

Les bases de connaissance à la croisée des chemins

Qui veut identifier ses ressources électroniques et y donner accès utilise un outil aujourd'hui incontournable : la base de connaissance. À l'heure où elle devient le cœur des nouveaux systèmes de gestion des bibliothèques, les questions qui se posent quant à la façon de l'alimenter et à la qualité des données sont plus que jamais d'actualité.

Une base de connaissance est une base de données qui rassemble de façon structurée les informations connues sur les ressources électroniques : titres, périmètres des bouquets, états de collection, URL d'accès, et, intégrés plus récemment, termes des licences, coûts, statistiques d'usage...

Elles sont apparues au début des années 2000 avec les résolveurs de liens. Basés sur la norme OpenURL, développée au départ par l'université de Gand, et commercialisés pour la première fois par Ex Libris sous le nom de SFX, ces outils ont pour objectif de faciliter l'accès aux ressources électroniques, depuis une référence bibliographique vers le texte intégral en ligne.

Progressivement, les bases de connaissance sont devenues le socle de nombreux autres outils : moteurs de recherche fédérée, ERMS (*Electronic Resource Management Systems*), outils de découverte et maintenant systèmes de gestion de bibliothèque de nouvelle génération.

LA BASE DE CONNAISSANCE EST-ELLE UN CATALOGUE ?

Cette documentation électronique que l'on sait expansive et mouvante n'a jamais pu se fondre dans le circuit du traitement du document en place dans les bibliothèques. Mouvante, car éditeurs, diffuseurs, agrégateurs cohabitent dans le paysage de l'édition scientifique et échangent des titres au gré des accords commerciaux. Expansive, car les modèles économiques sont également très variés, du « *big deal* » au titre à titre, en passant par les accès couplés papier + électronique, et plus récemment le PDA (*Patron Driven Acquisitions*)¹, ou l'achat de forfaits pour les *ebooks*. La précision des modalités d'accès prévaut sur la description du document souvent réduite au minimum. Le couple titre/identifiant numérique (ISSN ou ISBN) constitue bien souvent le corps de la notice qui n'est pas une notice descriptive, mais une notice qui a essentiellement vocation d'identification. La base de connaissance a pour but de répondre à la question « à quoi peut accéder l'utilisateur et comment ? ».

Les fournisseurs de ressources électroniques réalignent fréquemment de nouvelles plateformes, pou-

vant justifier ainsi des hausses tarifaires, et la syntaxe des URL change. Les éditeurs imposent des embargos sur les publications, nécessitant une mise à jour régulière des états de collection. La bibliothèque n'a pas d'autre choix que de sous-traiter ce travail de maintenance des liens et de vérification du périmètre des bouquets, impossible à suivre manuellement.

Dans une enquête récente², les fournisseurs déclarent entre 8 et 29 équivalents temps plein (ETP) consacrés aux mises à jour de la base de connaissance. Dans les SCD, on est loin de ces chiffres. La réflexion reste à mener sur la redéfinition et la redistribution des tâches en interne dans les bibliothèques. Le catalogueur devient un « gestionnaire de métadonnées », terme que l'on peut juger trop politiquement correct, mais qui recouvre bien une nouvelle réalité : il ne gère plus des stocks, mais des flux, et ces tâches de veille et de mises à jour permanentes restent encore sous-estimées.

BASES DE MÉ-CONNAISSANCE ?

Il faut dire que les fournisseurs rivalisent d'arguments d'exhaustivité et de simplicité pour vendre leurs produits : chacun vante sa base de connaissance comme étant la plus complète et la plus ergonomique. Le bibliothécaire n'aura qu'à activer en un clic ses abonnements pour faire fonctionner son résolveur de liens ou générer un abécédaire des revues électroniques et booster l'usage de ressources chèrement acquises. Pourtant, ouvrir une base de connaissance peut donner le tournis, car exhaustivité ne rime pas toujours avec rationalité et précision.

Le tournis devient vertige quand on ouvre une deuxième base de connaissance et que l'on constate que les données qui devraient *a priori* être les mêmes, puisque transmises par les éditeurs, diffèrent souvent grandement d'une base à l'autre. Par exemple, d'un côté on ne signale que les revues, dotées d'un ISSN, de l'autre on recense aussi la littérature grise, rapports, actes de conférences, qui viennent grossir la liste des publications. Les problèmes peuvent venir aussi du format dans lequel les fournisseurs transmettent leurs données. À l'initiative de l'UKSG

[1] Modèle consistant à acquérir uniquement les ressources qui sont réellement empruntées par les utilisateurs.

[2] Breeding, Marshall, *Knowledge Base and Link Resolver Study : General Findings*, mai 2012, www.kb.se/dokument/Knowledgebase_linkresolver_study.pdf



(United Kingdom Serials Group) et de la NISO (National Information Standards Organization), un groupe de travail a produit des recommandations pour l'échange de données entre les différents partenaires³. Au final, on est loin d'une norme, mais ce rapport souligne l'importance de la structure des données. Il préconise une méthode finalement un peu artisanale – transfert par FTP de fichiers tabulés – mais qui a le mérite d'être facilement applicable par tous.

À cette confusion intrinsèque s'ajoute celle générée par la nécessité de personnaliser les données. Les éditeurs proposent des collections standard, mais les bibliothèques soumises à de fortes pressions budgétaires veulent sortir des « *big deals* » pour reprendre la main sur les contenus souscrits. De plus, les portefeuilles historiques sur lesquels sont encore basés de nombreux modèles économiques diffèrent d'un établissement à l'autre. Ils doivent donc personnaliser la liste de leurs publications dans leurs instances locales. Le même travail est répété pour les ressources absentes de la base de connaissance, que chaque bibliothèque crée aussi localement.

VERS UNE BASE NATIONALE ?

Il paraît légitime d'imaginer que ce travail puisse être mutualisé dans une base nationale partagée. « *Do once and share* », c'est le slogan de la base de connaissance britannique « Knowledge Base+ », opérationnelle depuis août 2012⁴. Mené par le JISC, ce projet partage la mise à jour des données entre

tous les acteurs de la chaîne, de l'éditeur au bibliothécaire, en passant par les agences d'abonnement ou les fournisseurs de bases de connaissance bien sûr. En France, il existe déjà plusieurs exemples de veille partagée sur les périodiques électroniques, mais rares (?) sont les bibliothèques qui font l'économie d'une base de connaissance commerciale. Les données de la Bimpe⁵ ou de Mirabel⁶ servent de référentiel, mais la réutilisation des données pour alimenter des applications extérieures reste ponctuelle.

Il semble donc que *flexibilité* et *interopérabilité* soient les maîtres mots d'un projet de base de connaissance partagée. Pour être réellement efficace, elle doit prendre en compte les évolutions structurelles (organisations consortiales à plusieurs niveaux), d'une part, et logicielles, d'autre part : les outils de découverte sont les catalogues de demain – voire d'aujourd'hui. Ils sont adossés à des bases de connaissance au départ conçues pour accéder au texte intégral et non pas pour l'indexer. Les bibliothécaires ont la sensation d'être face à une boîte noire. Le projet de hub de métadonnées de l'ABES, qui reprend la main sur les données pour les enrichir, paraît donc une solution appropriée pour sortir de la dépendance vis-à-vis des fournisseurs de bases de connaissance.

CÉCILE CLARAC

Responsable de la gestion des ressources électroniques
SICD de Toulouse

cecile.clarac@univ-toulouse.fr

[3] KBART : Bases de connaissance et outils associés, rapport préparé par le groupe de travail NISO/UKSG KBART, janvier 2010, traduit en français par Thomas Porquet et Karine Delvert, www.couperin.org/groupe-de-travail-et-projets-deap/acces-aux-ressources-cat/recommandations-kbart

« La qualité d'une base de connaissances repose en grande partie sur les données que les fournisseurs de contenu (éditeurs, agrégateurs, etc.) fournissent aux développeurs de la base de connaissance. Les erreurs contenues dans ces données se retrouvent souvent dans la base de connaissance. En outre, du fait de l'absence de norme pour ce type de données, les développeurs de base de connaissance doivent consacrer beaucoup d'efforts à convertir les listes de titres des différents fournisseurs en un seul et même format, ce qui peut à son tour être la cause de nouvelles erreurs et rendre le pistage de ces erreurs difficile. » [page 7].

[4] www.kbplus.ac.uk/kbplus

[5] http://bimpe.free.fr/bimpe_accueil.php

[6] www.reseau-mirabel.info