

# La reproduction de sauvegarde des documents patrimoniaux

Partie 1 – Numérisation et  
conservation

---

Dominique MAILLET

Experte numérisation  
Bibliothèque nationale de France  
Département de la conservation

Depuis le début du XXI<sup>e</sup> siècle, la numérisation s'est progressivement substituée à la microforme argentique (microfilm ou microfiche) pour la reproduction des documents écrits et graphiques.

## Introduction

---

Les immenses possibilités offertes par cette technique en progrès constants, en lien avec l'expansion d'Internet, font que son usage est beaucoup plus étendu, largement axé sur la diffusion et la valorisation. Elle demeure toutefois un outil capital et fondamental pour la conservation des documents, à condition de respecter des règles précises dans sa mise en œuvre et dans le stockage des fichiers réalisés. La présente fiche vise à préciser ces règles dans ses deux premières parties. Une troisième partie reviendra brièvement sur le microfilm qui, s'il n'est plus guère produit pour les établissements culturels français, a permis de constituer des filmothèques de conservation et peut être utilisé en connexion avec la numérisation pour améliorer la pérennité de cette dernière technique.

# 1. Quelle numérisation pour quel(s) objectif(s) ?

En plein essor depuis près de vingt ans, la numérisation s'est progressivement substituée à la micrographie comme technique de reproduction des documents écrits et graphiques dans les bibliothèques et les archives. Les multiples possibilités qu'elle offre font qu'on l'emploie pour des objectifs divers. La première question à se poser pour la constitution d'une bibliothèque numérique est donc de définir pour quel (s) usage (s) et à quel public elle est destinée.

- La conservation du document original est l'un des premiers objectifs. Le document original n'est plus proposé au public (ou a minima pour des consultations spécifiques) et par conséquent, s'il bénéficie dans son magasin de conditions de conservation physique adéquates, il se conservera mieux.

- La diffusion : L'évolution croissante des technologies informatiques et les puissances de débit réseau ainsi que l'évolution des capacités de stockage permettent de proposer des modes de diffusion démultipliés, sur l'intranet d'une institution ou plus largement sur internet. Les possibilités vont d'une communication par simple substitution électronique d'un document original à des organisations plus abouties de corpus constitués sur des thématiques données. La diffusion d'une version numérisée permet dans le cas de documents de grand format (affiches, cartes, estampes) ou particuliers (maquettes en volumes, plaques de verre) ou encore en mauvais état (papier acide, encre délavée, etc.) de donner accès à des contenus qui ne seraient pas accessibles à un public via sa forme originale : on rejoint donc là les objectifs de la conservation, cette fois curative et non plus préventive. Sans parler de la possibilité de démultiplication d'accès pour un même objet numérique sans recourir à l'acquisition de plusieurs exemplaires originaux pour satisfaire la demande du public.

Ces objectifs ne sont donc pas, et loin de là, exclusifs les uns des autres. Pour être atteints, ils impliquent de bien connaître ses collections, de bien connaître son public et de bien définir les usages.

On notera cependant que ces objectifs complémentaires peuvent être parfois difficiles à concilier et que la politique documentaire peut être différente ou compliquée pour y parvenir : ce ne sont pas forcément les mêmes documents qui nécessitent d'être numérisés à des fins de conservation ou à des fins de diffusion. Mais même la numérisation dont l'objectif premier est la diffusion peut et doit avoir l'effet d'améliorer la conservation des documents originaux.

# 2. La numérisation : quelques préalables importants à connaître

Les objectifs prioritaires de la numérisation entraînent des choix de numérisation et donc des objets numériques différents essentiellement sur 3 aspects : la fidélité à l'original ou non, les règles de prises de vue, le choix des formats de stockage.

- L'original : S'il est relativement simple de définir ce qu'est un original lorsque l'on parle de monographies, de manuscrits ou de partitions de musique, la question devient plus compliquée lorsque l'on a entre les mains une série de photographies organisées en album dans lequel elles ont été placées dans un ordre précis et annotées d'une légende, tout cela effectué non par l'auteur des photographies mais par le collectionneur ou un conservateur. Quel original va-t-on choisir : l'album ou le lot de photographies ? Il en est de même pour une série de correspondances montées sur onglet dans des volumes reliés : la même question se pose sur la définition de ce que sera l'original.

- La question de la fidélité à l'original : cette question est très importante lorsque l'on parle de numérisation à des fins de conservation ; elle devient cruciale surtout dans le domaine des collections 'spécialisées' à une époque où la couleur est présente partout et où le public a désormais une exigence de plus en plus grande sur le niveau de qualité auquel il s'attend. C'est un exercice compliqué voire impossible de chercher à faire des images esthétiques qui plaisent au public tout en respectant l'original.

La fidélité à l'original est une notion toute relative.

D'ailleurs, de quel original parle-t-on ?

Celui que l'on voit objectivement lorsque le document est placé sur le berceau du numériseur ?

Ou celui qu'il était, ou que l'on suppose qu'il était, lorsqu'il a été créé il y a des années, des dizaines d'années voire des siècles ?

Tenter de recréer ce que l'on pense qu'il était relève de la plus totale subjectivité du photographe qui opère, et il est plus sage de s'en tenir à ce que nous observons au moment où le document va être traité.

Bien sûr, il faut garder en tête que cette fidélité à l'original est une notion très volatile car si le document avait été numérisé 10 ans plus tôt, ou dans 50 ans, le résultat serait certainement différent.

- La question des formats : on n'utilise pas le même format de numérisation si l'on vise d'abord la conservation ou la diffusion.

Un document numérisé à des fins de conservation devient un master à conserver dans des formats dits de conservation, c'est-à-dire des formats ouverts, indépendants, facilement interprétables et garantissant zéro compression de données, ou une compression minimale qui n'affecte pas les images produites. A contrario, un document destiné d'abord à être diffusé pourra, selon l'usage et les modalités de la diffusion, être encapsulé dans des formats de fichiers différents à des taux de compression différents, plus élevés que pour la conservation, l'avantage étant un poids de fichier moindre et donc un coût de conservation moins élevé (cf. partie II, paragraphe 3.1).

C'est ainsi que, autant que possible, les bibliothèques numériques doivent faire l'objet de deux organisations de stockage différentes : un espace dédié aux documents archivés à des fins de conservation et un espace dédié à la consultation de la bibliothèque numérique, la seule différence entre ces deux espaces étant le format de fichier, davantage compressé pour l'espace de diffusion.

Dans certains cas, les documents stockés dans l'espace de consultation peuvent faire l'objet de corrections chromatiques destinées à « embellir » l'original en saturant par exemple les couleurs ternies par le temps ou en améliorant le contraste des images un peu fades. Mais les documents archivés à des fins de conservation doivent rester aussi fidèles que possible aux originaux.

### 3. La numérisation : les paramètres techniques qui qualifient une image et les choix à faire

Un paragraphe technique sur les paramètres de la numérisation est ici nécessaire, car bien les choisir est un gage de qualité de ce qui sera produit, et d'adéquation avec le ou les objectifs fixés.

#### La résolution et la définition

Ce sont 2 termes importants qui permettent de définir la qualité de numérisation et la finesse d'analyse.

La résolution est une unité de mesure : on la mesure en points par pouces (ppp) ou plus souvent en dot per inch (dpi). Elle indique le nombre de points (ou pixels) analysés dans un carré de 2,54 cm de côté.

Cette information seule n'est pas suffisante pour savoir si la qualité de numérisation est correcte.

On entend majoritairement parler de résolutions à 300, 400 ou 600 dpi.

300 dpi est une résolution historiquement utilisée par les fax et les photocopieurs, ou encore les imprimantes au format A4.

Ce choix de 300 dpi n'a bien sûr pas été fait au hasard et il découle de l'analyse des qualités optiques de l'œil, dont celle qui nous intéresse sur ce sujet peut se résumer comme suit : l'œil possède un pouvoir séparateur qui lui permet à une distance d'observation donnée de distinguer 2 points proches l'un de l'autre. Il est universellement observé qu'à 2,5 mètres, un œil sain distingue 2 points de 1 mm séparés de 1 mm. Au-delà, les 2 points se confondent. C'est en extrapolant cette donnée physique que l'on peut affirmer qu'à 300 dpi on obtient une image numérisée que l'œil humain sait parfaitement lire et qui contient suffisamment d'informations pour que le produit numérisé soit estimé conforme à l'original.

Numériser à 400 dpi, soit 400 X 400 points analysés sur un carré de 2,54 cm de côté et non plus 300 revient donc à capturer des pixels qui ne seront pas utiles à la lecture car notre œil n'a pas les moyens physiques de les distinguer. Le seul avantage est de pouvoir zoomer dans le document comme si nous le lisions avec une loupe.

Partant de ces principes physiques et de l'expérience, on peut affirmer qu'une numérisation à 300 ou 400 dpi est suffisante pour la quasi-totalité des documents.

Les seules exceptions sont les suivantes :

- les documents de grand format conçus pour être regardés à des distances plus importantes qu'une lecture traditionnelle (affiches), qui nécessitent une numérisation autour de 100 dpi ;
- les documents de petit format (inférieur au A6) qui doivent être numérisés à 600, voire 800 ou même 1200 dpi pour obtenir une restitution précise des détails du document.

<p><b>En résumé : résolutions conseillées</b></p> <ul style="list-style-type: none"> <li>- 300 ou 400 dpi pour la majorité des documents</li> <li>- 600 dpi ou plus pour les documents de taille inférieure au format A6</li> <li>- 100 ou 200 dpi pour les documents de grand format type affiche</li> </ul>
---

Il faut garder en tête, pour résister aux exigences de plus en plus fortes et souvent inutiles des éditeurs ou imprimeurs, que si une image numérisée permet une impression de qualité sur un livre d'art format standard A4 ou supérieur, alors sa résolution suffit pour imprimer une affiche de 4 mètres sur 3 ou une bache de chantier !

La définition : Contrairement à la résolution qui est une unité de mesure, la définition est propre au document. Elle indique le nombre de pixels que contient le document. On indique la définition en donnant le nombre de pixels analysés dans la hauteur du document multiplié par le nombre de pixels analysés dans la largeur du document. La définition permet de connaître la finesse de numérisation et de savoir quel usage on peut faire du document (taille d'impression, capacité de zoom à l'écran, etc.)

Taille du document, résolution, définition sont les 3 informations nécessaires pour qualifier l'image numérisée. Et chacune des 3 se déduit des 2 autres.

## Profondeur de codage

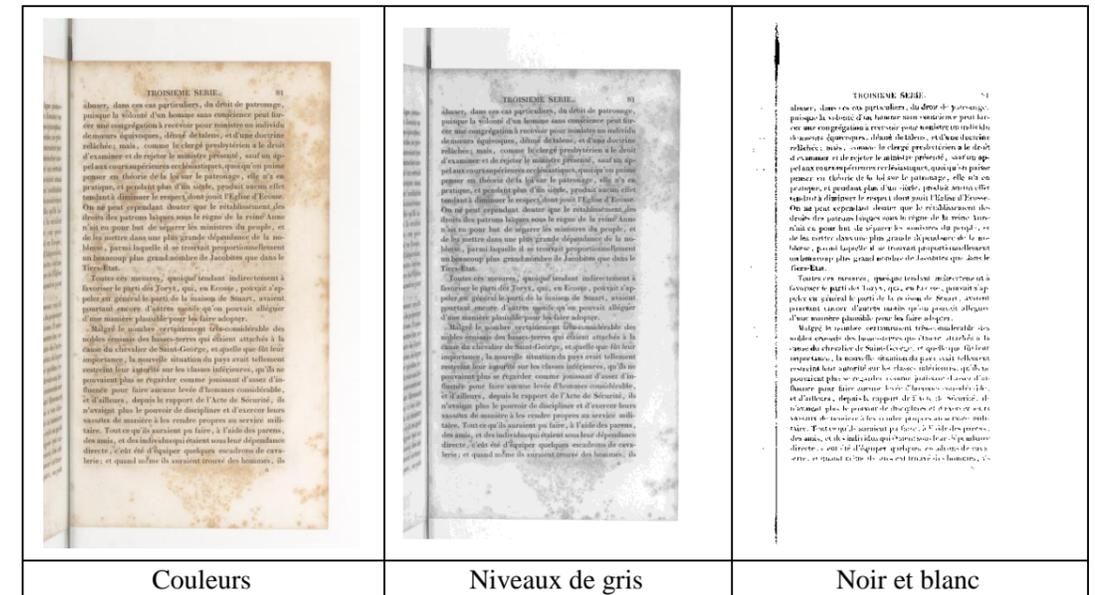
Appelée également profondeur de couleur, elle reflète et exprime le choix du type de numérisation qui est fait :

- numérisation en noir et blanc : codage sur un bit ;
- numérisation en niveaux de gris : codage sur 8 bits ;
- numérisation en couleur : codage de chaque canal de la couleur (rouge-vert-bleu) sur 8 bits, soit 24 bits en tout qui permettent de coder 16 millions de couleurs.

La question ne se posait pas lors de l'apparition de la numérisation, car la plupart des numériseurs étaient conçus pour une capture en noir et blanc ou en niveaux de gris. Mais l'évolution très rapide des techniques de numérisation, et celle tout aussi rapide des capacités de stockage, font qu'à présent tous les numériseurs du marché produisent nativement des vues en couleur.

Pourquoi s'en priver alors ?

Effectivement, si l'on parle de numérisation à des fins de conservation, la couleur comme le montre le tableau de comparaison ci-dessus est le seul mode de codage qui permette une restitution fidèle du contenu du document mais aussi du support papier et des traces de son histoire (plis, taches, lacunes, déformations, etc.)



Mais il faut toujours garder en tête que ce choix à un coût puisque les volumes à stocker sont multipliés par 3 entre un fichier couleurs et Niveaux de gris et par 24 par rapport à un fichier Noir et blanc.

Si l'on peut exclure assez aisément la numérisation en Noir et Blanc qui décontextualise vraiment le texte du support, la question du choix Niveaux de gris et Couleur peut se poser si la contrainte du volume de stockage existe.

Il est donc nécessaire de se poser la question de l'intérêt qu'il y a à numériser en couleur une collection de monographies ne comportant que des textes dépourvus d'illustrations colorées, dans lesquelles les seules traces de couleur viennent du jaunissement du papier et (parfois) des tampons d'appartenance de la bibliothèque.

## La gestion des couleurs

Sujet complexe, la gestion de la couleur doit permettre de maîtriser au mieux la chaîne de numérisation. En d'autres termes, permettre qu'un document numérisé avec n'importe quel scanner soit conforme à l'original et que l'image produite puisse être restituée fidèlement sur n'importe quel écran de consultation ou par une impression.

Ce domaine est trop vaste et trop technique pour qu'on l'aborde ici dans le détail.

Il faut simplement garder en tête quelques notions importantes qui seront à creuser le moment venu :

- L'espace couleurs dans lequel s'expriment les couleurs d'un périphérique donné ;
- Les valeurs Lab du CIE lab qui sont massivement utilisées pour coder les couleurs ;
- La caractérisation d'un scanner, ou la calibration d'un écran ou d'une imprimante qui permettent de définir l'espace couleur propre à ce périphérique et de définir ses défauts (écarts de couleurs). Le résultat de cette caractérisation ou calibration donne un profil ICC, carte d'identité du périphérique ;
- Le profil Adobe 98 : crée en 1998 par un consortium de constructeurs, il définit le plus petit dénominateur commun des équipements de qualité. Sans être une norme, il est reconnu internationalement et utilisé massivement pour coder les images dans un espace couleur commun à tout le monde.
- Le profil SRGB : Créé par un consortium de constructeurs, il définit le plus petit dénominateur commun des équipements bureautiques. Il est utilisé massivement pour coder les images qui sont destinées à un usage bureautique ou à des applications web.

## Les formats des fichiers

On distingue :

- Le format de conservation ou format d'archivage. Ce sont des Masters et les 2 formats référence absolue dans ce domaine sont le TIFF et le JPEG2000
- Le format de consultation. Ce format se décline à partir du format de conservation. On peut, selon les usages, décliner plusieurs formats de consultation à des taux de compression différents. Les principaux formats de consultation sont JPEG2000, JPEG, PDF et PNG.

En résumé :

Les choix de la résolution et de la profondeur de codage (couleur ou niveaux de gris) ne sont pas neutres en matière d'espace de stockage et de dimensionnement des réseaux.

Ainsi, une page au format A4 numérisée en niveaux de gris à 300 dpi occupera 8,8 MO (mégaoctets) contre 15,4 MO à 400 dpi, soit le double, et en couleur la même page occupera 26,4 MO à 300 dpi et 46,4 MO en 400 dpi.

On comprend bien que les choix de départ sont primordiaux et lourds de conséquence puisque, selon les décisions prises, le poids informatique d'une page numérisée peut varier de 8,8 à 46,5 MO, soit d'un facteur 1 à 5.

Si l'on numérise en couleur, le profil ICC Adobe 98 est indispensable pour donner l'information qui permettra d'afficher des couleurs cohérentes.

Le choix du format de fichier de conservation est très important. Il peut permettre de réduire des volumétries de stockage trop importantes si les choix de compression sont opérés de manière avisée (cf. partie 2 La Préservation numérique).

## 4. La chaîne de numérisation

La numérisation s'inscrit dans une chaîne comportant 11 étapes. Cette technique, qui a remplacé la micrographie (microfilm et microfiche), est donc plus complexe que cette dernière qui se limite à des opérations plus élémentaires (cf. Partie III Micrographie et conservation au XXI<sup>e</sup> siècle)

Les causes de cette complexification sont doubles :

- Le résultat de la numérisation est un produit informatique qui est archivé sur des serveurs et transmis via des réseaux. La gestion de cet objet virtuel, démultipliable à l'infini contrairement à l'objet physique microfilm ou original papier, nécessite une organisation stricte sous peine de stocker un document sans espoir de le retrouver, ou au contraire de diffuser trop largement des documents sous droits.
- Alors que la traditionnelle caméra microfilm sur son banc de reproduction imposait le plus généralement une mise à plat des documents, l'évolution des techniques a permis le développement de nombreux modèles de numériseurs, souvent complémentaires, équipés de porte-documents plus perfectionnés qui permettent une prise de vue adaptée à presque tous les types de documents présents dans les collections de bibliothèques et d'archives, à plat ou en ouverture restreinte selon les besoins. Il s'agit seulement de s'équiper des matériels adaptés aux collections à traiter et d'orienter les documents vers les numériseurs adéquats.

On peut décliner comme suit les étapes de la chaîne :

- a. Catalogage : sans catalogage à minima, il n'y a pas de numérisation possible. En effet, s'il est possible de se promener dans les magasins d'une bibliothèque et prélever des documents sans l'aide de notices, une fois le document numérisé et stocké sur des disques informatiques sans une notice minimum qui l'identifie, il n'y a aucune possibilité d'y accéder.
- b. Droits : attribution des droits qui vont permettre de définir les règles de diffusion.
- c. Expertise des documents : Elle doit permettre d'identifier la filière la plus adaptée à un traitement efficace des documents. Il peut s'agir d'un choix de numériseur mais également d'une filière (atelier interne à l'institution, marché de prestations installé en interne, externalisation) choisie en fonction de la fragilité d'un document, de sa préciosité, etc.
- d. Préparation avant numérisation : Elle regroupe toutes les interventions physiques éventuelles de restauration, maintenance, consolidation qu'il convient de mettre en œuvre pour que le document soit manipulable lors de la prise de vue et qu'on obtienne l'image la meilleure et la plus complète possible. On peut citer aussi d'autres types fréquents d'intervention comme le dépoussiérage, le désempoissage de

fascicules de presse reliés pour faciliter la numérisation, le doublage total ou partiel de feuillets trop fragiles, etc.

e. Numérisation : La prise de vue doit s'effectuer conformément aux paramètres techniques de numérisation qui ont été définis au démarrage du projet (résolution, profondeur de codage, formats de fichiers, etc.) mais également selon les consignes de numérisation établies en même temps (couverture complète ou partielle, mode de cadrage, numérisation en une ou 2 pages/image, sens de stockage des images, utilisation ou non d'un cache, etc.).

Pour sa part, la BnF a élaboré plusieurs référentiels de numérisation, notamment pour les documents opaques et pour les documents transparents. Guides de bonne pratique, ils constituent les documents de référence communiqués aux prestataires, aux ateliers internes, aux conservateurs qui choisissent les documents à numériser et aux contrôleurs de la qualité de production. Ils permettent d'évaluer sans ambiguïté la qualité de la production par rapport à la qualité attendue.

f. Post-production : Cette étape est étroitement liée à la prise de vue. Elle permet de saisir toutes les informations signalétiques et descriptives du document en cours de numérisation, qui permettront de faciliter les parcours de consultation. Selon le niveau de complexité du fonds, le niveau d'exigence des outils de consultation et le budget alloué à ces opérations, on peut se contenter d'encapsuler un ouvrage imprimé dans un fichier PDF multipages qui sera fourni au lecteur ou, au contraire, typer et nommer toutes les pages, identifier les tables des matières et d'index, les illustrations, les pages de titre, et tout type de méta données qui semblent importantes pour permettre une navigation aisée et fluide.

Un référentiel de saisie des méta données a été élaboré par la BnF.

g. OCRisation (de Optical Character Recognition) : Autrement appelée en français Reconnaissance optique de caractères, cette opération réalisée au moyen d'un logiciel spécifique permet de transformer une page imprimée numérisée en une séquence de caractères, comme si elle avait été créée avec un logiciel de traitement de texte. Pour certaines catégories de documents, cette étape est très importante pour compléter l'indexation élaborée en post-production. Au-delà de la reconnaissance de caractères, la reconnaissance des polices, des styles, de la segmentation et de la structuration des pages sont autant d'éléments qui permettent de recréer la page telle qu'elle a été conçue.

Le résultat obtenu dépend du niveau d'exigence demandé, de la qualité typographique du document original, de sa complexité et de la qualité de la prise de vue numérique. Réaliser un véritable livre électronique ou epub demande un travail considérable, qu'un OCR seul ne peut résumer et qui oblige à une saisie manuelle du texte. À l'inverse, un OCR brut, non affiché mais associé à une numérisation en bit map fera très bien l'affaire pour une indexation destinée à compléter les données bibliographiques de premier niveau.

h. Traitement physique éventuel des documents : Les documents ayant fait l'objet de modifications physiques pour faciliter leur numérisation (ou ayant été involontairement dégradés) peuvent le cas échéant devoir être remis dans leur état ou leur conditionnement d'origine une fois les opérations de numérisation et de contrôle achevées.

i. Contrôle qualité : C'est une étape indispensable pour vérifier que l'objet numérique produit est en tous points conforme aux spécifications techniques demandées.

Ce contrôle se décline en 2 étapes :

- Un contrôle automatique qui doit être exhaustif et systématique sur toute la collection produite. Il s'agit d'un contrôle totalement automatisé qui ne nécessite pas d'intervention humaine sauf pour la compréhension des erreurs identifiées par le logiciel de contrôle (structure incohérente des fichiers, des en-têtes TIFF ou JPEG2000, des métadonnées, etc.). Il permet que l'ensemble des produits numériques obtenus soit archivé de façon homogène et organisée.

- Un contrôle visuel : Selon la nature des collections, les enjeux de conservation et de diffusion, les volumes produits et les moyens humains disponibles, le niveau d'exigence variera d'un contrôle exhaustif à un contrôle par échantillonnages plus ou moins importants.

i. Le contrôle visuel doit permettre de vérifier la qualité de l'image et de son cadrage, la complétude du document numérique, et plus généralement le respect des règles de prise de vue telles qu'énoncées dans les référentiels.

ii. Concernant la qualité des images, il faut garder en tête une règle absolue : il n'y a pas de qualité universelle. La qualité d'une image numérisée est liée au document, au numériseur utilisé, aux consignes données et parfois même à l'opérateur qui était aux commandes du numériseur. C'est la connaissance de l'ensemble de ces données qui permettra de valider ou d'invalider une qualité d'image. Dans certains cas, la comparaison avec le document original sera indispensable pour statuer sur la qualité des images.

## 5. Les moyens de production

### Les numériseurs

L'évolution des nouvelles technologies impacte fortement le domaine de la numérisation où l'offre de matériels de prise de vues augmente et se diversifie sans cesse.

Cette révolution toujours en cours a permis aux constructeurs de numériseurs de diversifier leur offre et de rendre de plus en plus accessibles les prix de leurs matériels, même si les gammes professionnelles de scanners nécessitent encore des budgets confortables.

On trouve actuellement 2 types de systèmes :

- Les numériseurs à balayage
- Les dos numériques

Numériseur à balayage : 1 barrette de CCD (charge couple device ; capteur de transfert de charge sensible à la lumière) balaie le document et capture pas à pas les pixels. On peut ainsi numériser des documents de longueur variable. La limite de taille du document est donnée par le mécanisme de balayage, de la longueur des barrettes et la table qui supporte le document.

Dos numérique : 1 capteur matriciel acquiert la totalité du document en une seule fois. Ce capteur est composé de CCD (charge coupled devices) ou CMOS Complementary Metal Oxide Semiconductor) plus populaires et mieux implantés. C'est le principe d'acquisition de l'appareil photo. La capture est instantanée. Du nombre de pixels de la matrice on déduit à une résolution donnée la dimension maximale des documents (hauteur X largeur) qui peuvent être traités avec ce matériel.

- Les numériseurs à balayage sont des machines homogènes généralement issues d'un travail de conception industrielle, qui se révèlent fiables et stables dans le temps. Après son installation et sa mise en ordre de marche par les techniciens du constructeur, ce type de matériel se dérègle rarement, est autonome, utilisable par une variété d'utilisateurs assez large et pas seulement par des photographes qualifiés.

Le temps de balayage est le point faible de ce type de matériel ; ce temps est d'autant plus long que le format du document est grand et la définition recherchée élevée. On trouve cependant de plus en plus de capteurs matriciels sur ces dispositifs.

Les modalités précises de la capture de l'image et de son développement ne sont jamais révélées par le constructeur, elles relèvent du secret industriel. Cette connaissance non partagée avec l'utilisateur prive parfois celui-ci des outils nécessaires pour paramétrer finement la qualité dans le cas de collections complexes.

- Les dos numériques : Comme indiqué ci-dessus, l'évolution des technologies du numérique et de l'informatique a permis d'augmenter de façon très significative la taille des capteurs des appareils photographiques numériques.

Contrairement aux dispositifs à balayage dans lesquels l'ensemble optique-lumière-support est intégré et piloté par un logiciel, les dispositifs de numérisation matricielle sont d'une nature plus 'artisanale' : Les dispositifs d'éclairage, les porte-documents associés et le logiciel de pilotage sont autant d'éléments d'un kit qui, modulable et assemblé selon les besoins de l'utilisateur, permet de constituer un dispositif de numérisation.

Un capteur matriciel de 80 millions de pixels permet à titre d'exemple de numériser à 400 dpi un document au format A2, ce qui permet de rendre attractif ce type de matériel par rapport aux numériseurs à balayage. Ce type de matériel demande cependant une maîtrise plus importante de la technique de photographie numérique qui n'est pas forcément à la portée d'opérateurs non qualifiés.

Outre le choix important à effectuer entre ces 2 grandes catégories de numériseurs, il faut également choisir avec soin le type de support ou de berceau à adapter aux numériseurs pour permettre de traiter les documents non pas en ouverture à 180 degrés mais en ouverture restreinte (80° à 120 °). Cette ouverture restreinte permet de limiter la pression sur les documents à la prise de vues et donc contribue à leur sécurité. Depuis maintenant quelques années, on voit apparaître de nouveaux dispositifs, conçus en général autour de dos numériques, qui permettent de numériser en ouverture restreinte et avec une double prise de vues les pages droite et gauche d'un même document. Ils permettent souvent un gain de temps important et une manipulation plus simple et moins risquée des documents.

## L'unité de production

La nature des collections, leur préciosité, leur état, mais également la volumétrie à produire, les moyens humains disponibles et les budgets alloués à ces opérations sont autant de paramètres à prendre en compte pour choisir la solution de production la mieux adaptée.

Les trois solutions possibles sont passées en revue ci-dessous, avec leurs avantages et inconvénients :

- Création d'un atelier interne par l'établissement : Cette solution implique l'achat des matériels de numérisation, qui n'est rentable que si les collections à reproduire sont relativement importantes, tout en étant assez homogènes pour que l'on puisse limiter le parc des scanners à acquérir. Elle nécessite également de former des personnels à l'utilisation des numériseurs, à la post-production des fichiers produits. La productivité obtenue est généralement nettement inférieure à celle d'une société spécialisée. Il faut inclure également dans les prévisions budgétaires le coût de maintenance des matériels, qui n'est pas négligeable (de l'ordre de 12% du prix d'achat chaque année).

- Marché de prestation de numérisation : La première question à se poser est l'externalisation ou non de la prestation : numérisation dans les locaux du prestataire ou demande faite au prestataire d'installer un atelier in situ.

- Externaliser la production permet de traiter à moindre coût des collections plus diverses, nécessitant un parc matériel plus important. Ce matériel étant généralement utilisé simultanément pour plusieurs marchés, les coûts d'amortissement sont également répartis ce qui permet d'obtenir des prix de revient plus faibles. La souplesse totale d'organisation laissée au prestataire (l'atelier peut fonctionner en 2 X 8 heures, 3 X 8 heures, etc.) permet également d'obtenir des coûts plus réduits. En revanche, la sécurité des collections qui doivent sortir de la bibliothèque est plus compliquée à assurer, au moyen de contrôles de complétude et d'intégrité au retour. Des exigences strictes en matière de sûreté lors du transport et du séjour des documents doivent être formulées dans le cahier des charges du marché. Malgré cela, cette solution exclut dans la pratique les documents trop fragiles, précieux, à forte valeur d'assurance. Elle permet en revanche d'alimenter la chaîne avec des documents aux formats très hétérogènes pouvant aller jusqu'au A0 voire davantage.

- Demander au prestataire la création d'un atelier in situ constitue la solution la plus souple si l'on souhaite alléger fortement la logistique et la sécurité des collections, puisque les documents ne quittent pas les emprises de la bibliothèque. En revanche, elle est plutôt déconseillée pour la numérisation des collections de grand format (à moins qu'elle ne porte sur des volumétries très importantes), pour éviter au prestataire d'investir dans des matériels coûteux pas forcément rentabilisés sur le projet, ce qu'il ne manquera pas de répercuter dans les tarifs proposés.

Dans tous les cas de figure, un cahier des charges doit décrire de façon exhaustive la prestation demandée et aborder tous les points importants de la chaîne de numérisation. Clair et précis, il doit permettre aux prestataires de répondre à l'appel d'offres avec une parfaite compréhension de la demande de l'établissement, et de s'assurer de la qualité de l'offre retenue, qui conditionne celle de la production.

Ce cahier des charges décrira précisément :

Le fonds à numériser en typologies, support et volumétrie,

Les conditions de sa mise à disposition au prestataire et de restitution du prestataire (conditionnement, le cas échéant transport, assurance, constat d'état, bordereaux descriptifs accompagnant le fonds, etc.),

Les règles de numérisation (paramètres techniques résolution, format, profondeur de codage, et les règles de prises de vues),

La nature, le type et le format des métadonnées à saisir autour de la numérisation,

La phase de test demandée préalable à la mise en production,

La définition du niveau de qualité demandée et les procédures de contrôle qualité mises en place pour vérifier la qualité de la production.

La livraison, ses modalités, les formats et le support,

Les modalités de réfections des documents déclarés à refaire.

Et demandera l'organisation mise en place par le prestataire, les moyens matériels et humains ainsi que les outils déployés pour atteindre l'objectif fixé.

# Conclusion

Après avoir passé en revue les paramètres qui permettent de monter un projet de numérisation et toutes les étapes qui jalonnent la chaîne de production, on insistera, en conclusion, sur les conditions à rassembler pour qu'un programme de numérisation constitue également une opération de conservation pérenne des collections concernées :

Le premier intérêt de la numérisation est, comme souligné en préambule, de limiter voire supprimer la communication de l'original, ou de s'y substituer si elle est devenue impossible.

-Le programme doit donc cibler, de façon prioritaire et significative, des fonds fragiles, endommagés, rares et précieux, et/ou fréquemment communiqués : collections spécialisées de manuscrits, estampes, photographies, etc., monographies et surtout presse des années 1880 à 1950 incommunicables et menacés de disparition par l'acidification de leur papier. Ou encore des fonds photographiques qui ne trouveront une pérennité que dans un changement de support.

Avoir constitué une cartographie documentaire et matérielle de son fonds est donc un prérequis indispensable pour savoir comment organiser et planifier la numérisation, seule ou combinée avec des traitements de réparation voire de restauration.

Une fois ce plan établi, il faut :

-Définir le niveau de qualité élevé nécessaire à une conservation et un usage pérennes du document numérique qui permettra de diversifier les usages futurs de ce document sans avoir à lui faire subir une nouvelle numérisation : les paramètres techniques de la numérisation, les règles de prises de vues et de post-production, doivent donc être conformes au meilleur état de la technique, précisément indiqués, de même que les priorités à observer si le document pose des difficultés (par exemple, choisir entre une numérisation légèrement incomplète et une reliure légèrement endommagée). Les numériseurs doivent être choisis selon ce niveau de qualité et les caractéristiques physiques des documents. La qualité des métadonnées adjointes aux images numériques est évidemment capitale, elle suppose obligatoirement que chaque document ait été correctement catalogué au préalable. Les formats de conservation retenus doivent être pérennes, permettant une compression minimale des images.

-Garantir une prise de vue qui ne dégrade pas les documents originaux (ou qui les dégrade le moins possible dans le cas de pièces déjà très dégradées), ce qui implique de former les opérateurs aux bonnes pratiques et au respect du document, de choisir des scanners aux supports adaptés : désormais, presque tous les modèles proposent des « berceaux » permettant une numérisation de qualité sans ouvrir à plat les volumes) ; souvent, les documents fragiles (papier acide), les reliures serrées nécessitent des réparations, consolidations ou préparations physiques préalables qui devront être organisées en conséquence.

- Il importe bien sûr de choisir les supports de stockage et mettre en place les outils de surveillance pour garantir l'intégrité de la collection numérique ; ces points seront développés dans la partie suivante.

- Enfin, la politique de communication de l'établissement doit être modifiée en conséquence pour maximiser les bénéfices de la numérisation sur la conservation des collections : tout document numérisé ne doit plus être communicable physiquement, sauf cas exceptionnel.

Au total, les professionnels des bibliothèques ou archives désireux de se lancer dans la numérisation de leurs fonds bénéficient donc d'un contexte très favorable, les évolutions rapides et constantes des techniques informatiques et électroniques permettant d'ambitionner une qualité très supérieure à tous égards à celle qu'ils pouvaient espérer au début du siècle.

Le parc de numériseurs à disposition permet désormais de reproduire la majeure partie des documents en ménageant leur intégrité quel que soit leur état de conservation ; le très bon niveau de qualité obtenu fait de la numérisation non plus seulement un produit de diffusion mais aussi un produit de conservation à part entière. Il est d'autant plus intéressant de l'utiliser à cet usage que, dans un contexte économique dégradé, la numérisation est souvent l'une des rares activités permettant d'obtenir des crédits. On n'oubliera pas pour autant qu'elle ne peut pas se substituer à une politique conséquente et suivie de conservation physique des collections ; celle-ci étant même une condition indispensable à moyen et même court terme pour permettre une numérisation rapide et de qualité à des coûts raisonnables.