Arguments bibliothéconomiques

Du catalogue enrichi au métacatalogue

Thierry Samain

SCD de l'université de Limoges
Section Sciences
thierry.samain@unilim.fr

La plupart des services électroniques traditionnels proposés par les bibliothèques, ainsi que l'offre en développement constant de ressources en ligne sont de plus en plus rassemblés en la porte d'accès qu'offre le site web. La difficulté à proposer de manière cohérente cet ensemble hétérogène impose une réflexion sur la manière dont on envisage à terme de fédérer ces accès.

L'approche la plus naturelle pour un bibliothécaire paraît offerte par l'enrichissement à la source du catalogue qui, dans son interface web,autorise un passage naturel et souple des références au contenu signalé. Si cette solution apporte de manière transitoire une aide concrète et simple à mettre en œuvre rapidement, il n'en reste pas moins nécessaire d'imaginer dès aujourd'hui des mécanismes d'intégration d'accès différenciés à un niveau supérieur. Le « portail » documentaire, métamoteur offrant des passerelles vers des bases structurées selon leur cohérence propre tout en respectant des normes et standards établis,semble se présenter comme l'un des modèles les plus satisfaisants.

Le point de départ

Tout débute par une mise à disposition de services mettant en jeu un accès à de la documentation électronique. Le développement de services proposés en ligne et notamment à partir du site web de la bibliothèque fournit le matériau propre à l'enrichissement du catalogue. Ce substrat peut se composer, en prenant exemple sur ce qui est disponible au

Thierry Samain *est responsable de la section Sciences du SCD de l'université de Limoges, après avoir été membre du Service des nouvelles technologies du service commun de la documentation de Lyon 3. Outre son mémoire de fin d'étude à l'Enssib sur* L'accès aux catalogues des bibliothèques par Internet, *il a publié un article concernant les périodiques sur le web.*

service commun de la documentation (SCD) de l'université Lyon 3 :

– de ressources numérisées par la bibliothèque (tables des matières, documents spécifiques à une classe d'utilisateurs seule habilitée à y accéder, documents internes…) ;

– de périodiques (liste « enrichie » permettant de proposer, outre les données locales propres à l'établissement, un accès à différents éléments gratuits ou payants disponibles sur les sites web d'éditeurs ou d'agrégateurs)[1] ;

– de sélections de sites web généralement sous forme de liste ; la recherche sur ces éléments s'effectue par le biais d'un moteur de recherche plein texte propre au serveur web sur lequel les pages sont hébergées ;

– de bases de données en ligne gratuites ou payantes (accessibles par le biais d'agrégateurs, ou indépendantes) ;

– de bases de données sur cédéroms ; l'intégration de ce support aux autres ressources en ligne a été facilitée par le développement de clients web et de l'architecture client léger/serveur d'application, qui permet de centraliser la plus grande partie de la maintenance au niveau serveur.

Un premier degré d'intégration

Les différentes ressources en ligne présentées très brièvement ci-dessus sont souvent proposées avec plus ou moins de bonheur au travers des pages web de la bibliothèque. La première solution d'intégration de ces différents éléments est offerte par l'enrichissement du catalogue à sa source. Il s'agit pour cela d'utiliser les possibilités d'Unimarc[2] et des facultés de paramétrage de l'affichage détaillé des notices offertes par le client web du catalogue.

Par commodité, nous distinguerons par la suite deux notions : la ressource électronique qui recouvre une entité fermée se suffisant à elle-même (un article en ligne, la version électronique d'une revue papier), et le site web, multiforme et voué à des développements perpétuels.

Comment enrichir le catalogue à partir des différentes ressources évoquées précédemment ?

Les périodiques

L'enrichissement du catalogue par l'accès aux périodiques en ligne est tout particulièrement intéressant, car le plus facile à mettre en œuvre à une grande échelle. Nous évoquerons trois cas.

– Celui des *ressources électroniques à part entière* (cas de revues n'existant que sous forme électronique). On utilise le champ 856 (localisation et accès à une ressource électronique), avec un premier indicateur (mode d'accès) 4 (pour http) et une sous-zone $u (pour URL)[3].

Pour les différents exemples cités en note, des réalisations concrètes peuvent être observées sur le site web du SCD Lyon 3[4]. Les différents cas de recherche proposés permettent de visualiser des exemples avec lien vers des ressources d'accès libre ou sur souscription, par le biais de sites d'éditeurs ou d'agrégateurs.

– Celui des *ressources électroniques reproduisant une version papier* « originelle ». Si l'exemplaire n'est pas déjà disponible à la bibliothèque (et donc catalogué), le catalogage est réalisé avec le champ 856 ; si un exemplaire papier est déjà disponible dans la bibliothèque, on utilise le champ 452 (autre édition sur un autre support) avec la sous-zone $u[5].

Ce procédé soulève quelques questions : quelle est la frontière entre une ressource offrant une simple copie électronique du papier, et une ressource électronique qui dispose d'une originalité propre (par exemple le *BBF* en ligne qui propose une recherche sur l'ensemble des articles) ? À partir de quel moment doit-on créer une notice spécifique ?

Par ailleurs, qu'en est-il des publications plein texte avec retard d'un numéro (par exemple le *Courrier de l'Unesco)* ?

Afin de proposer l'amorce d'une réponse, on peut observer que le site de revue « enrichi » (contenu ou mise en page) est généralement le fait d'éditeurs qui ne publient qu'une revue (champ 856). Dans le cas des grands éditeurs, on dissocie nettement l'accès aux articles de chaque revue (champ 452) et l'interrogation d'une base globale portant sur l'ensemble des revues publiées.

– Celui des *ressources électroniques offrant une partie de la version papier* (sommaires ou résumés par exemple). On utilise le champ 464 (niveau de dépouillement) avec la sous-zone $u pour l'URL[6].

Le lien vers des parties électroniques de documents peut offrir un réel service ; il n'en reste pas moins important de suivre les évolutions futures, dans la mesure où aucune règle définie ne semble se dégager

1. Pour plus d'informations sur le service développé, se reporter à Thierry Samain, « Périodiques sur le Web : vers une intégration des ressources électroniques », *BBF,* 1999, n° 1, p.114-115.

2. *Unimarc Manual : Bibliographic Format,* München, Saur, 2000.

3. Voir : http://www.abes.fr/formatex.htm et, pour une consultation en ligne des éléments du champ 856, http://www.multimania.com/vacher/profess/cours/mediadix/catalre/856.htm

4. http://www-scd.univ-lyon3.fr/cgi-bin/frameset_cataloguerecherchemots.asp Effectuer une recherche avec titre = revue surfaces.

5. Effectuer la recherche avec titre = BO revue, ou titre = oxford journal of legal studies, ou titre = european journal of economic and social system.

6. Effectuer une recherche avec titre = revue droit société, ou titre = Beaux-arts magazine.

(en ce qui concerne le problème des tables des matières, *cf.* ci-après le cas des monographies). Les sommaires deviennent par ailleurs de plus en plus accessibles par le biais de bases de données globales (dépouillement, agrégateurs, fédérations d'accès), et vont nécessiter une intégration à un niveau supérieur (*cf.* dernière partie).

Les monographies

En ce qui concerne les monographies électroniques, le même type de catalogage peut être opéré à partir du champ 856[7].

Même s'il semble recommandé de pointer vers les ressources sur leur site de publication, il peut parfois s'avérer plus prudent de récupérer une copie de la ressource sur son propre serveur dans le cas de documents hébergés sans garantie de stabilité, en respectant, il va de soi, le droit d'auteur[8].

Les tables des matières peuvent enrichir le catalogue de différentes manières par :

– une numérisation en local (ou récupération sur sites web d'éditeurs), un stockage dans une base spécifique et un lien aux notices, dans la mesure où le client catalogue gère ce type d'accès[9] ;

– un repérage sur le web ou un stockage en local et un lien dans la notice par le biais du champ 464 ;

– une intégration directe à la notice par récupération auprès de fournisseurs de notices (comme c'est déjà le cas pour les résumés) : un service à venir ?

Les sélections de sites web

A priori, il est possible de cataloguer les sites web selon les mêmes règles que les ressources électroniques. Les avantages de l'enrichissement du catalogue à la source par le catalogage des sites web sont :

– une procédure unique de traitement des documents « traditionnels » et des sites web ;

– la disponibilité d'un vocabulaire contrôlé et structuré bien traité par les SIGB (systèmes intégrés de gestion de bibliothèque) ;

– l'utilisation d'un moteur de recherche unique au niveau du SIGB ;

– l'utilisation d'un format riche qui assure la conversion future vers d'autres types de format (Dublin Core).

Ce dernier avantage indique cependant qu'il est d'ores et déjà inévitable de réfléchir à une structuration autre pour une base de données propre aux signets. Le format Unimarc apparaît en effet souvent lourd et inadapté au signalement de ressources web.

Ceci dit, la difficulté à appliquer Unimarc aux ressources en ligne tient essentiellement au faible développement des normes de description et à leur manque d'illustration par des exemples variés. Concernant le format USMarc, la documentation la plus fournie est offerte dans le cadre du projet Corc d'OCLC[10] qui se donne pour objectif un catalogage partagé des ressources en ligne (il est à noter que les deux formats Marc et Dublin Core ont été retenus pour le signalement).

Par ailleurs, le PUC (Permanent UNIMARC Committee de l'IFLA) semble plutôt privilégier une extension du domaine de la zone 856 pour gérer les liens vers des URLs au détriment de la stricte logique du format UNIMARC qui prévaut à l'utilisation des zones 4XX (sous-zone $u)[11].

D'autres inconvénients apparaissent aptes à en déconseiller l'usage :

– *l'absence de règles éditoriales précises propres à la création des pages web* qu'un manque de normes détaillées de catalogage rend difficile à traiter. Ainsi des informations primaires telles que la mention d'édition ou le titre (faut-il employer le titre le plus visible sur la page d'accueil ou se référer à la balise <TITLE> du document HTML ?) sont souvent difficiles à repérer ;

– *la volatilité des liens* pose le problème de leur mise à jour ; si, dans les pays nord-américains, où le catalogage Marc des ressources électroniques est plus développé, les associations d'utilisateurs défendent auprès des développeurs de SIGB l'ajout de fonctionnalités permettant une veille automatisée des liens brisés ; aucun logiciel en France ne propose à notre connaissance cette fonctionnalité ;

– *la nécessité de se poser la question du catalogage dans une perspective nationale et internationale ;* en effet, dans le cadre de l'intégration des catalogues locaux à des bases collectives, il semble peu raisonnable que chaque bibliothèque catalogue de son côté les mêmes ressources. L'approche du signalement exhaustif des sites web ne peut se concevoir, il nous semble, que partagée et dans le cadre d'un projet national. En ce qui concerne les bibliothèques universitaires, le projet WebDoc développé par l'Agence bibliographique de l'enseignement supérieur (ABES)[12] a pour vocation principale de permettre, au sein du Système universitaire, le référencement par chaque bibliothèque de ses propres ressources électroniques et la gestion de l'accès éventuel à des abonnements en ligne contractés par les établissements ;

– *de nombreux organismes documentaires ont entamé un signalement des ressources électroniques de leur domaine* (bibliothèques publiques, universitaires, nationales, centres de documentation, éditeurs, libraires, etc.), mais ils ne sont pas les seuls à œuvrer pour le défrichage de la Toile… Tous ces organismes n'ont

7. Effectuer la recherche avec titre = état des lieux de la communication.
8. Effectuer la recherche avec titre = rapport bibliothèque nationale.
9. Effectuer une recherche avec titre = new zealand poetry.

10. http://www.oclc.org/oclc/corc/
11. Groupe de travail UNIMARC. http://www.abf.asso.fr/enrichi/unimarc/

12. http://www.abes.fr/docelec.htm

pas une culture du Marc, et il apparaît décisif d'utiliser un standard moins « confidentiel »,susceptible de favoriser la coopération entre différentes professions.

Les bases de données bibliographiques

Pour l'instant, peu de solutions sont proposées, mais il pourrait être envisagé, dans le cadre de l'enrichissement du catalogue à sa racine,le chargement de notices d'articles, avec un lien vers le texte intégral disponible en ligne. Cela nécessiterait,il est vrai, un travail lourd et régulier pour la mise à jour de la base locale, et apparaît, en conséquence, applicable essentiellement à de petites structures ne nécessitant l'accès au dépouillement que d'un faible nombre de revues.

L'avenir de l'intégration de ce type de données semble passer par une abolition de la frontière que crée localement l'hétérogénéité des supports (principalement cédéroms ou DVD et disques durs), ainsi que par une insertion de l'accès à un niveau supérieur, permettant ainsi de décentraliser la gestion d'une masse de données souvent imposante. Un niveau intermédiaire d'intégration est déjà proposé par certains fournisseurs de données par le biais d'une implémentation Z39.50. L'interrogation de bases d'articles peut ainsi être associée à celle des catalogues de bibliothèques supportant le même protocole.

Perspectives de développement

L'objectif devrait être l'intégration à un réel système d'information (portail constitutif d'un véritable intranet, voire Internet documentaire).

On le voit,l'intégration de l'accès aux ressources électroniques par le biais de l'enrichissement du catalogue propose une solution particulièrement intéressante dans la période de transition et d'évolution rapide que nous vivons, mais ne résout pas tous les problèmes.

Une autre approche pourrait être l'intégration de l'accès aux différentes ressources en ligne à un niveau supérieur : chaque type de données serait ainsi traité selon le modèle le plus approprié à sa spécificité structurelle, en respectant les standards internationaux, et la recherche unifiée sur l'ensemble de ces ressources s'opérerait par le biais d'un « portail » documentaire, véritable ante-serveur basé sur une passerelle principale pour l'interrogation de l'ensemble des données.

Actuellement,l'interrogation des ressources en ligne s'effectue souvent de manière séparée par :

- le catalogue de bibliothèque : moteur du SIGB ;
- les ressources locales : moteur du serveur web ;
- les ressources locales hors web ou ressources externes : moteur spécifique à chaque base.

Le moteur de recherche général du portail aurait vocation à fédérer les catalogues, les bases de documents numériques créées localement et les bases en ligne externes.

Le site web Enrichi *

La concertation sur l'information bibliographique enrichie est née de discussions informelles entre bibliothécaires et fournisseurs de données. Le constat était simple. Il existe un besoin d'élargissement de la notion classique d'information bibliographique pour élargir les catalogues en ligne d'informations affichables sur écran (qu'il s'agisse d'extrait ou d'images comme la couverture et la table des matières, ou du document lui-même s'il est électronique). Il existe une offre de fournisseurs de données bibliographiques ou de fournisseurs de documents pour proposer de telles données. Mais il manque, dans le contexte classique des formats d'échange bibliographique, les standards permettant de les prendre en compte, ou, s'ils existent, ceux-ci sont insuffisants.
Une première réunion organisée à la Rochelle à l'occasion du congrès de l'Association des bibliothécaires français, en mai 1999, a permis à des bibliothécaires, des fournisseurs de données et des fournisseurs de logiciels de gestion de bibliothèque de convenir qu'il y avait matière à concertation.
Dès la seconde séance tenue en octobre 1999, trois pistes de travail se sont dégagées :

1. Il est nécessaire de permettre à court terme, dans le cadre du format d'échange Unimarc et de la norme ISO 2709, la fourniture de notices enrichies ou liées à des fichiers de données numériques associées. Un groupe de travail Unimarc s'est d'abord constitué, qui a transmis ses missions au Comité français Unimarc constitué le 31 mars 2000. Ces initiatives pourraient déboucher sur des recommandations propres.

2. Il est indispensable de préparer dès maintenant la mutation vers des structures de formats d'échange et de stockage plus adaptés aux nouveaux besoins, aux nouvelles offres, et à l'état de l'art en matière de structuration de l'information. Dès 1999, le modèle XML a été reconnu comme de nature à répondre à moyen terme à ces défis. Un groupe de travail XML a donc été constitué. Dans un premier temps, il organise des séances d'information et de sensibilisation à diverses applications de XML, notamment des différentes DTD (définition de types de documents). Ces séances sont suivies par des bibliothécaires et documentalistes, des fournisseurs de données et des fournisseurs de systèmes de gestion de bibliothèque.

3. Le signalement des ressources en ligne se développe dans les bibliothèques, mais fait apparaître la nécessité d'une coopération sur le partage et la coordination du travail, et sur la question des outils et formats à utiliser. C'est l'objet du groupe de travail *Signalement partagé des ressources en ligne.*

La concertation pour une information bibliographique enrichie constitue un cadre informel de travail, d'information, de formation et de diffusion réunissant à la fois les professionnels de l'information au sens large et leurs fournisseurs (de données, de systèmes, de documents). Elle contribue à réunir ces divers partenaires dans une gestion commune de la phase de transition que nous sommes en train de vivre, et se caractérise par une mise en cause des cadres traditionnels de normalisation et de standardisation au profit d'outils et concepts communs à toute la chaîne de l'information numérique.
Créé dès 1999, le site web Enrichi, hébergé par l'ABF, permet d'annoncer les réunions, de publier leurs comptes rendus et de proposer des liens, notamment sur XML et les métadonnées. Une liste de diffusion privée permet la diffusion des informations et d'éventuels échanges d'idées et d'expérience.

Dominique Lahary

* http://www.abf.asso.fr/enrichi/

Les catalogues

Le volet du moteur permettant l'interrogation répartie des catalogues de bibliothèques reposerait sur la structure d'échange des données bibliographiques propre à la norme Z39.50[13].

Dans certains cas (constitution de portail dans le cadre d'une coopération étroite entre différents organismes), on pourrait tout à fait imaginer se passer de Z39.50 en opérant la conversion de l'ensemble des notices (ouvrages,articles,sites web,...) dans un format jouant le rôle de plus petit dénominateur commun (par exemple le Dublin Core) et en interrogeant une base d'index centralisée (toutes les données restant stockées de manière répartie).

Les Signets des bibliothèques

Coopération et intégration à un système global

Dans le cadre du groupe Enrichi-Signalement partagé des ressources électroniques[1], une réflexion a été initiée sur la constitution d'une base nationale signalant les ressources disponibles sur le web. Ce travail est envisagé en deux temps.

- *Coopération au niveau des listes statiques :* une solution « légère » d'évaluation de l'existant et de création d'une page par discipline renvoyant vers les sélections de chaque établissement pour ce domaine[2].
Le fait de regrouper le travail déjà effectué permettra de repérer les acteurs potentiels d'un signalement partagé des ressources par discipline et d'obtenir des propositions de maintien d'une page centralisant l'accès à ces différentes listes pour chacune d'entre elles (Centres d'acquisition et de diffusion de l'information scientifique et technique électroniques).

- *Mise en place d'une base de données :* une solution « lourde » de constitution d'une base nationale de signalement des ressources disponibles sur le web. Cette base serait alimentée par des participants identifiés pour chaque thématique. Le format des données serait celui préconisé par le groupe de travail XML (DTD à définir ou incluant les métadonnées Dublin Core). Le moteur de recherche sur les données pourrait être issu des travaux existant actuellement au niveau européen[3], et auxquels participe la Bibliothèque nationale de France, ou acquis auprès d'une société.

1. *Cf.* encadré page précédente.
2. *Cf.* le travail déjà réalisé pour les bibliothèques universitaires à : http://www-scd.univ-lyon3.fr/busignets/
3. Voir par exemple le logiciel ROADS développé dans le cadre du projet DESIRE (Development of a European Service for Information on Research and Education) à : http://www.desire.org/

Les bases de documents numériques (sommaires, signets…)

Il s'agit des bases créées localement et structurées selon des standards établis (XML pour le format de transmission, Dublin Core pour le format des métadonnées par exemple ; *cf.* encadré page précédente). Un exemple de base pouvant être constituée de cette manière est à l'étude dans le cadre d'un groupe de travail consacré au signalement partagé des ressources électroniques (*cf.* encadré ci-contre).

Comme indiqué précédemment, l'on pourrait par ailleurs imaginer faire subir ce type de traitement à nos bases bibliographiques (dans un but de consultation publique),afin d'éviter le développement d'une passerelle spécifique pour l'interrogation des catalogues dans le cadre du portail.

Les bases en ligne externes

L'intégration de l'accès à ces bases à un niveau supérieur nécessite le développement spécifique d'une passerelle pour chacune de ces bases en fonction du format des données, d'où l'importance de l'utilisation progressive par les éditeurs de formats communs et standardisés (notamment SGML, avec des DTD communes).

Les agrégateurs peuvent également jouer un rôle en offrant une porte d'accès unique (de plus en plus susceptible de s'intégrer à un portail documentaire dans la mesure où un accès Z39.50 est possible) à de grosses bases de données.

Il semble enfin primordial de réfléchir dès maintenant à la mise en place de structures nationales offrant de manière centralisée le stockage, pour une conservation raisonnée,et l'accès,afin de faciliter l'intégration de ce type de données.

En conclusion, le cheminement proposé a pour modeste objectif de dégager quelques pistes qui nous semblent incluses dans un processus d'intégration croissant des différents accès aux ressources électroniques, de l'organisation « rationnelle » des pages web en fonction des services que la bibliothèque se donne pour mission de rendre à la constitution d'un véritable « portail » documentaire, en passant par la centralisation qu'offre le catalogage à la source.

Si ces différentes pratiques sont pour un temps encore vouées à cohabiter, il nous semble important de réfléchir dès maintenant aux logiques dont elles procèdent et à leur place respective dans les systèmes d'information à venir.

Novembre 2000

13. *Cf. Z39.50 Maintenance Agency Page.* http://lcweb.loc.gov/z3950/agency/ et Dominique Lahary, *La norme Z39.50.* http://www.multimania.com/vacher/profess/cours/mediadix/z3950/