

# Éplucher, émincer, cuire...

## Des notices qui divaguaient en liberté...

Avant le milieu du <sup>xx</sup>e siècle, les bibliothèques décrivaient les documents selon leurs propres règles. Pendant des siècles, elles ont travaillé avec des inventaires, des listes de titres, plutôt que des catalogues, qui permettent de rassembler les ouvrages d'un même auteur ou sur le même sujet. Depuis quelques décennies, les bibliothèques cataloguent leur fonds selon des règles internationales : en 1971, la publication de la première norme « ISBD (M) » (International Standard Book Description for Monograph Publications) donne aux bibliothécaires un langage commun. Mais si cette normalisation suffisait pour harmoniser des catalogues sur papier, elle ne suffit plus à l'ère de l'échange de données informatisées : il faut aussi que les données aient un format compris de tous les programmes. On ajoute donc à une description qui suit les mêmes règles un format international d'échange. Ce format, l'UNIMARC (pour UNiversal MARC) remonte à 1977, et a été adopté en France comme « format national d'échange » en 1993.<sup>1</sup>

Les bibliothèques peuvent échanger leurs données pour que l'utilisateur ait accès à tous leurs documents ; un grand chantier reste cependant ouvert : la « conversion rétrospective » des notices rédigées sur un support papier (fiches ou registres) en notices dématérialisées au format UNIMARC. Quelques sociétés se partagent le marché, et si les techniques diffèrent – saisie manuelle ou numérisation directe – le résultat obtenu est toujours différent de ce que serait une notice créée directement sur un logiciel de catalogage, selon les règles en vigueur aujourd'hui.

## ...aux notices génétiquement modifiées

Pour un catalogue informatisé, les notices de conversion rétrospective sont parfois aussi indigestes que la tambouille d'une gargote : elles sont marquées du code « i » (notice partiellement conforme à l'ISBD) ou même « n » (non conforme à l'ISBD) dans le label UNIMARC, en position 18.

En quoi cela est-il un problème de santé pour un catalogue (et pour ses gestionnaires) ? Parce que l'information sur la publication décrite n'est pas bien répartie dans les zones ad hoc, qui n'existaient pas lors de la rédaction de la notice : les notices sur papier ont été saisies telles quelles, ou presque, et « reformatées », c'est-à-dire découpées selon le format UNIMARC.

« Mais encore ? » dira celui pour qui la gestion des catalogues est une sinécure. En fait, les programmes informatiques, bien moins avisés que les bibliothécaires, ne savent pas toujours retrouver les informations : par exemple ils ignorent le nombre de pages d'un document s'il est ailleurs que dans la sous-zone UNIMARC dédiée. Il faut préciser ici que ce format se découpe en zones et sous-zones, où une zone est identifiée par une « étiquette » à trois chiffres, et une sous-zone par un code (une lettre) introduit par le signe dollar « \$ ».

Le nombre de pages d'un document devrait se trouver dans la zone 215, sous-zone \$a, exclusivement. Et lors d'un chargement, lorsque l'on compare aux notices du Sudoc celles que l'on veut y faire entrer, il faut, pour que la comparaison ait un sens, que les éléments de chaque notice soient dans les zones dédiées du format.

Le défi de l'intégration dans le Sudoc de ces notices de conversion rétrospective non-ISBD conduit à élaborer des recettes originales pour éviter une forme grave d'intoxication : la « doublonnite aiguë ». Le service *Données* de l'ABES utilise à cette fin des moulinettes qui recomposent les notices au moment de la conversion.

Les moulinettes d'importation sont une obligation : le format interne du Sudoc n'est pas l'UNIMARC, mais un format propriétaire, le Pica+. Il faut donc convertir les notices du format national d'échange au format interne. On profite de l'occasion pour rapprocher autant que possible les notices reçues des notices produites actuellement. Comment savoir où se trouvent la mention de responsabilité, le lieu d'édition, la collation, ces éléments indispensables de l'identification d'une publication ?

Pour y parvenir, il faut mouliner les seuls éléments disponibles : la ponctuation et le vocabulaire.

## Éplucher le titre

Retrouver le début du titre et mettre à sa place la mention de responsabilité est la première étape de la recette. C'est en ISBD un élément localisé dans la « zone du titre et de la mention de responsabilité », en UNIMARC la zone 200.

L'identification du début du titre, qui marque le premier caractère indexé, permet de ne pas avoir de mots parasites qui craquent sous la dent, tel l'insecte dans la salade. Par ailleurs, à l'autre extrémité du titre, les mots qui peuvent introduire la mention de responsabilité, doivent être repérés, pour être les marqueurs de la sous-zone spécifique. Mais la sauce ne peut être liée qu'à l'aide d'une ponctuation décisive et identique d'une notice à l'autre.

Pour rendre ces lignes moins indigestes, on ne décrira pas l'identification des autres éléments de la zone, ni les indicateurs.

### Où commence le titre ?

Identifier le début du titre propre à l'air simple et dépourvu de pièges : c'est compter sans le fait qu'en UNIMARC, le titre doit comporter des caractères de non-tri encadrant l'article initial présent. La BNF s'est servi de ces caractères de non-tri d'une façon originale, dans les notices dites « CGI » (Catalogue général des imprimés).

Les fiches saisies commencent parfois par le nom de l'auteur, au lieu du titre. Dans ce cas, l'auteur est encadré de caractères de non-tri, et la moulinette sait où commence vraiment le titre propre.

Par exemple

**\$a<Nicodème Paillon.>Aux origines de la socca<sup>2</sup>**

Où < et > représentent les codes des caractères de non-tri.

Le programme de conversion utilise cette chaîne de caractères « Nicodème Paillon », qui doit être ignorée, pour trouver le premier caractère du titre : il la remplace dans la mention de responsabilité, c'est-à-dire dans le \$f, tout en traitant correctement l'article si besoin est.

Et si la zone ne peut pas fournir de \$f (si utile quand le titre se résume à « Œuvres complètes »), on peut aller plus loin dans la notice, chercher s'il existe une zone 700 (auteur personne physique) ou 710 (auteur collectivité) ou 720 (auteur nom de famille), et l'y copier, en inversant les éléments du nom de personne.

**\$a@Aux origines de la socca\$fNicodème Paillon**

Parfois les notices n'ont aucun caractère de non-tri : il faut alors scruter le premier mot de la zone titre, et si c'est un article à éliminer (présent dans une liste noire) c'est après lui que l'on place le caractère qui marque le vrai début du titre.

Voici la structure « farcie » que l'on rencontre dans un lot de notices de thèses.

**\$a**Université de Nice. Faculté de cuisine. La pissaladière, sa préparation, sa cuisson

Ici, il faut tester le premier mot du titre : si c'est « Université », avancez jusqu'au point suivant. Puis regardez le mot qui suit le premier point : si c'est « Faculté », avancez encore jusqu'au point suivant. Cela relève autant du jeu de l'oie que de la moulinette à notices. Et enfin, n'oubliez pas que **La** est sur la liste noire, c'est un article.

Le résultat sera :

**\$aLa** @pissaladière, sa préparation, sa cuisson

où @ marque le début du premier mot indexé du titre.

### Où finit le titre ?

Les mots utilisés par les bibliothécaires pour introduire une mention de responsabilité entrent dans une liste longue mais, heureusement, pas infinie.

En éminçant la notice selon la ponctuation, et en analysant si dans un des fragments obtenus on lit **par** on retrouve la mention de responsabilité.

#### Deux exemples

La ratatouille dans tous ses états, **par** Charles Nissart  
scindée en deux tronçons par la virgule, peut donner

**\$aLa** @ratatouille dans tous ses états**\$f**par Charles Nissart  
où **\$f** marque le début de la mention de responsabilité.  
La version ci-dessus est rendue tout à fait digeste pour le catalogue mais  
La ratatouille niçoise par Charles Nissart  
sans aucune ponctuation, donnera au mieux

**\$aLa** @ratatouille niçoise par Charles Nissart  
où le contenu du **\$a** englobe le nom de l'auteur : la ratatouille est trop lourde, la chaîne de caractères fait une longueur double de ce qu'elle devrait faire.

Car imaginez que l'on ait ce titre  
La culture de la blette par les paysans de l'an mil, rédigé par Alphonsine Trognon

Si l'on suppose que **par** suffit pour trouver la mention de responsabilité, on obtient  
**\$aLa** @culture de la blette**\$f**par les paysans de l'an mil, rédigé**\$g** par Alphonsine Trognon

Alors que si la ponctuation et les mots clés sont reliés dans l'analyse de la structure de la chaîne de caractères, on retrouve un titre correct  
**\$aLa** @culture de la blette par les paysans de l'an mil**\$f**rédigé par Alphonsine Trognon

À cet aromate doux-amer qu'est la ponctuation (absente / présente) s'ajoute une décision à prendre à propos de la casse des mots utilisés comme marqueurs : faut-il les chercher uniquement en minuscules, ou étendre la recherche aux éléments comportant des majuscules ? Seule une analyse poussée du fichier à convertir peut le dire.

## Trouver la bonne adresse

Une autre zone nécessaire à l'identification du document est celle de l'adresse. La ponctuation ISBD permet de distinguer « lieu : éditeur, date », que l'UNIMARC codifie **\$a**, **\$c**, **\$d**. Mais dans des notices de conversion rétrospective, on peut trouver un état mixte entre ponctuation et marquage par sous-zones, comme : « lieu, éditeur**\$d**date ». Il faut donc pouvoir distinguer les deux premières sous-zones par la présence de la virgule.

C'est trop facile, car un programme informatique balaie la notice dans l'ordre où elle se présente, sauf si on le force à revenir en arrière. Donc, on sait que la première virgule rencontrée introduit l'éditeur. La date sera l'élément qui suit le **\$d**. Elle peut être entourée de parenthèses, suivie d'un point, ou d'une date plus précise, tous éléments à identifier et éliminer proprement.

Mais tout se complique, hélas, après la date de publication.

Dans une bonne partie des notices, les éléments de la collation ne sont pas dans la zone prévue (**215**), mais directement à la suite de la date, en **210\$d**, comme ceci :

**210\$a**Barcelona, Libreria Bosch**\$d**1947. In-8°, XII-463 p., fig. et pl.

## Prendre une petite collation

Une fois de plus, il faut recourir à l'alliance du vocabulaire et de la ponctuation.

**.In-** signale le début de la zone de la collation, qui commence par le format, et non pas par le nombre de pages, comme dans les notices standard. Mais ce peut être aussi **.Gr. in-** et pourquoi pas **.Grand in-**. Gare à celui qui voudra faire le malin en réduisant la recherche à **Gr**, chaîne commune aux deux expressions : une sous-zone commençant par **Grenoble** sera traitée comme de la pagination, et non comme le lieu d'édition !

Comment retrouver cette pagination ? Il faut que la rondelle obtenue contienne **p**. Oui, mais si ce sont des colonnes, et pas des pages ? Un test de plus.

Quant aux illustrations, elles sont dans le fragment de zone qui n'entre ni dans la catégorie **pagination**, ni dans la catégorie **format**.

« Il suffit » donc de découper la zone. Mais en fait, il ne faut pas moins de cinq étapes de conversion des tronçons de zones dans des sous-zones intermédiaires pour arriver au bout de la zone de la collation.

**.In-8°, XII-463 p., fig. et pl.**  
devient

**215 \$aXII-463 p.\$cfig. et pl.\$dIn-8°**

## Éplucher, émincer, cuire...

Illustration de l'auteur M. T.  teissedre@abes.fr



## Éviter les fausses notes

En continuant le balayage des zones ISBD, on arrive aux notes : UNIMARC zone **300** (note générale), morceau de choix ! On y trouve tout ce qui n'a pas sa place dans le corps de la notice : mention de collection (**225**), titre traduit par le catalogueur (**541**), mention d'extrait d'un périodique (**305**), liste des œuvres contenues (**327**), résumé de la carrière de l'auteur (pourquoi pas **330**), supplément à un périodique (**311**), mention de thèse (**328**), ou encore des notes relatives à un exemplaire, qui n'ont pas leur place dans le corps de la notice d'un catalogue collectif.

Une salade pas très niçoise, cette zone **300** !

## Cuire à feu vif

Les méthodes maintenant éprouvées permettent une cuisson à feu vif, sans que les notices attachent au fond du serveur.

Faire gratiner est tout à fait néfaste, car le temps de traitement d'un lot serait augmenté de la durée de récurage du programme.

## Servir tiède

Une fois que la recette est élaborée, on fait venir les goûteurs : le mets est présenté sur trois plats, trois formats différents de la même notice (format reçu en entrée, format interne Pica +, format de catalogue du Sudoc). Les trois plats doivent avoir la même saveur, même si l'aspect en est différent. La réaction des goûteurs est essentielle : leur digestion est surveillée une semaine durant. Au moindre signe de malaise, il faut revenir sur le programme de conversion, ajouter un peu de basilic sur la collection, écumer les notes surabondantes, lier la mention de responsabilité.

Ce n'est que lorsque le cuisinier est sûr que les goûteurs ont survécu à l'ingestion des notices que le catalogue est alimenté à son tour.

**200 \$a**<L'Abbé Desanges.>De l'art de faire des beignets de fleurs de courgettes

**210 \$a**Aspremont, Librairie du Château**\$d**1925 (25 octobre). **In-8°**, 25 p., ill. portr.

**300 \$a**Collection Cuisine du terroir

**200 \$a**@De l'art de faire des beignets de fleurs de courgettes**\$f**l'Abbé Desanges

**210 \$a**Aspremont**\$c**Librairie du Château**\$d**1925

**215 \$a**25 p.**\$c**ill. portr.**\$d**In-8°

**225 \$a**Collection Cuisine du terroir

M. Teissèdre

 donnees@abes.fr

<sup>1</sup> Pour en savoir plus sur les formats et leur utilisation :

[www.rmbm.org/rencontres\\_2004/leresche-marc.pdf](http://www.rmbm.org/rencontres_2004/leresche-marc.pdf)

<sup>2</sup> Toute ressemblance d'un exemple avec des notices existant ou ayant existé ne serait qu'une amusante coïncidence

## Miroir de faille dans la vallée du Cinca en Aragon – Espagne

Cette surface subverticale, engendrée par la cassure, a été polie par le frottement des deux compartiments. Elle donne, lorsque l'érosion altère un des compartiments, un miroir de faille sur lequel on voit des stries qui matérialisent la direction du mouvement – faille normale.

Photo prise par Philippe Razin, maître de conférences à l'EGID – voir p.10

 Philippe.RAZIN@egid.u-bordeaux.fr



## Terre, océans et environnement...

### Sciences de la terre, des océans et de l'environnement terrestre

tel est l'intitulé des champs d'activités du CADIST implanté à la bibliothèque interuniversitaire scientifique Jussieu – BIUSJ. Arabesques a emprunté à ce centre d'acquisition et de diffusion de l'information scientifique et technique son appellation, pour en faire le thème de son numéro 38 ; thème qui avait été arrêté avant le 26 décembre dernier. Les éléments se sont chargés de rappeler, à qui s'occupe de documentation universitaire dans ces disciplines, toute leur actualité.

Et, en effet, l'eau, la terre et le ciel, le climat, les catastrophes naturelles mobilisent de plus en plus d'enseignants et de chercheurs... et sont donc des centres d'intérêt privilégiés pour les bibliothèques de l'enseignement supérieur concernées par ces sujets.

Les collaborations à ce numéro le soulignent, les exemples donnés ne pouvant être que des aperçus ; on remarque l'étendue et la diversité de la documentation, la multiplicité et l'originalité des recherches, la multitude et la densité des informations et des « sites ».

Arabesques a ouvert ses pages à des « anciens » du Système universitaire de documentation, au « CADIST de la BIUSJ » qui a hérité des collections de la Sorbonne et au service commun de la documentation de l'Université du Maine qui fut un des sept sites pilotes ; mais aussi à une petite bibliothèque associée qui voudrait l'intégrer et, enfin, à deux « nouveaux » : une école à organisation régionalisée, l'École nationale supérieure d'arts et métiers, et une bibliothèque de finistère qui a pris le nom d'un arpenteur de mer, La Pérouse.

S. Salvit  salvit@abes.fr