

Le Catalogue général des manuscrits des bibliothèques publiques de France

La conversion rétrospective

Parmi les monuments bibliographiques présents dans nombre de bibliothèques françaises, le *Catalogue général des manuscrits* (CGM) occupe certainement une place de choix. Lancé sous la Monarchie de Juillet¹, ce catalogue a pour but de décrire l'ensemble des collections manuscrites des bibliothèques françaises à une exception notable, la Bibliothèque nationale de France (BnF) pour laquelle une série indépendante de catalogues se réalise à peu près à la même époque².

Fabien Plazannet

Direction du livre et de la lecture
fabien.plazannet@culture.gouv.fr

La parution du *Catalogue général* s'étale sur plus d'un siècle et demi³. Il constitue par conséquent un ensemble important⁴ et logiquement disparate ; l'étude de ses différentes strates constitue un bon reflet des progrès de la normalisation. Après

la parution d'une première série de sept volumes de 1849 à 1881, la fameuse série in-quarto consacrée à de nombreuses bibliothèques de province, la fin du XIX^e siècle voit la mise en place du format in-octavo et de deux séries parallèles (1885), l'une numérotée, consacrée aux bibliothèques départementales, la seconde dévolue aux fonds des établissements parisiens. En l'espace de cinquante ans, ce sont plus de quatre-vingts volumes qui sont ainsi publiés, signe d'un intérêt soutenu de l'État et des bibliothécaires pour les fonds patrimoniaux.

Les années 1950 voient la relance d'une publication interrompue pendant près de vingt ans (depuis 1933). Ce temps n'a cependant pas été perdu : on peut mesurer dans les nouveaux volumes les progrès de la description et la plus grande rigueur qui préside désormais à la rédaction des notices. Cette évolution est l'aboutis-

2. Rappelons pour mémoire que le premier catalogue spécifique des manuscrits de la Bibliothèque royale fut réalisé sous l'impulsion de l'abbé Bignon et publié en 4 volumes dans les années 1740. Ce catalogue est encore utilisé pour certaines tranches de cotes. Par ailleurs, la publication des catalogues des fonds latins et français fut entreprise sous le Second Empire. Pour plus de précision, nous renvoyons aux ouvrages classiques de Lydia Mériqot et Pierre Gasnault, *Les Catalogues du département des manuscrits de la Bibliothèque nationale*, Paris, 1974, de Werner Paravicini, *Die Nationalbibliothek in Paris*, Paris, 1981, ainsi que de Léopold Delisle, *Le Cabinet des manuscrits de la Bibliothèque impériale*, Paris, 1868-1881.

3. Sur les origines de l'entreprise et l'état de la collection en 1950, nous renvoyons à l'avant-propos du volume 49 du CGM sous la plume de Julien Cain lors de la reprise de la publication du *Catalogue général* en 1951.

4. Près de 110 volumes, environ 55 000 pages et 175 000 notices.

1. Circulaire de François Guizot, ministre de l'Instruction publique, du 22 novembre 1833 chargeant les bibliothécaires de lui adresser un catalogue « des manuscrits de tout genre confiés à leur garde ».

La conversion rétrospective du *Catalogue général des manuscrits...*

Archiviste-paléographe, **Fabien Plazannet** est conservateur de bibliothèque au Bureau des politiques documentaires de la DLL, après avoir occupé divers postes à la BnF. Il a collaboré au Mémento de l'action culturelle (Weka, 2003).

sement des travaux de normalisation importants qui se sont effectués à bien des niveaux depuis l'entre-deux-guerres⁵. Pourtant, seulement douze volumes paraissent de 1951 à 1981. Tous les fonds sont désormais décrits dans une même série.

À partir de 1981, la responsabilité de la collection puis de l'édition échoit à la seule Direction du livre, mais la publication s'essouffle. La forme de la publication, la lenteur de sa parution⁶ ne correspondent plus aux exigences nouvelles de la recherche documentaire qui se tourne vers les bases de données. Se pose enfin le problème des corrections, des ajouts et des mises à jour qui s'avèrent de plus en plus nécessaires et pour lesquels le cadre des suppléments imprimés apparaît désormais inadapté.

Un contexte en pleine évolution

Des réalisations en France

À première vue, le domaine des catalogues informatisés d'imprimés a connu de grandes réussites depuis vingt ans – citons les ouvertures successives de BN-Opale Plus, du Système universitaire de documentation (Sudoc), du Catalogue collectif de France (CCFr) –, tandis qu'assez peu de chose se passait dans le secteur du manuscrit.

Cette présentation est cependant un peu réductrice et doit être nuan-

5. L'ISA, ancêtre de l'ISO, et l'Afnor furent créées l'une et l'autre en 1926.

6. Six volumes parus seulement depuis 1981 ; le dernier (n° 66, *Suppléments de la bibliothèque de l' Arsenal et de la bibliothèque de Reims*) en 1993. Par ailleurs, plusieurs volumes n'ont pas été édités pour des raisons techniques et financières.

Le *Catalogue général des manuscrits* sur un siècle et demi

Date	Format, vol.	Éditeur	Resp. scientifique
1849-1885	4°, 7 vol.	Imprimerie nationale	Institut de France
1885-1933	8°, 82 vol. – 48 vol. départementaux (dont 9 de suppl.) – 34 vol. parisiens	Plon	Institut de France
1951-1981	8°, 12 vol.	Bibliothèque nationale	Direction des bibliothèques
1981-	8°, 6 vol.	CNRS puis DLL	DLL

cée : du côté des manuscrits, on peut mentionner l'existence ancienne de la base des manuscrits médiévaux *Medium* de l'Institut de recherche et d'histoire des textes (IRHT)⁷, la base *Mandragore* du département des Manuscrits de la BnF élaborée par le Centre de recherche sur les manuscrits enluminés⁸, la création du *Répertoire des manuscrits littéraires français du XX^e siècle* en 1995⁹, une expérience menée en 2001 par la Bibliothèque nationale de France pour l'informatisation du dernier volume de la série des *Nouvelles acquisitions françaises*¹⁰, l'ouverture récente des bases de données iconographiques *Enluminures* et *Liber Floridus*¹¹...

7. La base *Medium* décrit dans un format propre l'ensemble des manuscrits médiévaux présents dans le *Catalogue général des manuscrits* dont elle indique les numéros de notices. Elle est désormais consultable sur le site de l'IRHT : <http://www.irht.cnrs.fr>

8. <http://mandragore.bnf.fr>

9. La base *Palme du Répertoire des manuscrits littéraires* est hébergée sur la base BN-Opaline de la BnF. Elle contient désormais près de 50 000 notices de documents, en format Interمارc intégré.

10. Le volume imprimé paru en 1999 a été converti en XML (DTD-EAD) en 2001 sous forme de cédérom. Cette expérience est l'un des modèles retenus pour la conversion du CGM.

11. À l'automne 2002, la première (enluminures.culture.fr) par l'IRHT et le ministère de la Culture et de la Communication (*enluminures* d'un certain nombre de bibliothèques municipales principalement, à partir des campagnes de numérisation réalisées par l'IRHT et financées par le MCC) ; la seconde (liberfloridus.cines.fr) par le ministère de la Jeunesse, de l'Éducation nationale et de la Recherche (correspond actuellement aux manuscrits enluminés de la bibliothèque Sainte-Geneviève et de la bibliothèque Mazarine). Cf. l'article de Marie-Hélène de La Mure, « Le fruit, le flacon et l'ivresse », *BBF*, t. 48, n° 2, 2003, p. 95-100.

Des réalisations importantes à l'étranger

À l'étranger, de grands catalogues de manuscrits ont vu le jour au cours des dernières années sous des formes variées : catalogue multimédia incluant de nombreux manuscrits dans le cas de la Library of Congress, catalogue collectif de manuscrits dans le cas allemand ou encore italien (projet Manus), catalogue des manuscrits de la British Library¹².

Il semblait donc souhaitable de réaliser en France une base de données de taille significative dans le domaine du manuscrit, susceptible d'une part de fédérer les diverses bases déjà existantes, d'autre part de compléter les bases étrangères.

Le choix d'un format bibliographique

La définition du format bibliographique est au centre du projet de conversion rétrospective. Le choix d'un format doit se faire en fonction des paramètres que sont l'adaptabilité aux fonds décrits, les possibilités d'exploitation par les systèmes informatiques et le souci de la meilleure interopérabilité entre catalogues équivalents.

12. Consultable à l'adresse <http://www.molcat.bl.uk>, ce dernier catalogue provient de la conversion de 70 volumes imprimés, soit 30 000 pages environ et 1 million de notices, et décrit la quasi-totalité des fonds de manuscrits entrés à la BL depuis 1753. La structuration des données s'est faite par un programme de balisage spécifique...

L'adaptabilité du format aux fonds décrits

Que décrit le *Catalogue général* ? Les fonds recensés sont d'une grande variété : on peut trouver au fil des volumes aussi bien des manuscrits médiévaux enluminés que des manuscrits littéraires contemporains, des correspondances administratives, des papiers d'écrivain, des documents comptables, des fonds relatifs à l'histoire de l'art, des listes généalogiques, etc. Bref, des fonds concernant presque tous les domaines, toutes les époques et sous des formes matérielles très diverses.

Depuis plusieurs années, des débats divisent les bibliothécaires sur la question du traitement bibliographique informatisé des manuscrits. La question centrale est celle du niveau de description des documents : description pièce à pièce ou description globale de la collection ou du fonds avec son corollaire, l'utilisation ou non d'un format adapté du Marc et donc conçu à l'origine pour des imprimés.

Le format Marc permet un assez bon traitement pièce à pièce. Les adaptations réalisées dans le cadre de l'Intermarc intégré à la BnF ont montré qu'un Marc raffiné permettait de décrire de nombreuses notions chères aux spécialistes des manuscrits. Mais ce format a aussi ses défauts, en particulier la difficile gestion de multiples niveaux hiérarchiques au sein d'une collection, cas assez fréquent dans le domaine du manuscrit, en particulier pour les fonds contemporains. D'où le rejet du Marc par certains.

L'interopérabilité et le choix de la DTD EAD

De telles discussions ont eu lieu au sein de la Direction du livre et de la lecture¹³ puis du comité de pilotage

13. Une première présentation de ces débats avait été donnée par Jean-Arthur Creff dans son article « Quelle informatisation pour le *Catalogue général des manuscrits des bibliothèques publiques de France* ? », *Gazette du livre médiéval*, n° 39, automne 2001, p. 41-45.

du CGM, constitué à l'hiver 2002¹⁴. Pour sortir de ce débat sur le format, certains préconisaient une mise en ligne du CGM avec une très faible structuration, voire en plein texte.

Une étude de faisabilité a été réalisée en 2001 par la société Doxulting afin de faire le tour des solutions envisageables. Comme bien souvent,

Depuis plusieurs années, des débats divisent les bibliothécaires sur la question du traitement bibliographique informatisé des manuscrits

c'est du côté des États-Unis qu'une solution fut trouvée. À côté du Marc, une autre option fut relevée : celle du XML¹⁵ et de la DTD EAD (*Encoded archival description*). Élaborée dans le cadre de l'université de Berkeley à partir de 1993 et promue aujourd'hui par la Société des archivistes américains, la DTD EAD est un mode de structuration de l'information, à la fois riche et souple, particulièrement adapté au traitement des fonds d'archives. Elle est notamment utilisée aux États-Unis par RLG (Research Libraries Group). En France, la Direc-

14. Le comité de pilotage réunit des représentants de la DLL, de la Sous-direction des bibliothèques et de la documentation au ministère de l'Éducation nationale, de la Bibliothèque nationale de France et de l'Agence bibliographique de l'enseignement supérieur. Il est assisté par un comité opérationnel chargé du dialogue avec les différents prestataires.

15. Le XML (*eXtensible Markup Language*) est un langage de structuration et de présentation des données créé en 1997 et issu du langage SGML, lui-même élaboré dans les années 1980 sous l'impulsion de l'armée américaine en vue de l'échange de données, notamment dans le cadre des programmes CALS de l'Otan. La DTD (structure type de document) la plus célèbre du SGML est le HTML pour les pages web. À la différence du SGML, le XML sépare nettement les informations sur les données des informations sur la présentation (feuille de style) d'où la garantie d'une grande pérennité des données.

tion des archives de France défend depuis plusieurs années son utilisation pour les instruments de recherche des archives nationales et locales¹⁶.

L'enthousiasme de nos collègues américains et de certains de nos collègues français, la garantie du soutien logistique de la Bibliothèque du Congrès, l'expérience concluante menée à la Bibliothèque nationale de France ont conduit le comité de pilotage à valider ce choix au printemps 2002.

Comment fonctionne la DTD EAD ?

Le langage XML suit le principe du balisage des données : dans un flux continu d'informations, chaque élément signifiant est séparé des autres éléments par deux balises, une balise ouvrante et une balise fermante, qui indiquent où débute et où s'achève l'information de tel ou tel type.

La DTD est la structure qui « déclare » quelles sont les balises employées pour organiser le flux de données¹⁷. Dans l'EAD par exemple, il existe une balise <title> pour le titre, une balise <c> pour désigner un niveau hiérarchique, une balise <dao> pour lier une image numérique à une notice, des balises <abstract>, <scopecontent>, <notes>, <bibref> pour indiquer les résumés, les notes, les références bibliographiques, etc. Le principal avantage de l'EAD par rapport au Marc est la possibilité de gérer un grand nombre de niveaux hiérarchiques¹⁸. Le passage d'une in-

16. Pour plus d'information, nous renvoyons au site Internet de la Direction des archives de France (www.archivesdefrance.fr) ainsi qu'au site de la Bibliothèque du Congrès, agence support de la DTD EAD (www.locweb.loc.gov/ead/). Il faut savoir par ailleurs qu'une traduction en français de la DTD EAD est en cours au sein d'un groupe de travail de l'Afnor.

17. Dans le domaine du manuscrit, on peut citer aussi la DTD TEI (Text encoding initiative), plus ancienne (1994) et développée au niveau international pour structurer des textes littéraires ; elle est employée notamment à la Bibliothèque vaticane.

18. Jusqu'à douze niveaux en théorie.

formation en EAD à une information en Marc n'est pas une opération très complexe : il suffit d'établir une table d'équivalence entre une balise EAD et une zone Marc¹⁹.

En pratique, ce travail n'est évidemment pas toujours si simple car la DTD EAD, comme tout format, évolue et une notion dans un format n'a pas toujours son strict équivalent dans un autre format (mais c'est un problème que l'on connaît déjà au sein des formats Marc !).

La réalisation d'un prototype

Un double prototypage

L'importance du projet et la nouveauté des choix envisagés ont conduit les partenaires à adopter une démarche prudente s'appuyant sur la réalisation d'un prototype expérimental dont la maîtrise d'œuvre fut confiée à la société Doxulting.

Plusieurs pistes s'avérant techniquement possibles, il fut même décidé de faire réaliser deux prototypes différents par les sociétés Jouve et AIS : un premier prototype s'attacherait à développer des procédures de balisage essentiellement automatiques, tandis que le second ferait davantage appel à un traitement humain. Dans les deux cas, trois étapes étaient à distinguer : la numérisation des volumes imprimés, le traitement des données par reconnaissance optique de caractères, le balisage pour finir.

Le double prototype fut réalisé à partir d'un corpus de volumes correspondant aux différentes tranches du CGM²⁰ : les pages du volume 1 concernant le fonds de la bibliothèque de l'École de médecine de Montpellier (1849), le premier volume de la bibliothèque municipale de Rouen (1888), les quelques pages du volume 31 concernant la biblio-

19. La Bibliothèque du Congrès a déjà réalisé une telle équivalence entre EAD et Marc 21.

thèque municipale de Montauban (1898), deux volumes de la bibliothèque de l'Institut de France (1928 et 1965), les trois volumes de la bibliothèque du Muséum national d'histoire

L'importance du projet et la nouveauté des choix envisagés ont conduit les partenaires à adopter une démarche prudente s'appuyant sur la réalisation d'un prototype expérimental

naturelle (le plus récent, non publié, ayant été achevé en 1995 ; ils présentaient l'intérêt d'avoir déjà été saisis sous Word).

Des spécifications à la mise en ligne

La rédaction des spécifications fut entreprise en novembre 2002 avec AIS et Jouve et achevée en février 2003. Ce travail délicat a montré la complexité du *Catalogue*, assez fortement structuré, mais riche en cas particuliers et autres « anomalies » potentielles. Par ailleurs, des contacts ont été pris avec plusieurs entreprises développant des logiciels gérant le XML en mode natif.

20. Dans cette phase essentielle des spécifications, je tiens à souligner le rôle capital de Fabienne Queyroux, responsable du fonds ancien de la bibliothèque de l'Institut et formée à l'EAD aux États-Unis, et de Pascale Heurtel, responsable des fonds manuscrits de la bibliothèque du Muséum, qui, antérieurement au projet, s'étaient livrées l'une et l'autre à des essais de structuration en EAD de leurs catalogues de manuscrits. Citons aussi la participation de Marie-Dominique Nobécourt-Mutarelli et de Pierre-Yves Cachard, de la bibliothèque municipale de Rouen, ainsi que de Danièle Duclos-Faure, de la Sous-direction des bibliothèques et de la documentation du ministère de la Jeunesse, de l'Éducation nationale et de la Recherche.

Ce sont finalement deux sociétés, Software AG et Xyleme, qui ont été retenues et mises en concurrence pour la mise en ligne des prototypes. Le développement des maquettes a nécessité quelques semaines, et les deux prototypes de Jouve et AIS ont donc été consultables sur Internet *via* deux interfaces différentes pendant plusieurs semaines en mars 2003.

À la fin de l'expérience, le cabinet Doxulting a rendu un rapport technique dressant le bilan des difficultés rencontrées et préconisant notamment la rédaction de spécifications par groupes de volumes selon la date de rédaction. Dans l'ensemble, le résultat du prototypage a été jugé très positif par le comité de pilotage qui a décidé le 11 avril le passage à la phase industrielle de l'opération.

D'une conversion rétrospective à un projet français autour du manuscrit ?

De la phase de production à un futur Catalogue national de manuscrits

En raison de l'importance de l'opération, la Direction du livre et de la lecture a proposé à la Bibliothèque nationale de France de prendre en charge la réalisation de ce projet à partir de l'été 2003 en lui affectant une dotation budgétaire spécifique²¹. Cette orientation s'explique aussi bien par un souci de convergence avec la démarche menée depuis 2001 par la BnF autour de l'informatisation de son propre catalogue de manuscrits²² que par le souhait de la DLL de placer la consultation du CGM

21. Le projet est évalué à 400 000 €. La Sous-direction des bibliothèques et de la documentation, partenaire du projet, financera la part du CGM correspondant aux collections des bibliothèques universitaires (évaluée à 1/5 du CGM).

22. Dans le cadre du futur *Schéma directeur de l'informatisation des collections spécialisées*, réalisé par le Bureau Van Dijk en 2002.

dans le cadre du Catalogue collectif de France.

La structure de direction du projet est cependant maintenue et le comité opérationnel, assisté d'un consultant, va préparer la rédaction du cahier des charges en vue de l'appel d'offres. Un prestataire doit être choisi dans les mois à venir ; la conversion rétrospective s'étendra sur au moins deux ans. L'achèvement de cette étape, qui sera en soi d'une grande importance, ne marquera cependant pas la fin du projet. Ce sera simplement un tournant vers de nouveaux chantiers plus ambitieux auxquels le comité de pilotage réfléchit dès à présent, en particulier la réalisation d'un futur *Catalogue collectif national de manuscrits* permettant la saisie de nouvelles notices, la modification des notices existantes ou encore l'échange de données entre un système central et des systèmes intégrés de gestion de bibliothèque (SIGB) fonctionnant en Marc. L'architecture de ce futur catalogue demeure, bien entendu, à déterminer.

D'autres projets autour du manuscrit

Il a enfin semblé à la Direction du livre et de la lecture que cette opération était l'occasion de lancer

d'autres actions autour du manuscrit. Rappelons que, depuis 2001, un cycle de formation à la description des manuscrits modernes et contemporains

Le succès de ce cycle et le souci manifeste de nombreux collègues de mieux se former au traitement de ces collections ont conduit la DLL à proposer une réflexion sur la création d'une norme Afnor de description des manuscrits modernes et contemporains

a été mis en place dans le cadre de la formation continue de l'École nationale supérieure de l'information et des bibliothèques (Enssib).

Le succès de ce cycle et le souci manifeste de nombreux collègues de mieux se former au traitement de ces

collections ont conduit la DLL à proposer une réflexion sur la création d'une norme Afnor de description des manuscrits modernes et contemporains. La recommandation actuelle, élaborée par la DLL en 1980, apparaît aujourd'hui un peu succincte et ne bénéficie pas de la reconnaissance apportée par l'Afnor. Un groupe de travail a donc été constitué dans le cadre de la Commission générale 46 de l'Afnor et prépare une étude préalable sur les enjeux et les orientations possibles de cette future norme²³.

Comme on le voit, plusieurs choix importants ont désormais été faits et des pistes mieux « balisées » s'offrent à nous... Il convient cependant de ne pas mésestimer tout le chemin qui reste à parcourir et les difficultés qui ne manqueront pas. Le soutien de nombreux collègues et les exemples de réussites similaires à l'étranger constituent pourtant de bons atouts pour le succès de l'entreprise.

Juin 2003

²³. Ce groupe de travail est placé auprès de la Commission de normalisation chargée des questions de modélisation, production et accès aux documents.