

Numérisation et patrimoine documentaire

Les opérations de numérisation des collections conservées par les bibliothèques ont pu être considérées, il y a quelques années, comme l'apanage des grands établissements. Il n'en est plus de même aujourd'hui, puisque les techniques sont à présent mieux rodées, les expériences mieux partagées, les articles et les ouvrages sur le sujet plus nombreux.

Gaëlle Béquet

Bibliothèque
de l'Institut international
d'administration publique

g.bequet@iiap.fr

Laure Cédelle

Direction du livre et de la lecture

laure.cedelle@culture.fr

La voie est à présent plus sûrement balisée pour une bibliothèque qui souhaiterait se lancer dans un tel projet. Néanmoins, ce dernier ne peut se réaliser sans un examen de ses modalités financières, une étude des choix techniques et une évaluation des projets en cours en France ainsi qu'à l'étranger. Le chef de projet, qui doit au préalable rassembler une documentation très éparse, trouvera dans cet article les premiers éléments nécessaires à sa réflexion.

Le financement de la numérisation

Il existe plusieurs programmes d'aide à la numérisation et l'on distinguera ici ceux qui sont dispensés au niveau central de ceux qui sont gérés sur le plan régional.

Le ministère de la Culture et de la Communication a mis en place depuis quatre ans un mécanisme de prise en charge à 100 % des frais techniques de numérisation des collections d'État, s'inscrivant dans le cadre d'un marché national. En 1996, 2,5 millions de francs ont été consacrés à cette opération, 2,65 en 1997, 4 en 1998 et 5 en 1999. Cette aide, réservée dans un premier temps à la

numérisation des documents de substitution - les microfilms ou les ekta-chromes, par exemple -, a été élargie depuis l'an dernier à la numérisation directe de documents aussi divers que les livres, les manuscrits, les cartes et plans, les estampes, les photographies et les documents sonores. Une aide complémentaire peut éventuellement financer l'inventaire préalable des fonds.

Chaque année, la Mission de la recherche et de la technologie lance plusieurs appels à projets. Les dossiers sont transmis aux Directions régionales des affaires culturelles, chargées d'informer les services de l'archéologie, de l'inventaire, des monuments historiques, les musées, les archives départementales et les bibliothèques. Les candidats doivent remplir un dossier comportant notamment des questions précises sur le personnel scientifique affecté à l'identification et à la préparation des fonds à numériser, ainsi qu'à leur traitement intellectuel. L'institution doit aussi disposer d'un matériel informatique de qualité suffisante et de logiciels adaptés pour récupérer les images numérisées livrées sur cédéroms, pour les contrôler et enfin pour les indexer et les mettre à disposition du public.

Jusqu'à présent, le programme ne concernait que les fonds d'État, mais, cette année, les fonds des collectivités locales pourront aussi bénéficier d'une aide qui atteindra jusqu'à 50 % du coût de la numérisation. Ces opérations ne seront aidées que si elles peuvent être menées dans un délai assez court avec une mise à disposition rapide des fonds traités.

En 2000, le ministère de la Culture et de la Communication consacra 12 millions de francs à la numérisation du patrimoine.

Depuis 1996, six bibliothèques municipales ont bénéficié de cette aide : Amiens et Troyes pour des manuscrits médiévaux, Valenciennes pour des incunables, Pau pour des documents iconographiques sur l'histoire du Béarn, Metz pour des atlas des XVII^e et XVIII^e siècles, Saint-Étienne pour des plans, dessins et cours relatifs à l'art des mines.

Au niveau régional, les bibliothèques municipales peuvent avoir recours à la deuxième part du concours particulier des bibliothèques de la dotation générale de décentralisation gérée par les Directions régionales des affaires culturelles. La circulaire d'application datée de mars 1999 prévoit en effet soit le financement de l'installation d'un atelier de numérisation dans les locaux de la bibliothèque, soit celui d'une prestation de numérisation réalisée par une entreprise.

Le premier, qui peut atteindre 50 % du montant des dépenses hors taxes, s'inscrit dans le cadre d'opérations destinées à améliorer les conditions de préservation et de conservation des fonds anciens, rares et précieux. Il est conditionné par la

présence au sein de l'établissement d'un personnel qualifié ayant reçu une formation en imagerie numérique.

Le deuxième financement porte sur la numérisation par un prestataire extérieur et peut atteindre entre 20 et 40 % du montant hors taxes des devis établis pour la numérisation en mode image, le passage au mode

texte pour les tables des matières, sommaires ou index, la production de supports de substitution (photographies, microfilms), la conversion rétrospective des fichiers ou catalogues décrivant ces documents numérisés et le contrôle qualité. S'il est plutôt recommandé de numériser des documents libres de droits, la numérisation de documents encore

protégés au titre de la législation sur la propriété intellectuelle n'est pas exclue. Il revient cependant à la commune de prendre en charge l'acquisition des droits afférents à l'exploitation numérique des documents par la bibliothèque. Outre le concours particulier, les Directions régionales des affaires culturelles disposent de crédits d'intervention propres susceptibles de financer quelques projets.

Les recommandations du ministère de la Culture et de la Communication

Depuis l'ouverture des appels à projet de numérisation, le ministère de la Culture et de la Communication s'est efforcé d'assurer la cohérence des différents programmes qu'il soutenait. À cet effet, une série de règles et de recommandations a été transmise aux bénéficiaires du plan de

numérisation du ministère, afin de garantir une homogénéité des documents numérisés et de faciliter la gestion ultérieure de ces fonds. Ces recommandations d'ordre juridique et technique s'adressent plus largement à l'ensemble des bibliothèques publiques auxquelles elles peuvent apporter des réponses dans leur choix de numérisation. Il est possible de les consulter sur le site du ministère de la Culture et de la Communication où la Mission recherche et technologies et la Direction du livre et de la lecture présentent des fiches d'information sur les différentes étapes de la numérisation¹.

Plutôt que de dresser un inventaire de ces recommandations qui peuvent être consultées en ligne, nous souhaiterions insister sur les éléments qu'il est essentiel de prendre en compte avant toute opération de numérisation.

En premier lieu, la question des droits de reproduction et de diffusion se pose avec une acuité toute particulière dans le cas des documents numérisés : il ne suffit pas de posséder un document dans sa bibliothèque pour pouvoir le numériser et le diffuser, même en intranet. Avant de procéder à la numérisation, il est impératif de s'assurer que la bibliothèque dispose de tous les droits d'exploitation des images. Le statut du document, au regard de la propriété littéraire et artistique, est le premier élément à apprécier. Seuls les documents tombés dans le domaine public sont libres de droit, les autres demeurent protégés au titre du droit d'auteur ; leur reproduction ainsi que leur diffusion sont à négocier auprès des ayants droit. D'une manière générale, les œuvres sont protégées du vivant de l'auteur et pendant une période de 70 ans à compter du 1^{er} janvier qui suit la

Avant de procéder à la numérisation de documents, il est impératif de s'assurer que la bibliothèque dispose de tous les droits d'exploitation des images

1. <http://www.culture.fr/culture/mrt/numerisation>

mort de cet auteur; cette période peut être allongée par les prorogations pour années de guerre. Dans le cas d'une œuvre de collaboration, où la part de chacun des auteurs est clairement identifiable, l'œuvre est encore protégée pendant 70 ans à compter de la mort du dernier collaborateur.

**Toute cession
de droits doit
faire l'objet
d'un contrat
précisant
les conditions
d'exploitation
des œuvres numérisées**

Les bibliothèques doivent rester vigilantes sur le respect des droits de reproduction et de représentation attachés aux documents contemporains. Cette remarque concerne aussi bien les livres que les photographies, les cartes postales, les cartes géographiques, les gravures ou même les périodiques. Ce dernier cas demeure particulièrement complexe, puisque tous les collaborateurs du journal (journalistes, photographes) représentent autant d'ayants droit. La bibliothèque doit donc vérifier si les articles et illustrations du périodique sont bien libres de droits, ou si la société de presse a négocié avec ses collaborateurs une cession des droits pour tout type de diffusion. Dans le cas où la bibliothèque serait dans l'obligation d'acquérir les droits nécessaires à leur numérisation, toute cession de droits doit faire l'objet d'un contrat précisant les conditions d'exploitation des œuvres numérisées. Les différentes aides du ministère de la Culture et de la Communication pour la numérisation des collections excluent l'acquisition des droits de propriété

intellectuelle, ces négociations demeurant de ce fait à la charge des établissements².

Une fois réglées les questions juridiques qui garantiront l'exploitation des documents par la bibliothèque, certains aspects techniques nécessitent une attention tout aussi grande si l'on veut garantir une bonne exploitation informatique des fichiers numériques créés. Sur ce point, les recommandations du ministère insistent particulièrement sur le format et le nom des fichiers numériques. La conservation des documents numériques et leur capacité à être exploités par différents systèmes sont largement tributaires de ces choix. Pour les seules images bitmap, on recense plus de 70 formats de fichiers. Il est fortement déconseillé d'adopter un format propriétaire de peur de ne pouvoir communiquer les fichiers en dehors d'un cercle restreint d'utilisateurs.

Le plan de numérisation du ministère de la Culture et de la Communication a imposé deux formats pour la numérisation en mode image : TIFF (Tagged Image File Format), format propriétaire mais devenu un standard de fait, et JPEG (Joint Photographic Expert Group), norme adoptée par l'ISO sous le numéro 10918-1 (1994). Ces deux formats, adaptés aussi bien aux modes couleur que noir et blanc ou niveaux de gris, offrent deux possibilités d'utilisation distinctes. Le TIFF représente un bon format de conservation des documents dans leur version non compressée ou faiblement compressée. Le JPEG trouve davantage son utilisation pour une exploitation des images en réseau ou sur l'Internet. L'usage de ces deux for-

mats est devenu suffisamment répandu pour faciliter les échanges de fichiers et prévoir leur conversion ultérieure sous d'autres formats plus performants.

Le dernier aspect sur lequel il faut insister concerne la structuration des noms de fichiers. Dans le cadre du plan de numérisation, des structures de nom de fichiers ont été conçues pour garantir leur cohérence, faciliter le repérage des documents et leur assurer une plus longue durée d'utilisation. Chaque image numérique créée reçoit effectivement un nom spécifique, unique, qui ne pourra être réutilisé pour une autre image. Plutôt que de recourir à des numéros purement séquentiels, il a été choisi d'utiliser des noms significatifs, constitués de caractères alphanumériques auxquels s'ajoute le « _ »; tous les signes diacritiques sont exclus. La structure du nom d'un fichier se compose de

**Il est fortement
déconseillé
d'adopter
un format
propriétaire
de peur
de ne pouvoir
communiquer
les fichiers
en dehors
d'un cercle restreint
d'utilisateurs**

deux parties : l'identifiant de la bibliothèque détentrice du document et l'identifiant du document numérisé. Le premier identifiant se compose du sigle des bibliothèques « B », suivi du code RBCCN délivré par les agences du Catalogue collectif national-Publications en Série (CCN-PS). Le deuxième élément décrit le docu-

2. Pour plus de précisions sur les questions juridiques (propriété intellectuelle, droit moral, droit des bibliothèques sur leurs documents numérisés...), on pourra se reporter à la fiche juridique présentée sur le site précédemment cité (note 1) dans les fiches rédigées par la DLL. On pourra aussi se reporter au *Code de la propriété intellectuelle*.

ment numérisé en le situant dans les collections de la bibliothèque, à l'intérieur d'un recueil ou d'un livre, voire au sein d'une page s'il s'agit d'un détail, la dernière information donnée indique le type de fichier (vignette, tiers d'écran, fichier de conservation). Ainsi le code d'immatriculation de l'image comprendra les éléments suivants :

- code ou numéro du fonds
- cote du document
- numéro du volume et numéro de page
- type de vue (pleine page, double page, détail)
- type de fichier (affichage vignette, tiers d'écran, plein écran, fichier de conservation).

Toutes ces informations trouvent leur intérêt quand une bibliothèque s'engage dans plusieurs programmes de numérisation de ses collections et qu'elle a besoin d'identifier avec certitude ses fichiers numériques pour les stocker de façon distincte, comme elle le ferait pour des documents traditionnels³.

Expertise, recommandations et plans de numérisation

La Commission européenne, à travers ses différents programmes (Télématique pour les bibliothèques de la Direction générale XIII jusqu'en 1998, INFO 2000 et Technologies de la société de l'information⁴ depuis 1999), a catalysé les initiatives dans le domaine de la numérisation en associant les secteurs recherche et développement de grandes entreprises et

des bibliothèques européennes. Le programme de travail Technologies de la société de l'information (ISTI) comporte une action-clé « Contenus et outils multimédias » (le premier appel en mars 1999 a recueilli 530 propositions dans ce domaine), dont l'un des volets est consacré à l'accès aux collections numériques de contenus culturels et scientifiques. Malheureusement, les résultats de cet appel ne sont pas encore officiels car la Commission européenne continue de négocier les contrats⁵.

Les projets aidés sont tenus de diffuser des rapports d'activité, qui peuvent ensuite servir de ressources à la communauté bibliothéconomique tout entière. Ils sont fréquemment consultables en ligne, et le site de la Direction générale XIII y renvoie⁶. Cette dernière continue par ailleurs d'alimenter la réflexion sur les synergies entre archives, bibliothèques et musées, perçues sous l'angle de la mise en commun des ressources *via* la numérisation.

La Direction générale X intervient, elle aussi, dans ce domaine en finançant des projets tels que SEPIA⁷, programme de recherche sur la numérisation et la conservation des photographies auquel participe le Centre de recherches sur la conservation des documents graphiques (CRCDG). Ce projet reçoit également le soutien de l'European Commission

on Preservation and Access (ECPA)⁸, autre instance qui s'intéresse à la question de la numérisation comme méthode de conservation.

Si l'on souhaite avoir une vision internationale des programmes de numérisation, il convient aussi d'évoquer le serveur⁹ de l'IFLA qui propose un ensemble de références consacrées aux bibliothèques numériques, parmi lesquelles des revues

en ligne comme *D-Lib*¹⁰, qui paraît mensuellement et permet de suivre l'actualité dans le domaine des documents numériques. C'est dans les « colonnes » de cette revue qu'a été évoquée très récemment l'étude¹¹ sur la conservation des documents numériques financée en mars 1999 par l'International Council for Scientific and Technical Information (ICSTI), à partir de l'évaluation de 18 projets, dont PANDORA¹², mené par la Bibliothèque nationale australienne, et EVA, mené par la Bibliothèque nationale finlandaise, projets qui visent tous deux à conserver des sites Internet. L'enquête a permis de montrer l'importance de la création de métadonnées sur les documents conservés, de la migration d'un mode de stockage à un autre (tous les 4 à 5 ans environ), de l'utilisation de formats de fichiers universellement reconnus (TIFF, HTML, SGML, PDF). Sur le plan européen, le projet NED-

Les projets aidés
sont tenus
de diffuser
des rapports d'activité,
qui peuvent
ensuite servir
de ressources
à la communauté
bibliothéconomique
tout entière

3. On trouvera un exemple de nom de fichier image détaillé dans la fiche « caractéristiques du fichier numérique » présentée sur le site du ministère.

4. Pour obtenir des informations sur les appels à projets « Technologies de la société de l'information », vous pouvez vous inscrire sur la liste de diffusion du Centre français du commerce extérieur, en contactant Christine Montagut (christine.montagut@cfce.fr).

5. Le site Evariste de l'ANVAR propose un forum sur la recherche et l'innovation en Europe (<http://www.evariste.anvar.fr/actu/schouller/>).

6. <http://www.echo.lu/digicult>

7. <http://www.knaw.nl/ecpa/sepia/>

8. <http://www.knaw.nl/ecpa/ecpatex/index.htm>

9. <http://www.ifla.org/ll/digilib.htm>

10. <http://www.dlib.org>

11. Gail M. Hodge, « Best practices for digital archiving : an information life cycle approach », *D-Lib Magazine*, vol. 6, n° 1, janvier 2000.

12. Cf. dans ce numéro l'article de Catherine Lupovici, p. 43-54.

LIB¹³, financé par la DG XIII et piloté par la Bibliothèque royale des Pays-Bas à La Haye, examine lui aussi les questions de conservation des documents électroniques (cédéroms et documents en ligne).

La numérisation dans les bibliothèques publiques françaises

En France, la numérisation des collections suscite certes un intérêt grandissant qui se manifeste par l'émergence de quelques initiatives, mais elle n'est pas encore pour autant une pratique courante. Malgré les formations et journées professionnelles organisées par la Bibliothèque nationale de France, l'École nationale supérieure des sciences de l'information et des bibliothèques ou les Unités régionales de formation à l'information scientifique et technique, les bibliothécaires sont encore hésitants à lancer des opérations qui demeurent assez lourdes tant par leur poids financier que par l'investissement en personnel qu'elles requièrent. Beaucoup craignent de numériser des documents qui l'ont déjà été ailleurs et il devient indispensable de diffuser une plus large information sur les projets initiés dans les bibliothèques publiques.

La Direction du livre et de la lecture a adressé en décembre dernier un questionnaire à des bibliothèques municipales afin de repérer les projets de numérisation. Ce questionnaire distingue les réalisations effectives des programmes en cours et ceux qui sont encore à l'étude. Il se donne également pour objectif d'identifier les techniques employées par les établissements. Les résultats de cette

enquête alimenteront une base de données qui sera hébergée sur le site du ministère de la Culture et de la Communication l'an prochain. Cet outil doit permettre aux bibliothèques de repérer les collections numérisées et les encourager à coordonner leurs politiques de numérisation.

Cette première enquête, menée à titre expérimental, a été adressée à cent bibliothèques municipales dont la sélection reposait sur trois sources d'information : l'enquête de 1998 sur l'équipement informatique des bibliothèques municipales révélait que dix-huit bibliothèques numérisaient des documents ; d'autres établissements avaient contacté directement la Direction du livre et de la lecture pour s'informer sur les aides disponibles ; enfin, le questionnaire a été envoyé à tous les membres de l'Association pour le développement des documents numériques en bibliothèque (ADDNB), où venait d'être créé un groupe de travail sur la numérisation. Les questionnaires n'ont pas tous été retournés, mais 51 bibliothèques ont déjà répondu¹⁴, ce qui rend possible une première analyse.

Parmi les bibliothèques ayant répondu à l'enquête, 53 % déclarent avoir un projet de numérisation¹⁵.

Tous ces projets ne sont pas cependant au même stade d'avancement. Si 23 % des bibliothèques interrogées proposent déjà des documents numérisés, 21 % n'ont pas encore achevé leur programme et 37 % étudient la possibilité de lancer un projet de numérisation¹⁶. Une analyse plus détaillée des réponses montre que les bibliothèques ayant déjà expérimenté la numérisation poursuivent le

plus souvent le traitement d'autres collections. Nous avons dénombré 16 programmes aboutis, 22 en cours et 40 à l'étude¹⁷. Les bibliothèques pionnières s'engagent donc davantage dans la voie du numérique, n'hésitant pas parfois à mener de front plusieurs projets.

Les documents anciens sont numérisés de façon privilégiée. Il s'agit, dans 50 % des cas,

de documents des XVI^e-XIX^e siècles, peu visibles du grand public, car souvent conservés dans les réserves des fonds anciens. Les collections contemporaines sont néanmoins très sollicitées (38 % des cas), à l'inverse des incunables et manuscrits médiévaux qui n'apparaissent que dans 10 % des projets. Plus qu'un manque d'intérêt pour les documents médiévaux, la faible représentation de ce

Les bibliothécaires sont encore hésitants à lancer des opérations qui demeurent assez lourdes tant par leur poids financier que par l'investissement en personnel qu'elles requièrent

13. Titia Van der Werf-Davelaar, « Long-term preservation of electronic publications : the NEDLIB project », *D-Lib Magazine*, vol. 5, n° 9, sept. 1999. Cf. également dans ce numéro l'article de Catherine Lupovici, p. 43-54.

14. Les chiffres qui suivent sont basés sur le dépouillement des réponses reçues au 27 février 2000.

15. Plusieurs bibliothèques, dont nous savons qu'elles ont mené des programmes de numérisation, ne nous ont pas encore répondu. Il faut en fait compter sur un taux de plus de 60 % à terme.

16. Parmi ces bibliothèques, beaucoup élaborent de nouveaux projets alors qu'elles ont déjà une réalisation à leur actif. En fait, seuls 17,6 % des BM ayant répondu en sont au premier stade de la réflexion sans expérience antérieure.

17. Il faut entendre par programme, la numérisation d'un fonds entier, par exemple le fonds des manuscrits de Citeaux, ou la numérisation de différents documents regroupés autour d'une thématique comme les arts de la mine à Saint-Étienne, ou les expositions virtuelles de la BM de Lyon.

Collections traitées	
Typologie des documents	Représentation dans les programmes de numérisation
Monographies	18 %
Manuscrits	16 %
Cartes et plans	13 %
Estampes	10 %
Photographies	10 %
Périodiques	9 %
Cartes postales	8,2 %
Dessins et aquarelles	8,2 %
Divers (affiches, tapuscrits, bois gravés)	6 %
Partitions	1,6 %

type de fonds peut s'expliquer par les lacunes du catalogage. Plus aisé, le traitement bibliographique des documents du XX^e siècle facilite l'accès ultérieur de leur version numérisée.

Si l'on dépasse cette distinction entre fonds ancien et fonds contemporain pour s'attacher à la nature des documents, on remarque que les collections iconographiques (49 % des collections numérisées) ont jusqu'ici été privilégiées par rapport aux documents écrits (43 %). Des collections

plus spécifiques telles que les partitions, les tapuscrits ou les bois gravés trouvent dans la numérisation un mode de diffusion et de consultation plus adapté. Le tableau présenté ci-contre donne un premier aperçu des collections traitées.

Enfin, il faut mentionner que la Bibliothèque nationale de France, à travers les pôles associés, prend également en charge la numérisation de certains documents qui viennent compléter ses collections. Les bulletins de sociétés savantes conservés par les archives départementales et certaines bibliothèques d'Aquitaine et de Lorraine sont numérisés dans le cadre de conventions passées avec les conseils généraux et les municipalités. Les partenaires locaux financent le microfilmage des

revues que la BnF fait numériser. À l'issue de ces opérations, les collectivités récupèrent des cédéroms contenant l'ensemble des collections de la région.

Conclusion

La mise en place d'un projet de numérisation est conditionnée par des facteurs financiers, techniques, juridiques et patrimoniaux que tout

maître d'ouvrage doit prendre en compte. Il est donc essentiel de suivre l'actualité dans ces domaines. Des possibilités d'aide financière sont proposées par le ministère de la Culture et de la Communication. Les bibliothèques trouveront auprès des conseillers pour le livre et la lecture les dates des appels à projets, ainsi que des renseignements sur le concours particulier des bibliothèques de la dotation générale de décentralisation.

Le ministère offre une aide en ligne à travers son site d'information, où conseils techniques et recensement des actions soutenues permettent aux bibliothèques de situer leur projet dans un ensemble plus vaste et de s'orienter dans leurs premiers choix. Ce site sera complété d'ici l'an prochain par une base qui recensera les projets dans les différents secteurs de compétence du ministère (archives, bibliothèques, monuments historiques, musées, etc.).

Enfin, il ne faut pas négliger les programmes étrangers qui suscitent une littérature professionnelle abondante. La consultation régulière de sites comme celui de l'IFLA, de la Direction générale XIII de la Commission européenne, ou encore de la revue *D-Lib* offre un panorama intéressant des recherches menées dans le domaine de la numérisation.

Mars 2000