Infoscience de l'EPFL⁷².

Cette intégration des références provenant de sources hétérogènes dans une couche supplémentaire, permet une plus grande souplesse : il n'y a pas le risque de toucher aux notices qui servent en même temps à la gestion, et éviter d'appliquer aux données importées les mêmes critères de qualité et de sécurité que pour les notices du catalogue. La base de données située dans cette couche (par exemple WorldCat⁷³, Dreiländerkatalog, TEL⁷⁴, etc.) peut alors proposer de nouveaux services et peut être ouverte aux contenus générés par les utilisateurs, dans la ligne des outils sociaux ou *web 2.0*⁷⁵. Dans ce sens il y a un nouveau marché qui se développe, avec des nouveaux outils orientés *web 2.0* qui sont proposés aux bibliothèques comme RLG⁷⁶, Primo⁷⁷, etc.

Texte intégral

L'année 2006 a vu le développement de la campagne de numérisation de livres à grande échelle, « Google Books »⁷⁸, entreprise par la société Google en 2004 en collaboration avec deux groupes de partenaires : d'un côté certains éditeurs et de l'autre un groupe restreint de bibliothèques⁷⁹. D'abord appelé « Google Print », le nom du programme est devenu « Google Books » en 2006, pour palier entre autres aux tensions entre Google et certains éditeurs.

Sans complexe, Google Books affiche l'ambition « *de travailler avec des éditeurs et des bibliothèques pour créer un catalogue virtuel complet de tous les livres et dans toutes les langues, dans lequel les internautes pourront effectuer des recherches.* »

Si ce projet reste encore très controversé (Salaün, 2005), il a eu le mérite de réveiller la communauté des bibliothèques qui, avec le Président de la Bibliothèque Nationale de France M. Jeanneney à la tête, a réagi de façon active en donnant une impulsion plus forte au projet de la bibliothèque numérique européenne⁸⁰. D'autres bibliothèques ont aussi annoncé des projets de numérisation à large échelle. La *British Library*⁸¹ a conclu un partenariat avec la société Microsoft⁸². De leur côté *Yahoo!* et la fondation *Internet Archive*⁸³ ont aussi annoncé le début de sa propre campagne « *Open Content Alliance* »⁸⁴. La *Library of Congress* est, quant à elle, en discussion avec l'UNESCO pour amorcer le lancement d'une Bibliothèque numérique mondiale⁸⁵.

Le projet de Google a aussi commencé à porter ses fruits et une grande quantité d'information contenue dans ces bibliothèques est devenue accessible, même si la qualité de numérisation peut être jugée décevante⁸⁶. A l'heure actuelle, 10'000 éditeurs et 13 bibliothèques font partie du projet dans son ensemble et, selon les chiffres donnés par Google, le nombre des livres dans son index avoisine le million⁸⁷.

Selon le contrat que les bibliothèques participant au volet « *Google Books for Libraries* » ont signé avec la société Google⁸⁸, les fichiers issus de la numérisation ne pourront pas être diffusés par la bibliothèque sans l'accord de Google, ce qui limite leur utilisation dans le catalogue. Malgré cette entrave commerciale, la bibliothèque de l'Université de Michigan⁸⁹ propose de visualiser le document numérisé⁹⁰ en format image, texte ou pdf dans sa propre plateforme digitale⁹¹ par le biais d'un lien figurant dans la notice complète du catalogue, parallèlement au lien sur le même document dans la plateforme de Google :



Source :

http://mirlyn.lib.umich.edu/F?func=find-b&find_code=MDN&local_base=MIU01_PUB&request=39015014807104

Selon ce même principe, les bibliothèques établissent aujourd'hui de plus en plus des liens vers le texte intégral des documents situés sur des serveurs externes au catalogue : articles de périodiques électroniques (en *open access* ou payants), documents disponibles en libre accès dans les bibliothèques numériques nationales ou internationales (*Gallica*⁹², *Projet Gutenberg*⁹³, *Biblioteca Virtual Cervantes*⁹⁴...), *e-prints* et thèses des archives ouvertes, etc. Cependant, dans la plupart des cas, ce lien est créé uniquement manuellement (avec les risques que cela comporte, comme la faible pérennité des liens quand on sort du cadre du DOI⁹⁵) et facultativement au moment de la création d'une nouvelle notice et non pas de façon rétrospective. Vu le rythme des campagnes de numérisation en cours, cela signifie que de plus en plus de notices dans nos catalogues resteront sans lien hypertexte avec la version électronique disponible pourtant quelque part sur Internet. Seul un outil performant de gestion des liens peut servir de solution à ce problème, comme nous l'évoquerons dans la deuxième partie, seulement. Après les difficultés rencontrées au moment vouloir intégrer les revues en format électronique au catalogue, nous risquons de nouveau de voir s'agrandir le fossé entre catalogue et ressources en texte intégral en ligne.

Concernant la recherche sur le texte intégral, étant donné que le stockage des documents numériques se situe généralement sur des serveurs

déconnectés du catalogue, la recherche simultanée dans les métadonnées et dans le texte intégral des documents ne peut pas être proposée dans l'OPAC, sauf si c'est le catalogue lui-même qui rejoint cette plateforme des documents numériques, comme nous l'avons vu plus haut dans l'exemple du CERN et de l'EPFL.

Dans le cas de l'Université de Michigan, de même que dans le service « *Search Inside* » d'Amazon⁹⁶, la recherche dans le texte intégral ne peut se réaliser que sur un seul document à la fois. Il faut alors passer par Google Books pour pouvoir effectuer une recherche sur l'ensemble du texte intégral de la collection en même temps que sur les métadonnées fournies par la bibliothèque.

Liens profonds : le rôle des identificateurs

Comme nous venons de constater plus haut, les catalogues de bibliothèques introduisent de plus en plus de liens profonds pointant sur le texte intégral du document répertorié ou sur la notice bibliographique résidant sur une base de données externe comme PubMed, dans le but de donner à l'utilisateur le plus grand nombre d'informations disponibles et de source sûre, concernant le document catalogué : résumé et liens offerts par PubMed, texte intégral ou résumé offert par l'éditeur de la revue ou sur une plateforme *Open Access*, nombre de fois que l'article est cité, offerts par une base de données comme Google Scholar, etc.

L'utilisation de ces liens profonds pose de nouveaux problèmes et de nouveaux défis aux catalogues, qui devraient en plus les maintenir à jour à l'aide des méthodes plus ou moins automatisés. Dans cette recherche de stabilité, seulement les liens profonds basés sur des identificateurs pérennes comme le DOI⁹⁷, le PMID⁹⁸ ou un identificateur OAI-PMH⁹⁹, ont de garanties de perdurer dans le temps. Il est donc évident qu'il faut utiliser ces identificateurs de façon préférentielle pour établir des liens dans le catalogue, et que nous devons les prendre en charge avec le même soin que nous appliquons à l'ISBN :

dans un champ à part bien identifié et avec une syntaxe cohérente et normalisée de type URN¹⁰⁰. Par exemple il serait préférable d'enregistrer le DOI ou le PMID dans un champ ad hoc et de générer l'URL à la demande, au lieu d'enregistrer cette adresse directement dans le champ dédié aux liens Internet :

- doi:10.1000/182 ->
<http://dx.doi.org/10.1000/182>
- pmid:1234 ->
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&list_uids=1234

La raison de cette préférence réside dans le fait qu'il est possible d'utiliser cet identificateur à d'autres fins que celui de construire un URL, comme par exemple l'utiliser dans un OpenURL destiné à trouver d'autres services associés au document ou bien dans un format d'export destiné à des logiciels bibliographiques. Bien qu'il soit toujours possible d'extraire automatiquement l'identificateur à partir de l'URL, c'est bien le cas contraire qui est plus aisé et canonique, d'autant plus que les URLs utilisés par CrossRef¹⁰¹ et par PubMed, bien que garantis pour un bon nombre d'années, ne sont pas des liens pérennes et peuvent varier dans le futur (on peut imaginer par exemple un nouveau protocole qui remplacerait le http, etc.).

Dans la même optique, le catalogage des notices d'articles (aussi appelées notices analytiques) devraient incorporer autant que possible ces identificateurs pérennes ainsi que l'ISSN¹⁰², seul élément d'identification fiable de la revue à laquelle ils appartiennent. Si ces informations, ainsi que la date de publication, le numéro du périodique, du volume, la page de début et celle de fin de l'article étaient codées de façon structurée dans ce type de notices, ce qui est loin d'être le cas aujourd'hui dans la plupart de catalogues de bibliothèques¹⁰³, l'utilisation d'un outil de gestion de liens serait beaucoup plus efficace et permettrait une plus grande précision à l'heure d'identifier l'existence du texte

intégral ou de proposer des services associés via un OpenURL en provenance de l'OPAC.

Contenu généré par les utilisateurs

Suivant l'exemple des sites commerciaux comme Amazon et des outils sociaux de plus en plus nombreux (*Wikio*¹⁰⁴, *del.icio.us*¹⁰⁵, *Flickr*¹⁰⁶, *Connotea*¹⁰⁷...), il est grand temps que les lecteurs puissent déposer des annotations, des tags ou des commentaires par rapport aux livres disponibles dans les catalogues des bibliothèques et, pourquoi pas, les laisser intervenir sur l'indexation des documents avec des tags, sans utiliser les ressources du SIGB, mais en offrant des applications s'appuyant sur des services web comme le fait par exemple *HubMed*¹⁰⁸

Cette possibilité de commenter ou tagger l'information est déjà implémentée dans certains catalogues ou archives ouvertes : WorldCat, Lamson Library¹⁰⁹, Dreiländerkatalog, CERN Document Server... Cependant, elle est absente de la plupart des OPACs classiques gérés par un SIGB.

Utilisation des données de gestion

En s'inspirant d'Amazon, les catalogues ont commencé à exploiter les données générées par l'activité des utilisateurs pour effectuer des suggestions de lecture :

- à partir des données anonymisées en provenance des emprunts, comme le fait la Ann Arbor District Library (AADL)¹¹⁰ : « *Les utilisateurs qui ont emprunté ce document ont aussi emprunté...* »
- à partir des données en provenance de la consultation web, comme le fait RERO DOC : « *Les utilisateurs qui ont vu cette page ont aussi vu...* »

La possibilité d'exploiter de façon dynamique et pondérée les données du document en cours de visualisation, peut aussi conduire à des suggestions de type « *Related articles* » de PubMed ou « *Related Books* » de WorldCat.

Autres services web bibliographiques

En dehors des services web d'Amazon, il y a malheureusement un choix très restreint pour le moment :

- XISBN¹¹¹ (OCLC) : ISBNs en relation avec l'ISBN envoyé (utilisé par le Dreiländerkatalog)
- CrossRef¹¹² : Métadonnées à partir d'un DOI et vice-versa
- PubMed¹¹³ : une des applications les plus avancées dans ce domaine, il met à disposition de la communauté plusieurs services web : métadonnées à partir d'un ou plusieurs PMIDs ou vice-versa, ainsi que les articles liés ou les informations des bases de données connexes (génétique, moléculaire, etc.) à partir de l'identificateur d'une ou de plusieurs références. Il offre aussi par service web la correction orthographique des termes de recherche.

D'autres services web généralistes pourraient être aussi exploités, comme par exemple l'affichage des localisations géographiques des bibliothèques d'un réseau qui possèdent un document en particulier, en utilisant l'API de Google Maps¹¹⁴. Un service web de ce type pourrait aussi être utilisé pour afficher l'emplacement précis dans lequel une image ou un film du catalogue a été prise. Ceci nécessiterait l'ajout des métadonnées géographiques (ou « *geotagging* »¹¹⁵) au catalogage des images fixes ou animées, comme cela se fait déjà sur des plateformes de partage des photos comme Flickr¹¹⁶.

Conclusion

Après cette énumération de services, d'outils et d'informations externes au catalogue dignes d'être incorporés dans cet outil, la question de la surcharge informationnelle pourrait être posée. Toutes ces options ne risquent-elles pas d'engendrer confusion et désorientation ?

La force du catalogue résiderait-elle alors dans son dépouillement ? Nous pouvons en douter. Les catalogues offrent actuellement peu de possibilités pour que les utilisateurs, y compris les professionnels de l'information eux-mêmes, s'approprient cet outil, ce qui est indispensable pour qu'il puisse trouver une place importante dans l'ensemble du paysage informationnel du web, devenu aujourd'hui notre plateforme de travail.

Certes, le risque est grand de se retrouver noyé sous une masse impressionnante d'applications et d'informations. Mais en offrant des données riches, bien structurées et ouvertes à l'extérieur, ainsi qu'une interface ergonomique, simple d'utilisation et d'appropriation par les usagers, le catalogue peut prouver à nouveau son utilité et redevenir ainsi un élément fort dans l'univers d'Internet. Pour y arriver, nous avons parcouru quelques pistes qui vont dans le sens d'une maîtrise des technologies du web par les bibliothèques et dans l'intégration d'éléments externes au catalogue. L'autre aspect clé de la question, la face opposée de la même monnaie, réside dans l'ouverture du catalogue à de nouvelles formes d'utilisation de ses propres données par des tiers, dans un changement de mentalité qui considérerait « le web » comme un utilisateur à part entière.

Dans la deuxième partie de cet article, nous traiterons donc les aspects suivants liés à cette ouverture du catalogue vers la réutilisation de ses données :

1. Citabilité et Permalien
2. OpenURL et COINS
3. RSS
4. Sitemaps
5. Indexation par des moteurs de recherche
6. Export XML pour Google Scholar
7. Open search et SRU/SRW
8. Services web et APIs

Notes

¹ Quelques exemples sortis de la « biblioblogsphère » : *Burn the catalog* <http://www.swarthmore.edu/SocSci/tburke1/perma12004.html> ; *Disintegration, disenchantment, distrust, and development*

http://www.polarislibrary.com/forums/blogs/techtidbits/archive/2006/12/05/Disintegration_2C00_-_disenchantment_2C00_-distrust_2C00_-and-development.aspx

² *If amazon sucked like our old opac* <http://library2.csusm.edu/amazon/index.htm>

³ Caractérisés par le phénomène connu sous le nom de « web 2.0 »

⁴ eXtensible Markup Language <http://www.w3.org/XML/>

⁵ <http://www.opensource.org>

⁶ <http://openaccess.inist.fr>

⁷ <http://www.openarchives.org>

⁸ <http://openurl.info/registry>

⁹ Uniform Resource Locator <http://www.w3.org/Addressing/URL/Overview.html>

¹⁰ <http://www.opensearch.org>

¹¹ Si ce terme est complètement nouveau, le concept ne l'est pas. Voir par exemple le projet « eXtensible Catalog (XC) » de l'Université de Rochester à New-York <http://www.extensiblecatalog.info/> et <http://www.rochester.edu/news/show.php?id=2518>

¹² Nouvelles Technologies de l'Information et la Communication

¹³ Pour suivre l'évolution et les nouveaux OPACs innovateurs, vous pouvez utiliser la liste de diffusion « *NGC4Lib - Next Generation Catalogs for Libraries* » et ses archives :

<http://dewey.library.nd.edu/mailling-lists/ngc4lib/>

¹⁴ SIBIL était à l'origine l'acronyme de « Système intégré pour les bibliothèques universitaires

de Lausanne » (Gavin, 1997)

¹⁵ <http://www.unil.ch/bcu>

¹⁶ MACHine-Readable Cataloging <http://www.loc.gov/marc/>

¹⁷ Conseil Européen pour la Recherche Nucléaire. Aujourd'hui le nom CERN

désigne l'Organisation européenne pour la Recherche nucléaire <http://www.cern.ch>

¹⁸ <http://www.slac.stanford.edu>

¹⁹ <http://news-service.stanford.edu/news/2001/april11/addris-411.html>

²⁰ Content Management System

²¹ Outil de publication web instantanée et ouvert aux modifications des utilisateurs. Il est utilisé par exemple pour le projet Wikipédia

<http://fr.wikipedia.org/wiki/Wiki>

²² Devenue norme ISO 23950 en 1998, ce protocole s'appelaient dans sa version d'origine « *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, ANSI/NISO Z39.50-1995* ». Il s'agit d'un protocole antérieur au web et qui spécifie des structures de données et les règles d'échange qui permettent à une machine client (nommé « *origin* ») de chercher des données dans un serveur (nommé « *target* ») et de d'obtenir les entrées résultant de cette recherche <http://www.loc.gov/z3950/agency/resources/>

²³ <http://www.loc.gov>

²⁴ La norme Z39.50, malheureusement très peu utilisée en dehors du domaine des bibliothèques, n'est pas exploitable à travers le protocole HTTP et il faut donc un logiciel spécifique pour l'utiliser. L'apparition de XML et les services web associés qui se développent un peu partout, sont en train de précipiter son déclin.

²⁵ Karlsruher Virtuelle Katalog

<http://www.ubka.uni-karlsruhe.de/kvk.html>

²⁶ Outil de publication web personnelle appelé aussi *weblog*, *carnet web*, *joueb*... Pour plus d'information voir l'article de Wikipédia :

<http://fr.wikipedia.org/wiki/Blog>

²⁷ Contraction de « iPod » et de « broadcasting ». Forme de flux RSS auquel on ajoute des fichiers sonores qui sont alors disponibles directement à partir du lecteur RSS ou téléchargeables automatiquement dans un baladeur numérique. Voir aussi la définition de Wikipédia : <http://fr.wikipedia.org/wiki/Podcasting>

²⁸ Aussi appelés « Link resolvers », ces outils permettent de générer dynamiquement une liste de liens cibles à partir des métadonnées de la source : lien vers le texte intégral ou vers le formulaire de commande, recherches par ISSN, auteur, titre ou descripteur dans les catalogues, etc. En Suisse l'outil le plus utilisé reste SFX <http://www.exlibrisgroup.com/sfx.htm> commercialisé par la société Ex.libris qui possède aussi le SIGB *Aleph* utilisé par le réseau suisse alémanique IDS.

²⁹ Les catalogues donnent pour le moment très peu des liens externes. Par exemple pour le catalogue collectif du réseau romand (RERO) seulement le 1% de notices portent un URL (environ 30.000 sur 3 millions)

³⁰ PHP est l'acronyme récursif de « PHP Hypertext Preprocessor ». PHP est un langage de script qui est très utilisé pour créer des sites web dynamiques. Site officiel : <http://www.php.net>. MySQL est un logiciel libre de gestion de bases de données de type SQL (Structured Query Language). Site officiel : <http://www.mysql.com>

³¹ « Google Docs & Spreadsheets » <http://docs.google.com> et « Think free » <http://www.thinkfree.com> sont déjà des bons exemples de cette évolution. Les SIGB open source disponibles sur le marché (PMB <http://www.sigb.net>, Koha <http://www.koha.org>, OpenBiblio <http://openbiblio.sourceforge.net> ...) sont pour la plupart déjà des systèmes « *full-web* » car, autant la gestion, le paramétrage, le catalogage que la consultation se font à travers le navigateur sans utiliser des clients lourds comme dans les systèmes propriétaires actuels, où seules la consultation et certaines options liées à la gestion des prêts et des lecteurs (inscriptions, réservations, prolongations, PEB, etc.) passent à travers le web

³² Voir les statistiques publiées par l'OFS « Utilisation d'Internet dans les ménages en Suisse : Résultats de l'enquête 2004 et indicateurs » et qui montrent l'augmentation très forte de l'utilisation d'internet au sein de la population suisse

ces dernières années : http://www.bfs.admin.ch/bfs/portal/fr/index/themen/kultur_medien_zeitverwendung/kultur/blank/publikationen.html?publicationID=2487. Ces chiffres suivent la tendance générale annoncée par l'UIT dans son rapport « digital.life » <http://www.itu.int/digitalife> et qui donnent, pour une personne entre 18 et 54 ans, une consommation moyenne hebdomadaire de médias numériques de 16 heures, tandis que de 13 heures pour la TV, 8 pour la radio, 4 pour les journaux et quotidiens papier et 1 pour le cinéma

³³ WYSIWYG est l'acronyme de la locution anglaise « What You See Is What You Get ». Les interfaces de ce type sont utilisées dans les logiciels de mise en page et surtout dans les plateformes de blogging comme outil pour pouvoir écrire facilement pour le web sans connaître le langage HTML

³⁴ Le fait de copier/coller du contenu en provenance d'une page web dans le corps d'un billet d'un blog peut poser des problèmes d'affichage car l'éditeur WYSIWYG cache le code HTML. Ce dernier peut pourtant contenir des balises pouvant interférer avec le code de la page du blog

³⁵ Généré automatiquement au moment de l'affichage

³⁶ <http://www.rero.ch/page.php?section=zone&pageid=856>

³⁷ C'est encore le cas des catalogues du réseau IDS et ce fut aussi le cas pendant longtemps pour les catalogues du réseau romand. Cependant, RERO a changé son approche et dans la nouvelle version de son OPAC <http://opac.rero.ch>, introduite depuis le 8 janvier 2007, les liens de la zone 856 sont désormais affichés déjà au niveau de la liste de résultats

³⁸ RERO est l'acronyme de « REseau Romand », et désigne le réseau des bibliothèques de Suisse occidentale majoritairement de langue française <http://www.rero.ch>

³⁹ Par exemple http://opac.rero.ch/get_bib_record.cgi?rero_id=R277678560

⁴⁰ Par exemple http://opac.rero.ch/get_bib_record.cgi?rero_id=R003636602

⁴¹ SAPHIR (Swiss Automated Public Health Information Resources) <http://www.saphirdoc.ch>. Base documentaire suisse spécialisée en santé publique et dont le CDSP est le responsable.

⁴² http://www.gbconcept.com/pro_alexandrie.html

⁴³ <http://www.refworks.com>

⁴⁴ <http://www.amazon.com>

⁴⁵ <http://www.electre.com>

⁴⁶ <http://www.syndetics.com>

⁴⁷ « International Standard Book Number ». l'ISBN est un identificateur international, défini par la norme ISO 2108, et qui sert à identifier sans ambiguïté chaque livre. Existant depuis 1972, les ISBN sont attribués et gérés par un réseau d'agences réparties dans 166 pays, avec une centrale à Londres. En suisse romande, c'était l'agence francophone pour la numérotation internationale du livre (AFNIL) qui gérait les numéros ISBN jusqu'à 1994. Depuis cette date, la gestion est assurée par la « Schweizer Buchhändler- und Verleger-Verband SBVV »

<http://www.swissbooks.ch/prestations/isbn/uebersicht.shtm>. D'abord constitué par 10 chiffres significatives, il a passé à 13 depuis janvier 2007 devenant ainsi compatible avec les codes-barre de la norme EAN 13 <http://isbn-international.org>

⁴⁸ http://www.amazon.com/AWS-home-page-Money/b/ref=sc_iw_l_0/103-1555994-9747843?ie=UTF8&node=3435361&no=3435361&me=A36L942TSJ2AJA

⁴⁹ <http://suchen.hbz-nrw.de/dreilaender/>

⁵⁰ Projet expérimental d'OPAC <http://www.plymouth.edu/library/opac/> basé sur la plateforme de blogging WordPress <http://wordpress.org>

⁵¹ Asynchronous JavaScript And XML. C'est un ensemble de techniques qui permet à une page web d'échanger des informations externes sans devoir être

actualisée. Voir l'article fondateur de Jesse James Garrett « Ajax: A New Approach to Web Applications »

<http://www.adaptivepath.com/publications/essays/archives/000385.php>

⁵² La technique est expliquée en détail sur le blog de la BiUM <http://www.bium.ch/blog/?p=106>

⁵³ Anglo-American Cataloguing Rules, 2nd Edition <http://www.aacr2.org>

⁵⁴ Comme par exemple « EZPump (EZP) » <http://www.ngscan.com/easypump/index.htm>, logiciel de pompage des notices bibliographiques avec client Z39.50, développé par un bibliothécaire de la Médiathèque Valais et utilisé par les bibliothèques de RERO

⁵⁵ Voir *La Lettre de RERO*, 2006-4 http://www.rero.ch/pdfview.php?section=lalettre&filename=LaLettre2006_04.pdf

⁵⁶ IDS est l'acronyme de « Informationsverbund Deutschschweiz » et désigne le réseau des bibliothèques de Suisse orientale, majoritairement de langue allemande <http://www.informationsverbund.ch>

⁵⁷ <http://www.nlm.nih.gov>

⁵⁸ <http://www.pubmed.org>

⁵⁹

http://www.nlm.nih.gov/bsd/bsd_key.html

⁶⁰ Optical Character Recognition

⁶¹ RSS est utilisé comme acronyme de « Really Simple Syndication », « Rich Site Summary », « RDF Site Summary » ou une autre variante de ces termes. Pour plus de détails, voir la page explicative faite par l'ADBS :

<http://www.adbs.fr/site/repertoires/outils/rss.php>

⁶² TOCRoSS http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_tocross.asp

X

⁶³ Par exemple <http://itde.vccs.edu/rss2js/build.php> ou <http://www.rss-to-javascript.com>

⁶⁴ « Open Archive Initiative and Protocol for Metadata Harvesting »

<http://www.openarchives.org/pmh/>

⁶⁵ <http://www.editeur.org/onix.html>

⁶⁶ <http://www.eprints.org>

⁶⁷ <http://arxiv.org>

⁶⁸ <http://hal.archives-ouvertes.fr>

⁶⁹ <http://doc.rero.ch>

⁷⁰ <http://infoscience.epfl.ch>

⁷¹ <http://cdsweb.cern.ch>

⁷² <http://www.epfl.ch>

⁷³ <http://www.worldcat.org>

⁷⁴ The European Library : <http://www.theeuropeanlibrary.org>

⁷⁵ Voir par exemple l'article de wikipedia http://en.wikipedia.org/wiki/Web_2 et l'article fondateur de Tim O'Reilly « What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software »

<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>. Version française :

<http://web2rules.blogspot.com/2006/01/what-is-web-20-par-tim-oreilly-version.html>

⁷⁶ <http://www.rlg.org>

⁷⁷

Primo <http://www.exlibrisgroup.com/webinar/1144862525.htm>

⁷⁸ <http://books.google.com> et depuis quelque temps seulement aussi consultable sur <http://books.google.ch>

⁷⁹ D'abord limité à 5, maintenant ce sont 13 bibliothèques qui fournissent les documents en échange des données obtenues par la numérisation : la bibliothèque de l'Université de Virginie, de Californie, de Harvard, de Stanford, du Michigan, du Wisconsin-Madison, de Texas, de Princeton, de Madrid, d'Oxford ainsi que la bibliothèque publique de New York et les bibliothèques de Catalogne et de Bavière. La bibliothèque du Congrès américain (Library of Congress) fait l'objet d'un autre projet à part : <http://www.washingtonpost.com/wp-dyn/content/article/2005/11/21/AR2005112101428.html>

⁸⁰ Voir le communiqué de presse de la commission européenne : <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/06/253&format=HTML&aged=0&language=FR&guiLanguage=en> et la page du programme de l'UE « i2010: Digital Libraries Initiative » : http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm

⁸¹ <http://www.bl.uk>

⁸² <http://www.microsoft.com>
⁸³ <http://www.archive.org>
⁸⁴ <http://www.opencontentalliance.org>.

Voir aussi l'article de CNET News.com :
« Yahoo to digitize public domain books »
http://news.com.com/Yahoo+to+digitize+public+domain+books/2100-1038_3-5887374.html

⁸⁵ Voir l'annonce fait par l'UNESCO :
http://portal.unesco.org/fr/ev.php-URL_ID=35949&URL_DO=DO_TOPIC&URL_SECTION=201.html

⁸⁶ Voir par exemple l'article et le billet de Lorcan Dempsey « Fingering volumes »
<http://orweblog.oclc.org/archives/001122.html>, « Le travail bâclé de Google Print »
<http://www.dsi-info.ca/moteurs-de-recherche/2005/11/le-travail-bcl-de-google-print.html> ou « Digitized by Google »
<http://e-benedictins.blogspot.com/2006/09/digitalized-by-google.html>

⁸⁷ <http://librariancentral.blogspot.com/2007/03/checking-in-with-google-book-search.html>

⁸⁸ Voir l'article de *LibraryJournal.com* « Release of Google Contract with UC Sparks Criticism »
<http://www.libraryjournal.com/article/CA6367340.html>

⁸⁹ <http://lib.umich.edu/>

⁹⁰ Par exemple, la page 7 du document « Versuch schweizerischer Gedichte » :
<http://mdp.lib.umich.edu/cgi/m/mdp/pt?seq=7&size=100&id=39015014807104&view=text>. Voir aussi à ce sujet l'article de *The Chronicle of Higher Education* : « U. of Michigan Adds Books Digitized by Google to Online Catalog, but Limits Use of Some »
<http://chronicle.com/free/2006/08/2006083101t.htm>

⁹¹ « Mbooks » :
<http://www.lib.umich.edu/mdp/>

⁹² <http://gallica.bnf.fr>

⁹³ <http://www.gutenberg.org>

⁹⁴ <http://www.cervantesvirtual.com>

⁹⁵ Document Object Identifier :
<http://www.doi.org>

⁹⁶ <http://www.amazon.com/Search-Inside-Book-Books/b?ie=UTF8&node=10197021>

⁹⁷ <http://www.doi.org>

⁹⁸ <http://en.wikipedia.org/wiki/PMID> ou
<http://pmid.us/>

⁹⁹ <http://www.openarchives.org/OAI/openarchivesprotocol.html#UniquelIdentifieur>

¹⁰⁰ http://fr.wikipedia.org/wiki/Uniform_Resource_Name

¹⁰¹ Organisation chargée de gérer les DOI
<http://www.crossref.org>

¹⁰² L'ISSN (International Standard Serial Number) est un numéro à huit chiffres non significatives de la forme 1234-5678 et qui identifie les périodiques, y compris en format électronique. La gestion des ISSN, qui compte plus d'un million aujourd'hui, est effectuée par un réseau mondial de 80 centres nationaux (dont la Bibliothèque Nationale Suisse) coordonnés par un centre international à Paris
<http://www.issn.org/fr>

¹⁰³ En effet dans la majorité de notices analytiques présentes dans les catalogues des bibliothèques universitaires, l'ISSN de la revue est absent. Les autres éléments clés pour identifier un article, tels le volume, et les pages de début et de fin, sont répertoriés dans une zone « In » peu normalisée et difficilement exploitables sans un traitement informatique

¹⁰⁴ <http://www.wikio.com>

¹⁰⁵ <http://del.icio.us>

¹⁰⁶ <http://flickr.com>

¹⁰⁷ <http://www.connotea.org>

¹⁰⁸ Interface alternative à PubMed utilisant les services web de cette dernière en ajoutant des nouvelles fonctionnalités comme le « tagging » ou la catégorisation par facettes :
<http://www.hubmed.org>

¹⁰⁹ <http://www.plymouth.edu/library/opac/>
¹¹⁰ <http://www.aadl.org/catalog>

¹¹¹ <http://www.oclc.org/research/projects/xisbn/>

¹¹² http://www.crossref.org/02publishers/op_enurl_info.html

¹¹³ http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html

¹¹⁴ <http://www.google.com/apis/maps/>

¹¹⁵ Voir l'article de Wikipedia
<http://en.wikipedia.org/wiki/GeoTagging>

¹¹⁶
<http://www.flickr.com/groups/geotagging/>

BIBLIOGRAPHIE

BBF (2005). Dossier : Mort et transfiguration des catalogues. *BBF : Bulletin des Bibliothèques de France* [en ligne], [consulté le 15 janvier 2007], T. 50, n° 4.
<http://bbf.enssib.fr/sdx/BBF/frontoffice/2005/04/sommaire.xsp>

BEARMAN, David (décembre 2006). Jean-Noël Jeanneney's Critique of Google: Private Sector Book Digitization and Digital Library Policy. *D-Lib Magazine* [en ligne], [consulté le 15 janvier 2007], vol. 12, n° 12.
<http://www.dlib.org/dlib/december06/bearman/12bearman.html>

BIBLIOGRAPHIC SERVICES TASK FORCE (décembre 2005). *Rethinking How We Provide Bibliographic Services for the University of California* [en ligne]. The University of California Libraries [consulté le 15 janvier 2007].
<http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>

BROUDOUX, Evelyne, GRÉSILLAUD, Sylvie, LE CROSNIER, Hervé, LUX-POGODALLA, Véronika (18 septembre 2005). Construction de l'auteur autour de ses modes d'écriture et de publication. *H2PTM'05* [en ligne], [consulté le 15 janvier 2007].
http://archivesic.ccsd.cnrs.fr/sic_00001552

CALHOUN, Karen, (2006). The Changing Nature of the Catalog and its Integration with Other Discovery Tools [en ligne]. Final report prepared for the Library of Congress, March 17, [consulté le 15 janvier 2007].
<http://www.loc.gov/catdir/calhoun-report-final.pdf>

ÇELIKBAŞ, Zeki (novembre 2004). What is RSS and how it can serve libraries. *E-prints in Library and Information Science* [en ligne], [consulté le 15 janvier 2007].
<http://eprints.rclis.org/archive/00002531/>

CHUDNOV, Daniel, CAMERON, Richard, FRUMKIN, Jeremy, SINGER, Ross, YEE, Raymond (avril 2005). Opening up OpenURLs with Autodiscovery. *Ariadne* [en ligne], issue 43, [consulté le 15 janvier 2007].
<http://www.ariadne.ac.uk/issue43/chudnov/>

DEMPSEY, Lorcan (2006). The Library Catalogue in the New Discovery Environment: Some Thoughts. *Ariadne* [en ligne], issue 48, [consulté le 15 janvier 2007].
<http://www.ariadne.ac.uk/issue48/dempsey/>

DEMPSEY, Lorcan (22 février 2005). The integrated library system that isn't. *Lorcan Dempsey's weblog On libraries, services and networks* [en ligne], [consulté le 15 janvier 2007].
<http://www.ariadne.ac.uk/issue48/dempsey/>

DUCHEMIN, Pierre-Yves (2005). L'enrichissement des catalogues ? Et après ? *BBF* [en ligne], n° 4, p. 21-27 [consulté le 15 janvier 2007].
<http://bbf.enssib.fr>

GARREAU, Angéline (19 septembre 2005). *Les blogs entre outil de publication et espace de communication : un nouvel outil pour les professionnels de la documentation* [en ligne]. Maîtrise des sciences de l'information et de la documentation, CAO, Université, [consulté le 15 janvier 2007].
http://memsic.ccsd.cnrs.fr/mem_0000073.html

GAVIN, Pierre (1997). *SIBIL : un bilan pour le passé, et quelques jalons pour le futur* [en ligne]. Lausanne : Nouvelle Association REBUS (Réseau des

bibliothèques utilisant SIBIL), [consulté le 15 janvier 2007]. <http://www.pierregavin.ch/documents/Sibil-bilan-jalon.pdf>

GAVIN, Pierre (2006). Les AACR2 : menace ou chance ? *Hors texte*, n° 80, p. 16-18

HAMMOND, Tony, HANNAY, Timo, LUND, Ben (décembre 2004). The Role of RSS in Science Publishing : Syndication and Annotation on the Web. *D-Lib Magazine* [en ligne], vol. 10, n°12, [consulté le 15 janvier 2007]. <http://www.dlib.org/dlib/december04/hammond/12hammond.html>

IRIARTE, Pablo (2006). La diffusion de l'information documentaire et des actualités en format RSS. In CHARTRON, Ghislaine [dir.], BROUDOUX, Evelyne [dir.]. *Document numérique et société : actes de la conférence organisée dans le cadre de la semaine du document numérique à Fribourg (Suisse) les 20 et 21 septembre 2006* [imprimé]. Paris : ADBS, [consulté le 15 janvier 2007]. P. 123-148. http://archivesic.ccsd.cnrs.fr/sic_00079211

LEBOEUF, Patrick (2004). Le jour d'après : où serez-vous ?... [en ligne]. *Journée d'étude Médiadix (Jeudi 21 Octobre 2004) : La fin du catalogage ?!* [consulté le 15 janvier 2007]. <http://netx.u-paris10.fr/mediadix/archivesje/leboeufweb.pdf>

LE MOAL, Jean-Claude [coord.], HIDOINE, Bernard [coord.], CALDERAN, Lissette [coord.] (2004). *Publier sur Internet : séminaire INRIA, 27 septembre - 1er octobre 2004, Aix-les-Bains* [imprimé]. Paris : ADBS. ISBN 2843650720

MARKEY, Karen (janvier 2007). The Online Library Catalog : Paradise Lost and Paradise Regained?. *D-Lib Magazine* [en ligne], [consulté le 15 janvier 2007], vol. 13, n° 1/2. <http://www.dlib.org/dlib/january07/markey/01markey.html>

MOFFAT, M (Mars 2006). "Marketing" with Metadata : How Metadata Can Increase Exposure and Visibility of Online Content [en ligne], version 1.0, 8 [consulté le 15 janvier 2007] <http://www.icbl.hw.ac.uk/perx/advocacy/exposingmetadata.htm>

SALAÜN, Jean-Michel (décembre 2005). Bibliothèques numériques et Google-Print. *Version non révisée par l'éditeur de l'article pour la revue Regard sur l'actualité* [en ligne], [consulté le 15 janvier 2007], n° 316. http://archivesic.ccsd.cnrs.fr/sic_00001576

SCHNEIDER, Karen G (20 Mai 2006). How OPACs Suck, Part 3: The Big Picture. *ALA TechSource* [en ligne], [consulté le 15 janvier 2007]. <http://www.techsource.ala.org/blog/2006/05/how-opacs-suck-part-3-the-big-picture.html>

STEPHENS, Michael (2006). Web 2.0 & Libraries : Best Practices for Social Software. *Library Technology Reports*, vol. 42, no. 4. ISSN 0024-2586

TENAILLEAU, Willy (14 mars 2006). Les services à distance d'une médiathèque - synthèse 3 : Les notices bibliographiques. *LaConjuration/notes* [en ligne], [consulté le 15 janvier 2007]. <http://www.laconjuration.net/notes/?2006/03/14/33-les-services-a-distance-d-une-mediathèque-synthèse-3-les-notices-bibliographiques>