

L'enjeu des métadonnées

dans l'édition et la commercialisation du livre

Depuis le début des années 2000, les éditeurs découvrent un nouveau métier : celui de... bibliothécaire. Ou, plus précisément, l'une de ses plus anciennes missions, le catalogage des livres. Jusqu'alors, les éditeurs ignoraient largement les techniques de production des données techniques, commerciales et bibliographiques servant à décrire les livres qu'ils publiaient – et *a fortiori* des enjeux liés à cette production. Mais l'essor spectaculaire des « librairies en ligne » (plus de 10 % du marché en 2011, souvent beaucoup plus pour certains secteurs), puis le développement de l'édition électronique depuis 2007, ont totalement changé la donne, obligeant les éditeurs à intégrer largement chez eux la production de ces fameuses « métadonnées ». Pour prendre la mesure de ce bouleversement, un bref historique (limité au cas français) n'est pas inutile.

Des bases de données conçues sans les éditeurs...

Depuis 1856, le Cercle de la librairie, réunissant libraires, éditeurs et imprimeurs, publiait la *Bibliographie de la France* (créée en 1811), dont la partie proprement bibliographique deviendra dans les années 1970 une publication annuelle, *Les Livres disponibles*. Au début des années 1980, la création par le Cercle de la base de données Électre (lancée en 1986), destinée à rendre ces données bibliographiques accessibles par voie informatique aux professionnels (libraires et bibliothécaires) comme aux particuliers (via le Minitel), se fit pour l'essentiel sans la participation des éditeurs, lesquels avaient déjà mis quelques années à se convertir à l'usage de l'ISBN, introduit en 1972. Électre fut d'ailleurs, dans un premier temps, un fiasco technique et financier, qui ne fut surmonté que par le soutien décisif du directeur du livre et de la lecture Jean Gattégno, puis par la scission en 1988 de la société – « Électre Bibliographie » (l'actuelle Électre) assurant les fonctions de catalogage et « Électre Transmission » (devenue ensuite Édillectre puis Dilicom, cogérée par les distributeurs et les libraires) celles de transmission par échange de données informatisé (EDI) des commandes des libraires aux distributeurs.

Au fil des années, ces deux structures interprofessionnelles sont devenues remarquablement efficaces, mais cela s'est fait pour l'essentiel sans les éditeurs. Ceux-ci se sont contentés de sous-traiter à leurs distributeurs la mise en forme et la (complexe) normalisation des données techniques et commerciales de leurs livres, via Dilicom et la Commission de liaison interprofessionnelle du livre (CLIL, constituée en 1991 pour gérer le transport du livre), qui ont créé en 2001 (avec Hachette) le « Fichier exhaustif du livre » (FEL), lequel comptait plus de 1 100 000 références en 2009. Tandis qu'Électre continuait à créer des notices

bibliographiques au format UNIMARC (1 200 000 titres en 2012, dont 500 000 épuisés et 14 000 à paraître), à partir des spécimens envoyés par les éditeurs, puis de leurs transferts par EDI de quelques métadonnées basiques.

En bref, un travail faisant largement double emploi avec celui du Catalogue général de la BNF (constitué à partir du dépôt légal), devenu lui aussi de plus en plus performant dans les années 2000, au point de concurrencer désormais Électre dans l'établissement rapide de notices bibliographiques des nouveautés. Sans parler des multiples autres bases de données bibliographiques et commerciales créées depuis les années 1990 par divers acteurs privés (comme Tite-Live ou Decitre) pour pallier les insuffisances des outils institutionnels (tarifs trop élevés d'abonnement à Électre, délais jugés trop longs de la BNF, etc.).

Depuis plus de vingt ans, nous avons été quelques-uns, bibliothécaires, libraires et éditeurs, à tenter de faire converger ces entreprises inutilement concurrentes, mais sans succès. Un exemple parmi d'autres : dans les années 1990, afin de faciliter le travail des bibliothécaires, les éditeurs de sciences humaines et sociales du groupe universitaire du Syndicat national de l'édition¹ (SNE) ont obtenu d'Électre qu'elle produise à l'avance des notices de catalogage (avec références Rameau et Dewey) qu'ils faisaient figurer sur la page de copyright de leurs livres, comme cela se faisait déjà de longue date dans nombre d'autres pays, des États-Unis au Brésil. Mais les responsables d'Électre de l'époque ont finalement renoncé à produire ce service, jugé trop coûteux...

Ce rappel historique indique à quel point la plupart des éditeurs français sont restés longtemps étrangers à la culture de la production de métadonnées, au mieux considérée comme une affaire de commerçants (les libraires, à faire traiter par leur distributeur) ou de conservateurs (les bibliothécaires). C'est surtout l'émergence imprévue du « commerce en ligne », à partir de 2000, qui les a amenés à se remettre en cause sur ce point.

...à la production de métadonnées par les éditeurs...

Les grands groupes d'édition, parce qu'ils avaient les moyens de ce lourd investissement, ont été les premiers à créer des bases de données commerciales et bibliographiques internes, directement renseignées par les éditeurs eux-mêmes, à partir de formulaires précis et normalisés. À La Découverte, nous avons ainsi été partie prenante du projet « Yod@ », lancé en 2001 par notre groupe Vivendi Universal Publishing (devenu Editis en 2004). Celui-ci consistait à créer, pour chaque livre à paraître, une fiche renseignant tous les champs pertinents pour le décrire : titre et sous-titre (provisaires puis définitifs), auteurs, argumentaire puis quatrième de couverture, image de cou-

verture, table des matières, pagination, prix, format, catégorie CLIL et des dizaines d'autres champs actualisés au fil de la vie du livre (avant et après parution).

Cela nous a conduits à revoir en profondeur nos procédures de fonctionnement interne, facilitant la vie de tous les collaborateurs, puisque les données d'un livre n'étaient plus saisies qu'une seule fois – c'était loin d'être le cas auparavant – et devenaient accessibles à tous, par exemple pour produire de façon automatisée nos catalogues imprimés ou pour alimenter en permanence notre site web, créé en 2002 (lequel, de plus en plus visité, nous a permis d'abandonner en 2007 notre bulletin trimestriel imprimé annonçant nos nouveautés, remplacé par une *newsletter* électronique).

L'amélioration progressive de cette base, étendue à l'ensemble des catalogues des éditeurs d'Editis, a permis à notre distributeur Interforum de la mettre progressivement à disposition (gratuitement) des libraires en ligne. Non sans maints « bugs » et difficultés, faute d'une normalisation interprofessionnelle des protocoles techniques de production de ces métadonnées. La norme Onix² française élaborée pour résoudre ce problème (à partir de son modèle d'origine états-unien), dont les prolégomènes avaient été pourtant engagés à l'initiative du Cercle de la librairie dès 2001, a mis en effet de longues années à s'imposer. Même si la « commission numérique » de la CLIL, créée en 2011, œuvre désormais pour généraliser cette norme, tous les grands diffuseurs/distributeurs français n'ont pas achevé en 2011 cette migration (Interforum l'a fait en 2011, sa base Yod@ devenant « Yonix »). Et les plus petits diffuseurs/distributeurs, au service de centaines d'éditeurs plus « artisanaux », restent encore derrière, malgré leurs efforts constants.

Car cette fois, on ne peut plus incriminer le désintérêt des éditeurs face à l'importance de produire, en même temps que les livres qu'ils publient, des métadonnées numériques associées de bonne qualité. La grande majorité d'entre eux, quelle que soit leur taille, ont bien compris l'importance de cet enjeu, redoublé par l'impératif de produire, simultanément à la version imprimée d'un livre, sa version

électronique au format ePub, avec ses métadonnées spécifiques. D'où d'ailleurs un nouveau vertige, né de la nécessité de passer de l'identification d'un livre par son simple ISBN à celle d'une « œuvre » identifiée par un code maître (l'ISTC)³, permettant de cataloguer toutes les versions (imprimées et électroniques) d'un même livre. Malgré les recommandations du SNE⁴, cette procédure est encore bien loin, en France comme ailleurs, de s'être normalisée. Le chantier est en effet fort complexe (et coûteux), impliquant de profonds bouleversements des pratiques éditoriales. Mais aussi interprofessionnelles, car dans cet univers en constante évolution, l'établissement de nouvelles normes en matière de métadonnées implique impérativement une collaboration étroite entre éditeurs, distributeurs, libraires et bibliothécaires.

François Gèze

¹ <http://www.sne.fr/sne/groupe/universitaire.html>

² <http://www.onixedit.com/fr/Accueil.aspx>

³ http://www.istc-international.org/html/international_agency.aspx

⁴ <http://xvm-12-234.ghst.net/dossiers-et-enjeux/numerique/normes-et-standards.html>

Éditions La Découverte

<http://www.editionsladecouverte.fr>

François Gèze est P-DG des Éditions La Découverte depuis 1982. En dehors de son activité d'éditeur, il a toujours exercé de nombreuses responsabilités dans les regroupements professionnels (président du groupe des éditeurs de sciences humaines et sociales du SNE, vice-président du CFC, membre fondateur de l'ADELIC...)

f.geze@editionsladecouverte.com

Qu'est-ce que c'est ?

ePub : format de production, d'échange et de lecture pour les livres numériques, basé sur des standards web. Le contenu structuré est interprété pour une mise en forme adaptée à la lecture, quelle que soit la taille de l'écran. Au fil des versions de la norme, l'ePub permet, hormis des contenus sémantiques simples (textes, images), l'intégration et l'interprétation de contenus enrichis (dont XHTML, CSS, SVG, images, etc.). L'ePub a pour vocation de permettre la production d'un seul format numérique de publication qui alimente la chaîne de distribution jusqu'aux lecteurs finaux, autour d'un format interopérable quels que soient les moteurs de lecture et les machines qui les hébergent.

Onix (Online Information eXchange) : norme internationale bibliographique pour la diffusion de métadonnées enrichies. Ses principes directeurs comprennent des spécifications de contenu, d'éléments de données, d'étiquettes et de listes de codes et une DTD type XML. L'objectif est de créer un standard permettant aux éditeurs d'offrir à leurs clients une information « à valeur ajoutée » très complète (plus de 200 données possibles).

Sources : SNE, Travaux du groupe Normes & Standards, Assises du livre numérique, Salon du livre, 2012

<http://xvm-12-234.ghst.net/img/pdf/Dossiers/Normes%20et%20Standards/Guide-du-numerique-usage-editeurs.pdf>