

Section de catalogage Catalogage et métadonnées : blanc bonnet et bonnet blanc ?

par Antoine Provansal*

Compte rendu et réflexions autour de l'intervention de Stefan Gradmann (Pica) Cataloguing vs metadata : old wines in new bottles ? Réunion 126, Section de catalogage, mercredi 19 août 1998, thème : « Catalogues de bibliothèques : répondre aux besoins des usagers ».

Stefan Gradmann plaçait son intervention sous le signe d'une référence néo-testamentaire implicite : « On ne met pas non plus de vin nouveau dans de vieilles outres ; autrement les outres éclatent, le vin se répand et les outres sont perdues. Mais on met le vin nouveau dans des outres neuves. » (Mt-9-17). La référence évangélique, bien que fort intéressante d'un point de vue œno-

* Conservateur, Bibliothèque nationale de France.

logique, ne fait, pour nous Français qui n'aurions pas l'idée de verser un Graves 1974 dans une bouteille étoilée, qu'embrouiller les choses. La question que pose Gradmann est simple : les métadonnées¹ sont-elles ou non une version relookée du bon vieux catalogage de Papa ? Autrement dit, la différence entre les métadonnées du Dublin Core, pour donner un exemple, et les notices ISBD est-elle une différence de nature ou simplement d'environnement : le Web pour les premières, le catalogue MARC ou papier pour les secondes ?

L'article de Rachel Heery² ainsi que nombre d'autres textes fréquemment cités définissent de façon lapidaire les métadonnées (*metadata* en anglais) comme « des données sur les données ». C'est assez satisfaisant d'un point de vue logique ; le préfixe « méta » est une importation du domaine scientifique, ainsi en linguistique où le concept de métalangage désigne le discours sur la langue, c'est-à-dire la linguistique elle-même. Les bibliothécaires sont contents d'y retrouver les notions bibliographiques traditionnelles d'information primaire et d'information secondaire. Mais c'est aussi extrêmement vague. Ainsi, elle ne dit rien des types de relation entre les données « signalantes » et les données signalées. S'agit-il de description, d'identification, de localisation, de contrôle... ? Ou bien tout cela à la fois ? D'autre part, quelle est l'utilité pratique des métadonnées ?

Revenons rapidement sur le catalogage, cette activité fortement consommatrice de temps, mais si utile pour départager les candidats dans les concours administratifs... Le catalogage ISBD est avant tout une description du document faite par un agent humain. Cette description mêle les différents aspects de la notion de document, à la fois support matériel de l'information (livre, CD...), œuvre de l'esprit (ou groupement d'œuvres), édition de cette ou de ces œuvres, exemplaire physique possédé par un établissement documentaire³. La question des accès, par sujet, auteur, titre uniforme, vient après : ils constituent une superstructure ajoutée à la description et qui a sa propre logique. Le formatage MARC des données ne modifie pas fondamentalement ce système.

Les métadonnées sont, au contraire, conçues exclusivement pour permettre la sélection d'un document par les robots de recherche du Web. Cette sélection inclut différentes opérations qui sont l'identification, la recherche proprement dite, le contrôle d'accès, la validation et la localisation. Cette dernière étape est essentielle puisque la validité du lien permet la fourniture du document. Dans le Dublin Core, qui est une grille minimale élaborée sous l'impulsion d'OCLC et du W3Consortium au cours d'une série d'ateliers interna-

tionaux, les éléments sont choisis et balisés par l'auteur du document en fonction de leur efficacité dans la recherche. Les métadonnées sont donc conçues en fonction de leur utilisation et de l'utilisateur (elles sont *user oriented*), même si celui-ci est une machine.

Il faut ajouter que les documents sur Internet ont des caractéristiques qui les différencient absolument des documents dits « manipulables par l'utilisateur ». On a justement souvent avancé des termes différents pour les désigner : ressources électroniques ou objets semblables à des documents (*document like objects*). Ces objets documentaires n'existent pour la plupart qu'en un seul exemplaire. Gradmann souligne que la relation entre les métadonnées et la ressource identifiée est de type bijectif (1 : 1). De plus, les versions sont constamment susceptibles d'être modifiées. Qui a vu un site Web a de fortes chances d'avoir vu un autre site que celui qui l'a vu il y a un mois. D'autre part, sur Internet, on a un mal fou à définir ce qu'est l'unité documentaire que l'on décrit, signale et identifie. Tout objet documentaire est susceptible d'être intégré dans un objet documentaire supérieur qui lui-même peut, parfois mais pas toujours, être assimilé à une publication en série, et lui-même intégré dans un objet supérieur... etc. Concrètement, si l'on interroge le site de la BNF (www.bnf.fr), s'intéresse-t-on à des informations sur l'établissement, à Gallica ou bien à la photographie d'Atget que contient ce dernier ? Quelle unité doit-on identifier ? Il y a des droits, des dates, des auteurs particuliers qui s'appliquent à chacun de ces niveaux. « Last but not least », les ressources du Web se distinguent par leur quantité : plus de 320 millions de pages Web courant 1998, alors que le catalogue OCLC ne comporte péniblement que 40 millions de notices⁴. L'inflation documentaire est multipliée par la connexion des ordinateurs et par la possibilité pour chaque internaute de devenir potentiellement éditeur. Ces chiffres rendent bien sûr illusoire la possibilité de cataloguer manuellement les ressources du réseau. Les recommandations du Dublin Core sont donc extrêmement basiques et pragmatiques, afin de permettre leur application par l'auteur, sans formation préalable voire de façon quasi automatique, et bientôt, le langage XML qui est destiné à remplacer HTML permettra l'inclusion de ces métadonnées dans le document lui-même.

L'idée de Gradmann est que le modèle du catalogage et celui des métadonnées sont de nature différente, même s'ils ont des points communs, comme celui de permettre tous deux l'identification. L'attitude qui consiste à dire des métadonnées « On peut appeler ça du catalogage » engendre du côté des bibliothécaires une mauvaise compréhension de la problématique des métadonnées et de ces enjeux. Les métadonnées servent à la recherche de l'information par des robots et répondent

à des besoins d'authentification et de sécurisation des données, ce qui n'est pas le cas des notices de nos catalogues. Par contre, certains des outils mis en place par les bibliothécaires, comme les fichiers d'autorité, pourraient être utilisés dans ce nouveau contexte ; mais reste encore à définir de quelle façon. La simple mise en relation cartographique des champs et sous-champs des formats MARC avec les éléments du Dublin Core, déjà faite pour USMARC et UNIMARC, n'intéresse que moyennement les promoteurs de ces travaux, qui précisément ont posé comme principe de départ de se situer en dehors de ce modèle.

Reste enfin à savoir comment et qui va commercialiser les métadonnées. Vont-elles être vendues par des fournisseurs extérieurs ? Qui va avoir le rôle de validation de l'information ? On imagine que les réponses à ces questions auront de grandes conséquences, notamment financières, pour les bibliothèques.

1. Si c'est la première fois que vous entendez parler de métadonnées, vous pouvez regarder le site de l'IFLA (<http://ifla.inist.fr/ifla//metadata.htm>) ou la liste des éléments du DC en français à <http://www-rocq.inria.fr/vercoust/METADATA/DC-french.html>

2. Heery, Rachel : *Review of Metadata Formats*. In : *Program*, Vol. 30, n° 4, October 1996, p. 345-373

3. Ces dernières données sont heureusement de plus en plus fréquemment séparées des notices bibliographiques.

4. Teasdale, Guy : *Métachronique XML*. In : *La Lettre du bibliothécaire québécois*, n° 11, mai-juin 1998. Accessible à : <http://www.bibl.ulaval.ca/info/pagepers/teasdale/LBQ/lettre11.htm>.