



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

août 22, 2005

Code Number: 124-F
Meeting: 133 SI - Bibliography

Recommandations pour une amélioration urgente des catalogues publics (OPACs) Rôle des agences bibliographiques nationales dans cette tâche

John D. Byrum, Jr.
Chief, Regional & Cooperative Cataloging
Library of Congress
Washington, D. C. 20540
USA

*Traduction : Bernard Dubourg,
tél : 06 22 94 08 78 ;
bernard.dubourg@bdic.fr*

Résumé

Aujourd'hui, conditionnées par les moteurs de recherche du Web, les personnes en quête d'information s'attendent à être satisfaites immédiatement. C'est une conséquence de la convivialité d'internet. A l'inverse, on remarque de plus en plus que les catalogues publics traditionnels des bibliothèques (OPACs) ne procurent pas les mêmes facilités d'usage ou d'accès à l'information. Les agences bibliographiques nationales (ABN) et, partout, les bibliothèques sont dans l'obligation de réagir face à cette disparité, en prenant des mesures destinées à introduire, dans leurs produits bibliographiques et leurs bases de données, bien plus d'informations que celles qui y sont actuellement saisies à titre de ressource. Dans le même temps, les agences bibliographiques nationales (ABN) doivent répondre à la nécessité d'une nouvelle génération d'OPACs, avec une offre de fonctionnalités largement améliorée qui puisse, pour l'essentiel, s'inspirer des critères classiques propres aux moteurs de recherche et aux librairies en ligne. Etant donné que d'autres choix s'offrent à la recherche documentaire, il s'avère nécessaire de considérer sans délai ces besoins, si les agences

bibliographiques nationales et les bibliothèques veulent qu'en ce vingt-et-unième siècle les usagers continuent d'être satisfaits.

Cette présentation débute par le constat qu'au cours des dix dernières années, les attentes des usagers se sont considérablement accrues en matière de satisfaction liée aux recherches documentaires menées après avoir utilisé des ressources électroniques. Actuellement, les utilisateurs de bibliographies nationales et de catalogues publics de bibliothèque espèrent avoir accès à une information qui va bien au-delà de ce que peut proposer une approche traditionnelle de la description bibliographique et de l'accès. Je vais considérer cette évolution selon deux perspectives : 1) la nécessité pour les bases de données bibliographiques de donner accès à bien plus de contenu que les éléments traditionnellement proposés ; et (2) la nécessité d'une nouvelle génération d'OPACs offrant des fonctionnalités considérablement améliorées, basées essentiellement sur les caractéristiques types des moteurs de recherche et des librairies en ligne. Du fait de leur position éminente, les Agences bibliographiques nationales (ABN) sont bien placées pour ouvrir la voie et répondre à ces nécessités, grâce à l'amélioration de services dont elles doivent amorcer le démarrage, grâce aussi à leur pouvoir d'influence sur les concepteurs et les vendeurs de systèmes intégrés de bibliothèques (SIGB) et outils analogues.

Au siècle dernier, les agences bibliographiques nationales ainsi que d'autres agences de catalogage ont cherché à contrôler leurs ressources sur le plan de l'inventaire et sur le plan bibliographique, en produisant des fiches bibliographiques. L'émergence des descriptions bibliographiques internationales standardisées (ISBD), identifiant des éléments et leur ordre pour composer la description ainsi que la standardisation de formats tels UNIMARC et MARC 21, ont favorisé un accroissement significatif des échanges de notices. L'apparition des Spécifications fonctionnelles des notices bibliographiques (FRBR) et l'évolution continue d'un code de catalogage international (ICC) représentent deux initiatives de l'IFLA qui visent à promouvoir l'objectif d'un contrôle bibliographique international. Aussi précieuses et importantes soient-elles, ces réalisations ou d'autres se concentrent en général sur la création et la communication de descriptions bibliographiques traditionnelles et n'ont répondu, pour la plupart, qu'aux critères de base des ressources documentaires ; en d'autres mots, ces descriptions ont souvent été limitées pour l'essentiel aux données obligatoires établies par l'ISBD¹.

Aujourd'hui, conditionnés par leurs pratiques sur Internet, les usagers s'attendent à accéder rapidement et facilement à l'information, puis à évaluer des réponses qui les aident à se prononcer sur le caractère utile de l'information renvoyée. Dans l'environnement de la bibliographie nationale et des catalogues de bibliothèque, ces attentes soulignent la nécessité d'enrichissements bibliographiques, lesquels ne sont pas fournis en temps normal. Toutefois, en raison de l'automatisation et d'autres évolutions qui lui sont liées, un ensemble d'informations peut facilement composer et, à peu de frais, la gamme des caractéristiques habituelles de nos bases de données bibliographiques. Permettez-moi d'expliquer pourquoi il en est ainsi, en faisant référence à la Bibliothèque du Congrès où nous avons mis en place avec succès une série de projets destinés à établir un lien entre chercheurs, catalogues et ressources électroniques ; à augmenter les contenus même des notices bibliographiques (en insérant par exemple des tables des matières et des compte-rendus) et à relier les descriptions bibliographiques aux ressources électroniques apparentées (par exemple, un accès Internet aux monographies d'une collection).

En 1992, le directeur du catalogage de la bibliothèque du Congrès, a approuvé la formation d'une équipe conseil pour l'enrichissement bibliographique (BEAT), composée exclusivement de volontaires en provenance de différents services, afin de mener des recherches et prendre des initiatives destinées à une meilleure complétude des notices bibliographiques. Encouragée à travailler « *en dehors de la boîte* », cette équipe a été dispensée des contraintes propres aux politiques et aux pratiques relatives aux activités conventionnelles de catalogage. BEAT avait reçu des instructions afin de créer et d'utiliser des méthodes automatisées pour mener son action avec un minimum d'incidences sur le travail du personnel chargé du catalogage traditionnel².

BEAT s'est efforcé dès le début, à introduire des tables des matières (TOCs) dans les notices bibliographiques, car les études, les unes après les autres, avaient conduit à la conclusion qu'une information de ce type améliorerait considérablement la fonction interrogation de la base de données ainsi que, en différents points, la pratique des catalogues par les usagers. A titre d'exemple, des recherches menées dans les années 1980 et 1990 ont démontré que les TOCs aident les usagers à mieux évaluer la pertinence de certains titres par rapport à leurs besoins en information, que les mots des TOCs améliorent l'efficacité de la recherche et qu'ils offrent les moyens de surmonter les habituelles contraintes de la recherche par sujet. Une étude longitudinale menée sur onze ans a en effet établi qu'« à une recherche par sujet se substitue une recherche par mots-clef », un élément en rapport directe avec les TOCs, puisqu'elles sont intégralement composées de mots-clef³.

Etant donné que la bibliothèque ne pouvait pas se permettre de laisser le personnel saisir manuellement les données TOC, le défi pour BEAT consistait à trouver une solution de programmation pour réaliser cet objectif d'enrichissement bibliographique. BEAT a en effet apporté en réponse plusieurs outils et projets. Le premier fut la création d'une application baptisée : *saisie de texte et conversion électronique* (TCEC), à utiliser pour le catalogage des publications que la Bibliothèque du Congrès reçoit en mode électronique de la part des éditeurs participant au programme de catalogage intégré (CIP). Utilisant le logiciel TCEC, le catalogueur signale la TOC du manuscrit électronique. Ensuite, le programme manipule les données de la TOC, puis ajoute le résultat dans le champ 505 de la notice de catalogage qui, dans MARC 21, indique l'information liée à la table des matières. Les manipulations réalisées par TCEC comprennent le formatage des informations concernant la table des matières suivant les prescriptions AACR2 relatives au catalogage des données TOC- par exemple l'effacement des chapitres, des sections, des mots et des nombres indiquant des parties, élimination de la pagination, et ajout de la ponctuation ISBD. TCEC convertit également tous les mots en minuscules, à l'exception du premier mot de chaque titre de chapitre. Aussi, tout ce que doit faire le catalogueur consiste à souligner tous les noms corrects à mettre en majuscules. Mais surtout, le transfert de l'information du manuscrit à la notice se réalise en quelques secondes et les données sont saisies avec une précision égale à celle que l'on peut voir dans le manuscrit électronique. A ce jour, un total de plus de 40 000 notices TOC ont été créées en utilisant TCEC⁴. Ce nombre est susceptible de s'accroître si les éditeurs continuent d'augmenter leurs propositions CIP en format électronique. Pour cette seule année, 35 000 manuscrits électroniques ont été réceptionnés pour le catalogage. 35% à 45% des notices bibliographiques qui en découlent présentent les conditions requises pour l'intégration de données TOC.

En ce qui concerne ce projet ainsi que d'autres efforts TOC, on évoque trois critères essentiels pour répondre à l'objectif de sélection des publications à traiter :

1. les titres de chapitre doivent améliorer une recherche par mots-clef en langage naturel ;
2. les titres de chapitre doivent apporter une meilleure compréhension des contenus du document que celle exprimée par le titre et la mention de responsabilité ;
3. les données TOC ne doivent pas demander une mise en forme manuelle trop importante pour être traitées par machine.

Heureusement, tout le monde ou presque est en mesure d'appliquer ces critères en prenant rapidement sa décision. Trois autres projets TOC fonctionnent principalement à l'aide de logiciels. Le premier, la table des matières numérique (D-TOC), a été amorcé à la fin des années 1990. Cette entreprise repose sur des personnels, principalement des opérateurs de saisie et des étudiants, pour examiner les tables des matières des publications sélectionnées selon des critères préétablis, comparables à ceux évoqués plus haut. A ce stade, un logiciel prend le relais. Les tables des matières sont converties en texte en utilisant la reconnaissance par caractère optique (OCR), puis un encodage en HTML et sont placées sur l'un des serveurs Internet de la Bibliothèque du Congrès.

Au cours de cette conversion, l'infrastructure MARC des notices est également modifiée de façon automatique, afin d'introduire des liens vers les TOC, établissant de ce fait, une réciprocité des liens entre ces deux sources d'information. Dans le but d'améliorer l'interrogation, les mots des zones du titre et des mentions de responsabilité figurent en début d'affichage de la table des matières. L'étiquette de méta-données de mots-clef à l'intérieur de fichiers TOC HTML inclut les termes des accès par sujets figurant dans la notice bibliographique. Le TOC en mode Web propose un lien direct vers la notice bibliographique et, réciproquement, cette notice bibliographique propose un lien du champ 856 vers la notice en mode Web de la TOC. De plus, Yahoo ! , Google ainsi que d'autres moteurs de recherche indexent ces notices D-TOC. Les utilisateurs d'Internet ne découvrent pas seulement des notices HTML. On leur donne aussi l'occasion d'avoir accès à l'OPAC de la Bibliothèque du congrès, grâce aux liens figurant dans les notices. Fait intéressant, le 6 juin 2005, une recherche effectuée sur Yahoo ! portant sur l'expression : « Contrôle tables des matières pour la Bibliothèque du Congrès » a généré un ensemble de réponses totalisant 262 000 entrées qui, toutes, comportaient des liens vers des notices D-TOC.

Ce projet a vu, à ce jour, la mise en ligne sur le Web de 300 000 notices. Ce rythme avoisine actuellement les 300 notices par semaine. Depuis peu, le projet D-TOC a été étendu à certaines collections rétrospectives, afin d'assurer la couverture documentaire des publications en langue étrangère, avec, pour commencer, l'allemand. La production de D-TOC ne devrait pas tarder à progresser. En effet, la bibliothèque possède des agences à l'étranger (Le Caire, New Delhi, Islamabad, Rio, Nairobi, Djakarta). Des efforts ont déjà été entrepris pour qu'elles s'investissent dans la création de fichiers TOC qu'elles transmettront à la Bibliothèque du Congrès.

Le plus productif et le moins onéreux des projets TOC de BEAT est l'application ONIX-TOC lancée en 2000. Cette entreprise implique l'extraction de données TOC à partir de fichiers ONIX fournis par des éditeurs. C'est un programme « Visual-Basic » qui gère ce projet en scannant un ensemble de fichiers ONIX, de façon à créer des tables des matières numériques mises en ligne par la suite sur le web. Comme pour le projet D-TOC, chacune de ces notices ONIX-TOC donne à l'utilisateur le choix de consulter, pour de plus amples informations, les

notices bibliographiques sur le catalogue en ligne de la Bibliothèque du Congrès. De la même façon, les notices bibliographiques dans l'OPAC de la bibliothèque sont enrichies automatiquement par des liens établis en zone 856 vers les fichiers ONIX-TOC. Certaines de ces notices ONIX-TOC comportent des photos de jaquettes. Actuellement, 50 000 notices D-TOC ont été créées. Ces chiffres, ajoutés à ceux des notices D-TOC représentent la saisie de plus de 6 millions d'occurrences d'ici juillet 2005.

La restriction la plus marquante liée aux projets D-TOC et ONIX-TOC vient de leur impuissance à jouer un rôle dans la récupération des notices bibliographiques à partir de l'OPAC, étant donné que ces notices prennent la forme de données disponibles sur le web. Toutefois, un progrès important est intervenu au début de cette année, avec la mise au point d'une solution programmée qui a rendu possible l'analyse grammaticale et l'extraction automatique de données nécessaires pour établir le corps textuel des TOC existant en mode numérique sur le web. Cette information est ensuite modifiée pour devenir conforme aux formats ILS (international library standards) des bibliothèques, permettant ainsi aux TOCs d'être ajoutées aux notices de catalogues. Du fait que ces notices ne sont pas totalement conformes aux prescriptions des AACR2, on les fait précéder de la mention : « *Note de table des matières* », générée automatiquement. Cette technique sera appliquée, non seulement à l'avenir, mais aussi, rétrospectivement, des personnels du BEAT mettent au point des programmes pour intervenir sur les fichiers des 30 000 notices D-TOC créées auparavant. Un travail va être entrepris pour vérifier si ce même genre d'approche peut s'appliquer aux ONIX-TOC, auxquels on ne peut, pour l'instant, accéder que par des moteurs de recherche Internet et les liens hypertexte des notices bibliographiques de l'OPAC.

C'est avec l'intention de bien faire comprendre le rôle joué par l'automatisation dans l'enrichissement des descriptions bibliographiques que je suis entré dans les détails en ce qui concerne le projet d'enrichissement grâce à la table des matières. Aucun de ces efforts n'est dispendieux en coûts horaires de personnel, une fois les programmes conçus et mis en application. En termes de frais réels de fonctionnement, les chiffres restent étonnement bas. Le coût moyen global pour la création d'une notice ONIX-TOC est estimé aux alentours de 0.10 dollars (US). Sans trop m'enliser dans les détails, j'aimerais que vous me permettiez d'en mentionner quelques autres. Tout d'abord, nous avons une série de projets basés sur des données ONIX, disponibles gratuitement. Une application ONIX mise en service par la Bibliothèque du Congrès assure aux utilisateurs du Web et de l'OPAC, un accès aux résumés créés par les éditeurs, afin de donner une indication sur le contenu de leurs livres. L'équipe BEAT se livre à une première estimation de ces résumés avant de décider de les traiter. Par exemple, si les descriptions d'un éditeur particulier donnent l'impression d'être excessivement publicitaires, elles sont écartées. Une fois que la source d'un de ces résumés est acceptée, chaque notice enrichie d'un lien vers elle porte une mention de désistement stipulant que « *l'information provient de données électroniques fournies par l'éditeur* », de façon à ce que les usagers comprennent que la Bibliothèque du Congrès n'est pas à l'origine de ces résumés. A l'heure actuelle, 100 000 notices ont été enrichies dans le but de fournir un accès à ces descriptions.

Dans un autre projet ONIX, des fichiers d'extraits représentatifs (en règle générale, un ou plusieurs chapitres ou une préface) fournis par l'éditeur sont mis en ligne sur le Web et des liens sont ajoutés dans les notices bibliographiques correspondantes. Certains de ces échantillons sont disponibles en mode HTML, d'autres sous fichier PDF. Dans le cas où l'un d'entre eux comporte des illustrations, cela donne la possibilité à l'utilisateur de consulter les agrandissements. A ce jour, plus de 17 000 de ces échantillons représentatifs ont été traités.

Enfin, amorcée l'année dernière, l'information biographique concernant les collaborateurs aux publications retenues, alimentée par des éditeurs, a été mise en ligne sur Internet, avec des liens vers les notices bibliographiques correspondantes. Cela représente aujourd'hui 12600 enrichissements réalisés.

Dans un autre domaine du travail d'enrichissement bibliographique, le BEAT a pris des dispositions qui ont permis la fourniture d'accès à des compte-rendus et commentaires savants concernant des acquisitions spécifiques. Dans certains cas, comme avec des projets TOC, on crée un accès vers des données qui sont saisies puis mises en ligne sur Internet, en ajoutant des liens vers les notices bibliographiques concernées. Dans d'autres cas, les commentaires et les comptes-rendus viennent compléter véritablement la notice bibliographique. L'importance matérielle du texte constitue le critère essentiel pour décider quelle technique appliquer. Lorsque les comptes-rendus s'avèrent plus concis, un programme extrait et insère automatiquement cette information dans les notices bibliographiques correspondantes. En revanche, lorsque les notices sont volumineuses ou contiennent des notes de bas de page ou des bibliographies, elles sont éditées sur le Web, avec création de liens vers les notices bibliographiques concernées. Dans tous les cas, la bibliothèque a pris des dispositions pour obtenir la permission de dupliquer les compte-rendus utilisés sur l'un de ses serveurs afin d'en assurer la longévité.

On attend des sources relatives aux commentaires et aux comptes-rendus qu'elles répondent à des critères spécifiques. Pour être sélectionnés, les compte-rendus doivent être disponibles gratuitement, stables, savants et sous forme de fichiers compatibles du point de vue des programmes. En raison du nombre important d'entreprises dont la publication de comptes-rendus constitue une source de revenus, le projet compte-rendu de BEAT a connu un développement moins rapide que les autres projets d'enrichissement. Mais, à l'heure actuelle, sept collaborateurs ont été recrutés :

- *Outstanding reference sources* du Service de référence et des usagers de l'American Library Association ;
- *Sites sélectionnés par les moteurs de recherche de cette association*, Commission des sites internet. Section des meilleures références gratuites ;
- *Handbook of Latin American Studies* , publié par la Bibliothèque du Congrès ;
- *H-Net Reviews in the Humanities and Social Science* (Revue savante);
- *Library Resources and Technical Services*, publié par "The Association for Library Collections and Technical Services";
- *Education Review* (Université de l'Arizona) ;
- *Bryn Mawr Classical Reviews*.

Deux projets supplémentaires BEAT qui fonctionnent pour l'essentiel sur la base d'applications informatiques, démontrent comment la Bibliographie Nationale et d'autres agences peuvent améliorer l'accès, en établissant des liens directs entre des notices bibliographiques et des ressources Internet en texte intégral. Le premier projet est : *Accès Internet aux publications en série*. Cette initiative a pour objectif de développer le caractère utile de la littérature grise, précieuse dans les domaines de l'économie et des affaires. De ce projet, en mai 2005, a découlé la production de liens donnant accès à des milliers de titres particuliers compris dans quelque 305 collections, pour la plupart d'un niveau de recherche et développement. Auparavant, presque 40 % de ces collections ne figuraient pas dans l'OPAC de la Bibliothèque du Congrès. Dans le but d'assurer la création de notices bibliographiques pour les monographies de ces collections éditées en mode électronique, on a établi un programme baptisé : *Assistant de catalogage sur le web*. Il crée des notices MARC en manipulant des données extraites des titres de monographies, au sein de collections sélectionnés pour être analysées. Muni de cet outil, un technicien consulte la page de résumé de la monographie et colle l'URL de cette page dans le programme. *L'assistant de catalogage sur le Web* trouve alors la page Web et, à partir de cette page, saisit les éléments du texte concernant le titre, les auteurs, les numéros de collection, les résumés et les mots-clef, s'il y en a. Ce programme rajoute aussi des informations essentielles sous la forme de notes normalisées. *L'assistant de catalogage sur le Web* introduit alors ces notices dans la base de données de la bibliothèque. Ce programme avertit aussi le personnel par le biais d'alertes postées en mode automatique sur la messagerie, lorsque de nouveaux contenus sont venus compléter le site Web que des entreprises ou autres ont créé pour leurs publications. Près de 3000 notices ont été créées de juin 2004 à juillet 2005, en utilisant *l'assistant de catalogage sur le web*. Cet outil a non seulement permis à la bibliothèque, à un moment on ne peut plus opportun, de fournir un accès à cette littérature grise, mais aussi de le faire en-dessous du coût normal prévu, même pour un catalogage de niveau minimal.

Le dernier projet de BEAT que je vais décrire, s'appelle *Accès Internet aux œuvres tombées dans le domaine public*. L'objectif consiste à récupérer le texte intégral, en mode électronique, de monographies imprimées figurant dans les collections de la Bibliothèque du Congrès. Ce projet est le fruit de plusieurs partenariats : des répertoires fiables fournissent à la bibliothèque du congrès une liste automatisée d'URL et de numéros de contrôle de la Bibliothèque du Congrès relatifs à des monographies numérisées pour la consultation publique. Un logiciel à la bibliothèque du congrès reçoit cette information, identifie la notice bibliographique correspondant à la version imprimée présente dans les collections de la bibliothèque, y ajoute l'URL de la version électronique. Il en résulte un accès immédiat à la version numérique pour le lecteur qui n'a plus besoin de se procurer la version imprimée. La bibliothèque partenaire est signalée dans une note qui l'identifie et, de ce fait, rend publique la collection dans laquelle figure la version numérique.

Le projet d'enrichissement bibliographique que je viens de décrire démontre la façon dont une bibliothèque nationale, en utilisant ses prestations anciennes à l'ère du numérique, capitalise sur celles-ci pour proposer des services nouveaux et perfectionnés. Ces projets sont d'organisation facile et peu coûteux à mettre en œuvre. Ils réclament juste un petit nombre de personnes motivées, principalement pour développer les programmes informatiques. Une fois en place, ces programmes fonctionnent surtout en autonomie, à l'arrière-plan. De nombreuses agences bibliographiques nationales détiennent l'expertise requise pour reproduire ces fonctionnalités et, par ce type d'action, elles sont à même d'ouvrir les portes de leurs bases de données, de leurs OPACs et de leurs collections. Ces activités, telles que je les ai décrites, apportent une réponse à des usagers de plus en plus habitués aux résultats immédiats de la

recherche par mots-clef et réclamant autre chose qu'une information bibliographique sommaire. En proposant ces enrichissements, les bibliothèques nationales répondent aux chercheurs dans leurs besoins accrus d'information, offrent des itinéraires structurés en fonction de leurs propres ressources documentaires et, ce faisant, incitent les mécènes à poursuivre cette mise en valeur de leurs services. Les enrichissements permettront également aux OPACs de mieux fonctionner en tant que mécanismes proposant des accès performants.

Cette réflexion nous oriente vers le deuxième sujet que je me propose d'aborder : la nécessité d'avoir une nouvelle génération d'OPACs avec des fonctionnalités considérablement développées, dont l'essentiel s'inspire des propriétés classiques particulières aux moteurs de recherche sur le Web et aux librairies en ligne. Pour commencer sur un point de vue positif, permettez-moi de reconnaître que les OPACs actuels accomplissent dans l'ensemble une longue liste de tâches de façon tout à fait performante. En effet, l'une des plus grandes réalisations du 20ème siècle, en matière de technologie de l'information, aura été le remplacement des catalogues de bibliothèque sur fiches par des catalogues d'accès public en ligne, pièces maîtresses des actuels systèmes intégrés de bibliothèque (ILS). Le nouveau code de catalogage international de l'IFLA précise cinq fonctions fondamentales pour cet outil : trouver, identifier, sélectionner, obtenir et naviguer. Et, nombreux sont ceux qui tomberaient d'accord pour reconnaître que la génération actuelle d'OPACs est compétente dans la réalisation de ces tâches.

Mais, de plus en plus, dans les écrits professionnels et ailleurs, s'exprime la préoccupation que nos catalogues en ligne ont fait leur temps et, plus précisément, qu'ils ne sont pas très adaptés à la récupération de ressources numériques en quantité sans cesse croissante. Comme la part des périodiques électroniques augmente un peu partout, dans les bibliothèques nationales ou dans les établissements de recherche, les OPACs trouvent facilement des informations bibliographiques concernant des titres de revues. Mais, en règle générale, ils ne permettent pas aux usagers de découvrir, à l'instant, les articles recensés. Les Agences bibliographiques nationales et les bibliothèques en général ne peuvent simplement pas se permettre de développer des descriptions bibliographiques en-deça de ce niveau. Aussi, en raison de contraintes économiques, les OPACs ne fourniront accès, au mieux, qu'à une sélection relativement restreinte de ressources électroniques, comprenant les bases de données de citations, des compilations de texte intégraux, des outils de référence en ligne et des sites internet. Parce qu'elles sont volumineuses et remplies d'éléments complexes, ces ressources se posent en défi à un contrôle bibliographique universel et moderne.

De moins en moins, les usagers dépendent des OPACs pour trouver les ressources qui les intéressent.⁵ En fait, une grande majorité s'est tournée vers des moteurs de recherche rapides et souvent précis, tel que Google, conçus pour répondre à la prolifération des contenus internet. Il apparaît incontestable, en effet, que cette grande majorité est satisfaite par de tels outils, même si elle se trouve confrontée à des ensembles de résultats abondants et imprécis⁶. Dans les années 2000, toutefois, de nouveaux systèmes de bibliothèque ont commencé à faire leur apparition, pour répondre à des demandes d'information de nature plus granulaire, certains de conception maison, d'autres, venus de l'offre commerciale. Ces outils intègrent des moteurs de recherche fédérés appelés parfois « *portails* », des séparateurs d'URL ouvertes, des systèmes de gestion des ressources électroniques (ERMS). Tous visent à mettre l'utilisateur en relation avec les ressources électroniques, comprenant celles qui sont disponibles dans le cadre des collections d'une bibliothèque, mais inaccessibles par l'intermédiaire de l'OPAC, et celles qui sont accessibles à distance sur le web⁷.

Dans l'ensemble, ces nouveaux outils s'efforcent de proposer des fonctionnalités qui permettent une recherche unique dans un grand nombre de bases de données de grande qualité, et de ressources en texte intégral, notamment la recherche dans l'Internet profond, orientant l'utilisateur vers de meilleurs contenus et fournissant des réponses détaillées par rapport à un thème ou une tâche – avec la facilité d'utiliser un moteur comme Google. Bien que, dans leurs premières versions, ces outils n'aient pas atteint leurs objectifs, toutes les grandes bibliothèques, ou presque, en ont fait l'acquisition, car ils offrent aux usagers un accès à l'information en tant que telle, et non pas à la description bibliographique de contenus similaires par leur intérêt. En effet, les sépareurs d'URL ouvertes, les applications de métarecherche fédérées, et les ERMS se sont avérés si utiles et la demande en est si forte, que la concurrence a incité les vendeurs à se dépenser en efforts généreux, pour fournir les évolutions les plus recherchées.

Malgré leur apparition et leur perfectionnement, ces outils indispensables se situent, pour la plupart, en dehors du champ des OPACs. Ainsi, dans son préambule à : *The Future of Integrated Library System : an LJ round Table* » Brian Kenneth remarque :

« les possibilités pour améliorer l'interopérabilité incitent les bibliothécaires à se montrer plus critiques à l'égard des ILS. Certains se demandent si le portail d'information du futur doit s'appuyer sur un ILS unique, ou, au contraire, représenter un ensemble de produits en provenance de différents prestataires d'ILS. Les vendeurs eux-mêmes sont à l'origine de cette manière peu cohérente d'aborder l'interopérabilité dans le marché des bibliothèques, avec des produits tels que *SFX* de Ex Libris, un instrument de liaison pour les références et *Endeavor's ENCompass*, un produit conçu pour la création et la gestion de contenus numériques. L'année dernière, ces systèmes autonomes pour la gestion des liens et la gestion numérique représentaient presque 13 % du marché des ILS⁸.

Cette remarque reste aussi vraie aujourd'hui qu'elle l'était au moment où elle fut écrite en 2003. La Bibliothèque du Congrès examine avec intérêt les évolutions d'une stratégie pour parvenir à un alignement des fonctionnalités des OPACs sur celles de ces nouveaux outils. En conséquence, elle finance un journal de recherche destiné à étudier les changements de nature du catalogue en ligne, où on peut trouver une réflexion sur une structure pour l'intégration de celui-ci aux autres outils de découvertes. Cette entreprise va fournir l'occasion d'étudier en détail une prévision de Sarah Thomas qui, dès l'an 2000, pressentait une « réinterprétation » possible du catalogue selon laquelle celui-ci allait devenir un service d'information qui recense, dans une organisation systématique, les publications et documents suscitant l'intérêt d'une communauté donnée, sans tenir compte de la forme sous laquelle ils paraissent. » Elle pronostique l'apparition imminente d'un « hybride » qui adoptera quelques unes des meilleures caractéristiques du catalogue, mais qui utilisera une infrastructure technologique de plus en plus sophistiquée, dans le but d'augmenter la production pour ceux qui sont à la recherche d'information⁹. Le projet de recherche de la Bibliothèque du congrès, qui va être entrepris en concertation avec de nombreux experts, a pour intention d'identifier ce qui pourrait être aujourd'hui réalisé pour se rapprocher d'un tel objectif. Ces recommandations, qui seront largement partagées, déboucheront ensuite sur des mesures pratiques.

Au-delà du travail qui mènera à l'intégration de ces fonctionnalités, la prochaine génération d'OPACs devra proposer beaucoup d'autres qualités, si les services des agences bibliographiques nationales veulent concurrencer avec succès les plus avancés des logiciels de découverte de ressources, représentés, sur le plan commercial, par les moteurs de recherche du Web et les librairies en ligne. Dans leur étude concernant « l'impact des moteurs de recherche Internet sur la recherche par sujet dans les OPACs¹⁰ », Yu et Young ont trouvé que

« la pratique des moteurs de recherche est suffisamment omniprésente pour avoir des répercussions sur les attentes de ceux qui utilisent un OPAC doté d'une infrastructure internet. » Leur recherche les a incités à prendre la défense d'un ensemble de caractéristiques pour la prochaine génération d'interfaces d'OPACs conçues sur les bases du web. Les fonctionnalités spécifiques recommandées par Yu et Young sont les suivantes :

Menu par séquence

Les études ont indiqué que les usagers choisissent les options de recherche en haut des listes. Il en résulte qu'on doit attentivement réfléchir aux accès placés en hauteur. Thorne et Whitlach ont trouvé que les menus devraient fortement encourager les usagers débutants à commencer avec des recherches par mots-clef. A l'université de Californie, Los Angeles, l'analyse de saisie transactionnelle laissait entendre que le succès des usagers était meilleur lorsque l'interface de recherche était repensée pour proposer la séquence suivante : mot-clef / expression , titre, auteur, sujet (spécifiant recherche par mots-sujet de la Bibliothèque du Congrès) et cote.

Caractéristiques de balayage

Hancock-Beaulieu a trouvé qu'entre 30 et 45 % des recherches en ligne étaient suivies par un butinage des rayons de la bibliothèque. D'après Yu et Young « une option balayage peut aider les usagers à trouver avec efficacité les documents désirés, en regroupant les documents apparentés par des mots du thésaurus.

Affichage des résultats

L'impact des recherches Internet sur les usagers les a incités à attendre des représentations graphiques, comme les icônes, pour l'affichage de l'information. Les affichages des OPACs devraient être l'objet d'une conception nouvelle où l'introduction de graphiques faciliterait la lecture de l'écran et la compréhension des étiquettes

Classements par pertinence

Les utilisateurs de moteurs de recherche Internet dépendent d'eux pour donner la priorité à certaines séries de résultats. Souvent, la majorité des utilisateurs ne va pas au-delà de quelques pages dans le déroulement des réponses. Yu et Young préconisent que les algorithmes de classement par pertinence incluent la date de publication, des vedettes-matière et des mots, ajoutant que « la popularité parmi les usagers et la fréquence des termes dans les tables des matières » pourraient également figurer dans les facteurs utilisés.

Fourniture d'indicateurs d'aide

Les OPACs devraient être autorisés à délivrer de l'aide en dehors des requêtes des usagers et proposer des outils et des concepts pour générer de meilleures stratégies de recherche. De plus, les systèmes devraient produire des conseils de recherche à la demande et, dans le cas d'une recherche n'obtenant aucun résultat, suggérer un autre choix d'orthographe ou de formulation.

Yu et Young plaident en faveur d'autres qualités fréquemment proposées par les moteurs de recherche Internet et les librairies en ligne et qui devraient faire l'objet d'une incorporation dans la prochaine génération d'OPACs, en raison de l'enrichissement qu'ils apportent à la pratique de ceux qui recherchent de l'information. De plus en plus, les écrits professionnels signalent les textes de nombreux autres auteurs en accord avec ces conseils.

La faculté de corriger l'orthographe

Comme les fonctionnalités de Google : « Est-ce que vous vouliez dire ... » Yu et Young suggèrent que les orthographes fautives, enregistrées lors des saisies pour les transactions dans le catalogue, pourraient être exploitées pour la constitution d'une base permettant la fourniture de ce service.

Le repérage de la popularité auprès des usagers

Tel qu'il existe sur Amazon.com. Les utilisateurs du catalogue se verraient donner la possibilité de commenter les choix dans le but de proposer un système de recommandations. Leurs commentaires ainsi que la diffusion des avis correspondants pourraient faire office de base pour alerter les usagers de la disponibilité de fonds comparables

Recherche en langage naturel

Telle qu'elle figure sur Google, avec une présentation dans une case unique dédiée à la recherche. Les utilisateurs du catalogue devraient pouvoir entrer des titres avec les articles de début, des noms d'auteur dans n'importe quel ordre, de la même façon qu'ils peuvent le faire en utilisant des moteurs de recherche.

La Bibliothèque du Congrès a récemment commandé auprès de Marcia Bates une étude qui a pour objet d'examiner les moyens d'enrichir les notices de métadonnées, en s'attachant à la fourniture de sujets complémentaires et d'autres mécanismes d'accès (par exemple, les thésaurus en accès frontal) et d'augmenter la granularité de l'accès et de l'affichage. Dans son rapport, Bates présente un ensemble de recommandations en vue d'accroître l'accès des usagers aux catalogues de bibliothèques et aux portails d'information¹¹. Dans son interprétation même, l'étude de Bates défend le principe qu'en ce qui concerne la recherche, la reconnaissance est beaucoup plus fiable que la mémoire, et qu'un thésaurus frontal donnerait aux usagers la possibilité d'entrer différents termes, au moment d'aborder un sujet sur lequel ils font des recherches, et leur permettrait de parvenir aux résultats escomptés. Elle a par conséquent élaboré une approche conceptuelle, dans le but de présenter aux usagers un vocabulaire par grappes, en relation avec les mots-clef entrés dans un système. Cette approche envisage la création d'un procédé de recherche par thésaurus, conçu comme interface frontale d'une base de données bibliographiques existante. Ses propositions détaillées devraient susciter un véritable intérêt auprès des vendeurs qui souhaitent perfectionner la prochaine génération d'OPACs. Dans le même temps, la Bibliothèque du Congrès examine les possibilités qu'elle aurait de mettre en œuvre, pour ses propres fins, certains aspects des conceptions de Bates.

Il est rassurant de constater qu'il s'agit de décisions déjà en cours qui démontrent leur capacité de réaction face aux nécessités d'apporter des améliorations aux pratiques des usagers devant l'OPAC. Aux Etats-Unis, par exemple, les principaux services bibliographiques publics prennent des initiatives pour enrichir leurs catalogues en ligne. Les efforts de RLG, avec son projet *Redlightgreen*, pour alimenter son catalogue, sont impressionnants. Cette action, menée avec le soutien de la Fondation Mellon, a pour but « d'offrir des références documentaires, riches, fiables, qui soient uniques en configuration Internet, et de proposer cette information par des procédés qui correspondent aux attentes des utilisateurs avertis du web ». *Redlightgreen* a été élaboré pour répondre aux besoins des étudiants de premier cycle. D'après Richard Parker de l'université de Warwick (Royaume-Uni) : « [Avec *Redlightgreen*], il est facile d'obtenir des résultats, il est facile d'affiner ces résultats et avec ces résultats, on peut faire beaucoup de choses. » Voici les exemples qu'il mentionne :

- La recherche est rendue claire en utilisant des mots clefs ou des expressions ;
- L'utilisateur peut peaufiner les résultats en utilisant les vedettes-matière de la bibliothèque du congrès ;
- Les résultats sont classés selon deux niveaux : la pertinence par rapport au(x) terme(s) de recherche et le nombre de bibliothèques offrant des résultats. Les ressources situées en haut de liste sont possédées par un plus grand nombre de bibliothèques que celles placées en bas. Il en ressort que plus une ressource se situe vers le haut, plus elle est susceptible d'avoir une crédibilité universitaire (et un étudiant de la bibliothèque aura plus de chance d'entrer en sa possession) ;
- Toutes les éditions d'un titre sont combinées dans un résultat unique. De ce fait, un texte de littérature populaire est facile à distinguer des oeuvres secondaires le concernant ;
- Les usagers peuvent vérifier si leur bibliothèque possède une ressource particulière qu'ils recherchent ;
- Les résultats sont liés à d'autres ressources en ligne qu'on peut y associer, telles que des articles, des compte-rendus, et à Amazon.com ou d'autres librairies. Les utilisateurs de *Redlightgreen* ont la possibilité d'éditer des bibliographies en ligne, de les mettre en forme dans l'un des quatre styles les plus répandus, puis de les télécharger ou de les envoyer par courriel¹².

De manière significative, *Redlightgreen* met en application la structure de l'IFLA pour les critères de fonctionnalités des notices bibliographiques (FRBR), dans la mesure où ce projet regroupe des œuvres pour lesquelles il existe un grand nombre d'éditions, dans un ensemble gérable d'occurrences qui correspondent aux termes de recherche d'un utilisateur. Des informations concernant les efforts en cours pour ajouter des fonctionnalités à *Redlightgreen*, ainsi que des commentaires sur son utilité sont disponibles sur le site internet de RLG.¹³

De la même façon, OCLC augmente ses efforts pour faire progresser le champ et l'utilité de son catalogue collectif en ligne, grâce à des initiatives comme *Open World Cat Pilot* fait pour « incorporer des notices de bibliothèques dans les sites de recherche Internet et tester l'efficacité du Web à guider les usagers vers des contenus possédés par les bibliothèques. » *Fiction Finder*, une autre initiative d'OCLC, est un système prototype conçu pour plus de 2,5 millions de notices bibliographiques concernant la fiction, qui fait l'objet d'un regroupement au niveau du travail, de telle façon que soient affichés un résumé, le genre, des informations sur le décor et le sujet. *Fiction Finder* agit également en tant que prototype dans le cadre d'une mise en œuvre plus vaste des concepts FRBR sur lesquels il se base. Par l'intermédiaire d'un projet baptisé par OCLC : *Curiouser*, cette application sera élargie aux notices d' *Open World Cat* hébergées sur Yahoo ! Par ailleurs, à la fin de cette année, OCLC prévoit d'afficher des notices par grappes, suivant les principes FRBR de *First Search*.

Au début de cet été, OCLC a annoncé que le site Internet *Open World Cat* contenait désormais des informations sur les outils susceptibles d'apporter une aide beaucoup plus pratique à la recherche de ressources documentaires, à partir du bureau d'un ordinateur personnel. Parmi eux figurent : 1) la barre d'outil de Yahoo ! qui fournit un accès permanent

aux notices d'*Open World Cat* via *Yahoo ! Search*. 2) la barre d'outil de Google dotée de la spécialité *Autolink* qui permet, elle aussi, un accès permanent aux notices d'*Open world Cat* et 3) les extensions de *Firefox Search*, à partir desquelles on peut entrer des termes de recherche et sélectionner le moteur de recherche à interroger.¹⁴

Des projets sont en cours pour la création d'un *World Cat wiki*. Selon l'un des concepteurs : l'idée, c'est d'avoir un wiki en complément de *World Cat*. Les gens peuvent ajouter des comptes-rendus, des couvertures, des commentaires, etc. et les relier aux notices bibliographiques. Nous espérons que ce système soit assez souple pour que les gens puissent réaliser des choses que nous ne prévoyons pas. Nous aimerions que le wiki soit disponible partout où se trouvent des notices *World Cat*.¹⁵

Ces exemples devraient suffire à encourager les agences bibliographiques nationales à lancer l'amélioration des processus qui rendent la navigation sur leurs bases de données en ligne plus réceptives aux usagers qui, dans leur pratique de la recherche documentaire, ont été influencés par leur expérience d'internet. Dans certains cas, les agences bibliographiques nationales comme RLG ou OCLC estimeront valable de s'investir elles-mêmes dans la fourniture de fonctionnalités supplémentaires qui ont été recommandées. Dans d'autres cas, les NBA pourront exercer leur influence, individuellement ou en association, pour réclamer de meilleures fonctionnalités aux distributeurs qui vendent et assurent la maintenance des OPACs qu'elles utilisent. Comme Roy Tenant l'a déclaré sans ménagement dans les colonnes d'un numéro récent du *Library Journal* : « Nous devons consacrer plus d'énergie à des transformations radicales et systématiques et non à des opérations cosmétiques. Si votre système propose une recherche plus difficile et moins efficace que sur Amazon.com (et lequel ne le propose pas ?), eh bien, vous avez du pain sur la planche. Arrêtez de demander aux vendeurs de minimales adaptations. Au fond, vous pouvez badigeonner un cochon de rouge à lèvres, il en restera toujours cochon pour autant.¹⁶

Véritables fruits d'un capital, nos bibliographies nationales et nos catalogues de bibliothèques traduisent des siècles d'efforts intellectuels. Vu les investissements majeurs consentis pour créer et maintenir de telles interfaces avec leurs collections, partout, les bibliothèques devraient rechercher les opportunités susceptibles d'offrir les caractéristiques indispensables, afin que les usagers soient incités à maintenir leur utilisation des OPACs, en tant que mécanismes fournisseurs d'accès performants. Du fait de leur position influente, les agences bibliographiques nationales sont les mieux placées pour améliorer et élaborer, à l'âge du numérique, de nouveaux moyens capables de capitaliser, grâce aux activités d'enrichissement bibliographique, sur les prestations traditionnelles et de préconiser une révision des OPACs, d'une absolue nécessité, pour satisfaire les usagers du 21ème siècle. Le moment est venu de proposer une action immédiate, de crainte que les utilisateurs ne laissent tomber, non seulement les ressources que nous avons rassemblées à leur intention, mais aussi les outils qui permettent leur accessibilité

Notes de fin de chapitre

1. Pour de plus amples informations concernant ces activités de l'IFLA, visiter la page du site de la section catalogage : <http://www.ifla.org/VII/s13/index.htm> [June 2005].

2. Pour plus d'informations sur l'équipe de l'enrichissement bibliographique et l'ensemble des projets décrits dans ce texte, voir la page du site BEAT : <http://www.loc.gov/catdir/beat/> [June 2005].

3. Pour les études et rapports qui apportent des éléments de justification sur les données TOC, voir : Pappas, Evan and Herendeen, Ann, "Enhancing Bibliographic Records with Tables of Contents Derived from OCR Technologies at the American Museum of Natural History Library," *Cataloging and Classification Quarterly*, 23:4 (2000), pp. 65-67; Winkle, R. Conrad, "An Analysis of Tables of Contents in Recent English-Language Books," *Library Resources and Technical Services*, 43:1 (1998), p.14.; Morris, Ruth C., "Online Tables of Contents for Books: Effect on Usage," *Bulletin of the Medical Library Association*, 89:1 (Jan. 2001), p. 29. Visiter aussi la page internet de RichCat : <http://www.loc.gov/standards/catenrich/> [Juin 2005].
4. Dans un second projet E-CIP TOC, utilisant des programmes mis au point par trois personnes de l'équipe BEAT, BEAT a créé une base TOC sur internet pour pratiquement 100 % des notices E-CIP contenant des données TOC. Ces données sont créées par programme avec un lien hypertexte pour chacune des notices concernées dans le catalogue de la Bibliothèque du congrès. Ces programmes, récemment modifiés traitent actuellement presque tous les signes diacritiques et enrichissent également l'affichage internet des TOC avec des vedettes-matières de la Bibliothèque du congrès qui appliquées par le personnel chargé du catalogage. A partir de mai 2005, environ 54000 notices électroniques CIPs (E-CIP) TOC ont été ajoutées sur le serveur Web.
5. Breeding, Marshall, "The many facets of managing electronic resources", *Computers in Libraries*, v. 24, no.1 (Jan. 2004). Disponible sur : <http://www.infotoday.com/cilmag/jan04/breeding.shtml> .
6. Achenbach, Joel, "Search for Tomorrow: We Wanted Answers, and Google Really Clicked, What's Next?" *Washington Post* Feb.15, 2004, D7.
7. Une vue complète de ces outils est disponible sur le site du Groupe des questions d'applications, portail de la bibliothèque du congrès : <http://www.loc.gov/catdir/lcpaig/paig.html> (Juin 2005)
8. Kenny, Brian, "The Future of Integrated Library Systems: An LJ Round Table," *Library Journal* (June 15, 2003), 37.
9. Thomas, Sarah E., "The Catalog as Portal to the Internet", *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web*, Washington, D. C.: Cataloging Distribution Service, 2001, Disponible à : http://lcweb.loc.gov/catdir/bibcontrol/thomas_paper.html [Juin 2005].
10. Yu, Holly and Young, Margo, "The Impact of Web Search Engines on Subject Searching in OPAC," *Information Technology and Libraries*, (Dec. 2004), pp. 168-180.
11. Bates, Marcia, "Task Force Recommendation 2.3 Research and Design Review...Final Report (Version 3), June 1, 2003" Voir : <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf> [2005].
12. Parker, Richard, "Promoting RedLightGreen at the University of Warwick", *RLG Focus*: (issue 71, Dec 2004). Voir : http://www.rlg.org/en/page.php?Page_ID=20480#article0 [Juin 2005].
13. Voir : http://www.rlg.org/en/page.php?Page_ID=13061 {[Juin 2005]

et : http://www.rlg.org/en/page.php?Page_ID=20500 [Juin 2005].

14. *OCLC Abstracts*, 8:23 (June 6, 2005). Voir :
<http://www5.oclc.org/downloads/design/abstracts/06062005/index.htm> [Juin 2005].

15. Thom Hickey, “Library Metadata Techniques and Trends”, posting to Outgoing (Blog), Mai 2005. Voir : http://outgoing.typepad.com/outgoing/2005/05/worldcat_wiki.html [Juin 2005].

16. Tennant, Roy, “Lipstick on a Pig,” *Library Journal*, April 15, 2005, p. 34.
