

**Isabelle Boudet**

*Bibliothèque de France  
Département Bibliothéconomie  
et relations avec les bibliothèques*

**Thierry Cloarec**

*Bibliothèque de France  
Département Bibliothéconomie  
et relations avec les bibliothèques*

# PROJET D'ENRICHISSEMENT POUR LE CATALOGUE DE LA BIBLIOTHÈQUE DE FRANCE

**L**E CATALOGUE sera au cœur de la Bibliothèque de France l'instrument indispensable aux lecteurs pour repérer dans la diversité des espaces et des collections ce qu'ils sont venus chercher. Quel sera le catalogue de la Bibliothèque de France ? Exhaustivité, innovation et enrichissement sont les trois clés de sa conception et de sa réalisation.

## Exhaustivité

Il sera d'abord une gigantesque base de données, issue de plusieurs entreprises qui vont converger :

- le catalogue informatisé de la Bibliothèque nationale, OPALE, fruit de travaux très importants menés depuis dix ans, et déjà riche de plus d'un million de références des vingt dernières années ;

---

Isabelle BOUDET et Thierry CLOAREC sont respectivement chef de projet Catalogue et délégué pour la conception informatique.

- les six millions de notices des catalogues anciens de la Bibliothèque nationale qui sont en cours de saisie informatique et qui signaleront tous les livres et périodiques conservés depuis la découverte de l'imprimerie ;
- le catalogue des acquisitions que l'Établissement public de la Bibliothèque de France (EPBF) constitue actuellement.

Le catalogue de la Bibliothèque de France fusionnera dans un même ensemble ces différentes bases d'imprimés et celles de la Phonothèque de la BN et du département de l'Audiovisuel de l'EPBF, ainsi que des données de l'Institut national de l'audiovisuel.

Il décrira donc toutes les ressources de l'image et du son, des produits éditoriaux de type nouveau, comme les documents électroniques, ou ceux qui, composites, associent de multiples supports.

Comme le recommandait le groupe Catalogue qui s'est réuni du mois de septembre 1990 au mois de mars

1991, dans le cadre des groupes de travail mis en place par l'EPBF, le catalogue sera unique, encyclopédique et multimédia.

## Innovation

Le catalogage courant dans la future Bibliothèque de France reprendra les acquis de la Bibliothèque nationale en matière de règles de catalogage, de format et de fichiers d'autorité.

Des évolutions sont d'ores et déjà en cours. En matière de format, par exemple, des travaux menés par le CCBT<sup>1</sup> et l'EPBF sur un format INTERMARC « intégré » visent à mettre en cohérence et à mieux adapter les descriptions des différents types de documents audiovisuels.

A l'issue d'une étude préconisée par le groupe Catalogue et menée par l'EPBF sur la liste d'autorité matières

---

1. CCBT : Centre de coordination bibliographique et technique.

RAMEAU<sup>2</sup>, un certain nombre d'actions pour développer la liste et renforcer les moyens et les outils de sa gestion vont être proposées aux différents responsables de RAMEAU au niveau national.

Le groupe de travail Accès aux documents a choisi la classification Dewey pour le système de cotation des ouvrages en libre accès et a recommandé d'indexer tous les ouvrages de la future bibliothèque de cette façon,

même s'ils étaient destinés aux magasins. Le service de catalogue de l'EPBF utilise déjà la classification Dewey pour indexer et coter les documents.

Le catalogue est aussi et surtout une des pièces clés du système d'information de la future bibliothèque, dont la conception est en train de s'achever. Le module de catalogage reprendra largement à son compte l'expérience et le savoir-faire de la Bibliothèque nationale en matière de production de notices. A ses fonctionnalités familières aux catalogueurs de la Bibliothèque nationale s'ajouteront des facilités nouvelles, comme l'usage de la « souris » ou des fonctions de traitement de texte de type « couper / coller ».

Le module de consultation, le futur OPAC<sup>3</sup>, conjuguera de puissantes capacités de recherche à une ergonomie qui en simplifiera l'utilisation.

## Enrichissement

Une autre recommandation du groupe Catalogue à l'EPBF était de mener une étude sur l'enrichissement des notices du catalogue avec des extraits de documents primaires.

Les techniques de description et d'analyse des documents ont en effet des limites qu'il est possible de dépasser pour fournir à l'utilisateur une information plus précise et plus complète à l'aide de données extraites des documents primaires. L'hypothèse est que des extraits significatifs donnent à l'utilisateur une meilleure représentation du contenu du document et l'aident dans ses choix. Ils lui permettent de vérifier l'adéquation d'un document à sa recherche et lui évitent d'inutiles recours au document lui-même.

C'est cette hypothèse pour le catalogue courant de la Bibliothèque de France que l'étude devait vérifier en ayant pour objectif de définir les différents extraits significatifs, d'en évaluer la pertinence et d'analyser la faisabilité d'une telle opération.

Deux limites ont été faites au champ de l'étude :

– d'une part, dans un premier temps, et dans le cadre de l'appel d'offres

pour la conception du système d'information, l'utilisation des extraits est réservée à la consultation, même s'il est envisageable par la suite de s'en servir comme points d'accès pour la recherche ;

– d'autre part, cette étude avait pour objet les documents imprimés, livres ou documents d'accompagnement des documents audiovisuels ; l'image et le son n'en faisaient donc pas partie.

Menée par la société TOSCA/AITIC, l'étude s'est déroulée du mois d'avril à octobre 1992. Un comité de pilotage regroupant des représentants de l'EPBF et de la Bibliothèque nationale l'a animée.

La notion d'extrait significatif a été testée sur un échantillon représentatif d'environ mille documents établi à partir de projections du secteur Politique documentaire. En fonction des différents types de documents retenus, et des différents types d'extraits (table des matières, quatrième de couverture, index et bibliographie), une grille d'analyse a permis d'évaluer l'importance de l'extrait et son apport d'information comparé aux notices bibliographiques décrivant les mêmes ouvrages.

2. RAMEAU : Répertoire d'autorité matières encyclopédique et alphabétique unifié.

3. OPAC : Online Public Access Catalogue.

Différents scénarios ont été étudiés du point de vue du choix des extraits et de l'organisation de la chaîne de traitement. Enfin, une partie très importante de l'étude était consacrée aux questions techniques et informatiques.

Quelles conclusions peut-on en tirer et quelles perspectives l'EPBF en a-t-il dégagées ?

Pour les éléments accompagnant les documents audiovisuels, la notion d'enrichissement s'est révélée peu pertinente, soit à cause de difficultés techniques d'extraction (pochettes de disques en couleurs, caractères fantaisie ne passant pas en OCR<sup>4</sup>, etc.), soit parce que les notices elles-mêmes fournissent les informations complètes (dépouillement des différentes parties).

En ce qui concerne les imprimés, un certain nombre d'orientations peuvent être retenues :

#### 1. *Choix des extraits : table des matières*

La table des matières est l'extrait par excellence : 99 % de taux de pertinence d'après l'évaluation réalisée par l'étude.

#### 2. *Type de documents : monographies à caractère documentaire*

Sont exclus la fiction, les ouvrages de référence (dictionnaires et encyclopédies dont la table des matières n'apporte pas d'informations significatives) et les périodiques (bases de données documentaires existantes).

#### 3. *Sélection de l'extrait : systématique*

La faible proportion de tables de matières non significatives parmi les documents retenus permet d'envisager une intégration systématique de cet extrait, sans tri préalable. L'organisation du traitement en est simplifiée et l'homogénéité du catalogue garantie. On peut mieux évaluer les volumes concernés, en fonction des accroissements prévus, et donc maîtriser les coûts.

#### 4. *Mode de stockage et de restitution : texte*

Les deux modes de stockage et de restitution possibles, analysés par l'étude informatique, sont le mode texte et le mode image. Le mode texte semble largement préférable. Malgré des coûts de production plus élevés, ses avantages sont décisifs :

– beaucoup moins d'espace de stockage dans le futur système, avantage économique et technique pour la consultation ;

– plus grande rapidité de transmission sur les réseaux ; le mode image crée, à l'inverse, des risques de surcharge à la consultation très importants : soit on doit constituer des réseaux à hauts débits très coûteux, soit les temps de réponse sont dégradés ;

– les extraits en mode texte peuvent être diffusés sur n'importe quel écran, y compris vidéotex (mode image sur écran graphique) ;

– à plus long terme, les extraits pourraient être utilisés pour la recherche, comme des points d'accès supplémentaires ; le mode texte est directement utilisable pour l'indexation et la recherche.

#### 5. *Organisation du traitement*

Les questions d'organisation du traitement sont en discussion.

Dans la mesure où le traitement de la table des matières est systématique, les opérations peuvent se situer avant les étapes de traitement intellectuel, au début du circuit du document. Après l'enregistrement d'une entrée, une photocopie de la table des matières est réalisée, sur laquelle on reporte l'identifiant de la notice.

La suite des opérations pourrait être effectuée en sous-traitance, pour ne pas alourdir davantage le circuit du document. Une société spécialisée pourrait effectuer des traitements par lots et fournir en retour un fichier qu'il reviendrait à l'EPBF d'intégrer dans sa base.

#### 6. *Coûts*

Pour obtenir un fichier en mode texte, différentes solutions techniques

peuvent être utilisées : soit saisie manuelle, soit scannérisation avec passage d'un logiciel de reconnaissance automatique de caractères.

L'estimation de cette dernière méthode, dans l'étude informatique, aboutit à un total de 9 F par page. Une table des matières ayant en moyenne quatre pages, le coût serait de 36 F par extrait.

#### 7. *Volumétrie*

La volumétrie des fichiers d'extraits a été évaluée par l'étude informatique à une moyenne de 3,2 K octets par table des matières en mode texte. On peut en déduire l'augmentation annuelle de la taille des fichiers d'extraits selon l'importance des accroissements prévus.

#### 8. *Scénario de démarrage*

Il est nécessaire de prévoir une montée en charge du projet. Un département thématique de l'EPBF semble l'entité la plus intéressante pour expérimenter l'organisation de la chaîne à une échelle réduite au départ et parce qu'il représente un ensemble intellectuel cohérent.

Le département des sciences exactes et appliquées pourrait être le premier : dans la perspective de travaux ultérieurs sur l'indexation des extraits, ce sont des disciplines dont les vocabulaires présentent moins de risques d'ambiguïté.

Dès la fin de 1994, un premier test pourrait commencer à l'EPBF, pour mettre au point l'organisation de la chaîne de traitement, et résoudre les questions techniques, en particulier l'intégration des données dans la base bibliographique.

Un premier bilan pourrait alors être fait, du point de vue économique et technique. Mais une véritable évaluation ne pourra être entreprise qu'après l'ouverture de la Bibliothèque de France, avec l'observation des pratiques du public. Car l'enjeu de ce projet d'enrichissement du catalogue est bien l'amélioration du service rendu aux lecteurs.

Février 1993

4. OCR : Optical Character Recognition.