

Module 9 - Section 2 :

Numériser les documents

PATRICK PERROT

08 mars 2016

Table des matières

Objectifs	5
1. Introduction à la numérisation	6
1.1. Définitions.....	7
1.2. La numérisation: une technique aux bases anciennes.....	7
1.2.1 1.2.1. La transmission des documents écrits ou dessinés.....	7
1.2.2 1.2.2. La transmission des photographies à distance.....	8
1.2.3 1.2.3. L'héritage actuel.....	8
1.3. La bonne numérisation.....	9
2. Notions fondamentales en numérisation	10
2.1. Image en pixels.....	10
2.2. Résolution.....	11
2.3. Modes colorimétriques.....	12
2.4. Analyse et typologie des documents à numériser.....	16
2.4.1 2.4.1. Analyse du document.....	16
2.4.2 2.4.2. Nature des documents déterminant la colorimétrie.....	16
-	17
3.1. Objectifs.....	17
3.2. Usage du fichier final.....	17
3.3. Portabilité.....	17
3.4. Répartition des documents numérisés selon l'usage.....	18
3.5. Usages mixtes.....	19
-	20
A propos de ce chapitre des formats.....	20
A. Les propriétés des formats d'enregistrement.....	20
4.1. Les qualités des formats pour les besoins archivistiques.....	22
4.2. Formats sans perte de données.....	23
4.3. Formats avec pertes de données.....	26
4.4. Choix des formats d'enregistrement selon l'usage.....	31
4.4.1. Création d'une photothèque numérique.....	31
4.4.2. Création d'une bibliothèque de documents à pagination multiple.....	32
4.5. Formats libres et formats propriétaires.....	35
4.6. Microfilm et numérisation : des champs d'application connexes.....	35
-	37

5.1. Phases des opérations de numérisation.....	37
5.2. Le choix du numériseur.....	38
5.3. La disparition de certains types de numériseurs.....	38
5.4. Analyse des caractéristiques des numériseurs.....	38
5.5. Inventaire des fonctionnalités.....	39
5.6. Numériseurs de petit format.....	39
5.7. Numériseurs de grand format.....	43
5.8. Numériseurs de films et microfilms.....	49
5.9. La pratique de la numérisation en centre d'archives.....	53
5.10. La mise en place des éléments du poste de numérisation.....	53
5.11 Ségrégation des documents.....	53
5.12 Post-traitement des images.....	54
5.13. Informations textuelles dans le fichier image.....	54
5.14. Trucs et astuces.....	56
5.14.1. Documents plus grands que la vitre d'exposition.....	56
5.14.2. Numérisation d'un ouvrage «à la baguette».....	56
-	58
6.1. La nature des manuscrits.....	58
6.2. Choix d'un procédé physique pour les reliés.....	59
6.3. Choix d'un procédé physique pour les feuilles volantes.....	59
6.4. Choix du mode colorimétrique.....	60
6.4.1. Le mode bitonal.....	60
6.4.2. Le mode niveaux de gris.....	60
6.4.3. Le mode couleurs RVB.....	61
6.4.4 Reliure en documents PDF.....	61
6.5. La numérisation de microformes.....	61
6.6. Importance de la taille des fichiers.....	62
6.7. La reconnaissance optique des caractères.....	64
-	66
7.1. Première approche en numérisation de photographies.....	66
7.2. Plaques de verre.....	66
7.3. Diapositives et négatifs de petit format.....	69
7.4. Photographies sur papier.....	70
7.4.1. Photographies monochromes.....	70
7.4.2. Photographies en couleur sur papier.....	71
-	73
8.1. Numérisation au bon mode.....	73
8.2. Cas spéciaux.....	79
-	80

9.1. Contenu des imprimés, procédés d'obtention.....	80
9.2. Principe de la similigravure.....	80
9.4. Principe de l'héliogravure.....	87
9.5. Lithographies et Chromolithographies.....	88
-	90
10.1. Plans tracés, cartes.....	90
10.2.Tirages de plans.....	91
10.3 Moyens de numérisation.....	93
-	95
11.1. Introduction aux choix faits dans les solutions de la DAF.....	95
11.2 Complément d'information sur le contenu des colonnes de solutions de la Direction des Archives de France....	96
-	99
12.1 Numérisation, sites en ligne.....	99
12.2 Bibliothèques en ligne.....	100
-	101
-	104
-	105
Solution des exercices	118
Glossaire	123
Bibliographie	124

Objectifs

Description du module :

Dans une politique de préservation bien conçue, la reproduction des documents, par microfilmage et numérisation, peut être considérée comme une véritable mesure de conservation préventive puisqu'elle permet d'arrêter la manipulation des documents originaux qui sont fragiles et précieux. La numérisation est à présent prépondérante dans le processus de consultation, prenant le pas sur la consultation des microfiches in situ. Un document numérisé est disponible au domicile du chercheur, quel que soit son éloignement du centre d'archives.

Le microfilmage reste le seul moyen pérenne de conservation des documents, et il est à présent considéré comme un moyen de sauvegarde et non plus de diffusion.

Le but du module est de :

présenter les techniques de base pour l'utilisation raisonnée des moyens de numérisation.

L'apprenant doit être en mesure de :

- reconnaître la typologie des documents vis à vis du processus de numérisation
- déterminer les moyens matériels de numérisation adaptés aux documents
- déterminer les modes colorimétriques en facteur de l'objectif de la numérisation et de la typologie du document
- pratiquer la numérisation de toutes sortes de documents écrits ou iconographiques
- déterminer les formats d'enregistrement en fonction de l'usage
- adapter les sous fichiers à la diffusion en ligne et faire des assemblages de pages multiples
- introduire des métadonnées dans l'image pour déclencher le fonctionnement des moteurs de recherche à partir de l'objet.

Positionnement :

Deuxième module du cours 4 sur les politiques et pratiques de conservation et préservation des documents.

Il suit tout naturellement le cours de gestion et traitement des archives et précède celui sur leur communication et mise en valeur.

Conseils d'apprentissage :

Prérequis :

Avant d'aborder ce module, il convient d'avoir pris connaissance des modules 2 (notions générales d'archivistique) et si possible 6 (traitement des archives définitives).

Prérequis pour la numérisation :

- connaissances de base en informatique
- pratique de la photo ou de l'image en général nécessaire pour les fonds iconographiques

Conseil :

Le cours ne donne pas les détails pas à pas pour la pratique de la numérisation sur tel ou tel équipement. Il faut suivre le cours par étapes et tenter d'appliquer la connaissance, à chacune de ces étapes, sur son propre équipement. Ce sont seulement les exercices pratiques qui permettent d'évaluer les difficultés de terrain en matière de reproduction.

Où aller en fin de module?

Le cours 5 sur la communication et la valorisation des archives est la suite logique.

Des images qui « parlent »

Si vous téléchargez des images de ce cours, elles contiennent pour la plupart des métadonnées et leur légende. Vous pouvez réutiliser librement ces images pour composer un cours condensé ou une présentation aux collègues du centre d'archives. Pour s'informer sur les métadonnées, consultez le chapitre 5.12

1. Introduction à la numérisation

1



Image 1 Introduction à la numérisation

La numérisation des documents est un moyen de diffuser ces documents de manière immatérielle.

Au départ, il existe une image réelle, dessinée, imprimée ou photographiée sur papier, c'est le document, parfois vieux de plusieurs siècles, auquel on va donner une nouvelle vie : il pourra se transmettre par des réseaux de communication (la toile «web»), et être restitué à l'autre bout de la planète sans perte de qualité.

L'exemple le plus significatif est celui des journaux. Ainsi pour l'impression quotidienne du *Times* il fallait 17 tonnes de plomb et des tonnes de papier qui étaient distribuées à New York et aux quatre coins du monde. Maintenant d'un simple clic de souris, de bon matin, le fameux journal apparaît à l'écran

1.1 1.1. Définitions



Image 2 introduction binaire

Avant d'entrer dans le détail de notre cours, nous devrions hiérarchiser les types d'archives **électroniques** que nous sommes appelés à manipuler. La famille d'archives virtuelles qui sont relatées dans ce cours sont générées par l'ordinateur, elles appartiennent aux archives numériques.

Archives numériques

Ensemble des archives numérisées (images des archives physiques) et des archives informatiques, conservées sur un support numérique. Elles font partie des archives électroniques. Une archive numérique est restituée de façon directe par le truchement d'un ordinateur.

On l'appelle numérique parce que l'ordinateur est conçu pour fonctionner avec un langage binaire caractérisé par deux chiffres 0 et 1 représentant un passage de courant ou le non passage de courant dans le processeur.

Les archives numérisées sont des copies de documents physiques existant bel et bien dans un centre d'archives.

Archives informatiques

Les archives informatiques sont les documents créés à l'origine par un ordinateur ou entrés en centre d'archives sous forme numérique, par l'intermédiaire d'un système informatique de transmission à distance ou d'un enregistrement sur support informatique (CD ROM, DVD....).

Ce sont par essence des fichiers texte, des bases de données, des résultats de calculs, des données bancaires, des fichiers de Publication Assistée par Ordinateur (PAO), de Dessin Assisté par Ordinateur (DAO), etc...

Certaines de ces archives peuvent contenir des fichiers de documents numérisés, mais dans ce cas on ne retiendra que la forme sous laquelle elles sont entrées dans le centre d'archives: la forme numérique (le centre d'archives ne possède pas le document physique d'origine)

Quant aux photos issues d'appareils numériques, on les fera entrer aussi dans cette catégorie des archives informatiques, car par essence le modèle original ayant servi à les générer (une scène instantanée de la réalité) n'existe plus, et les éventuels tirages sur papier en sont les copies (et non les sources).

Terme anglais : *computer data files*

Archives numérisées

Ce sont les formes numériques d'archives physiques **reproduites** au mode image par points (pixels), et dont il existe **un original physique**.

Ces documents numériques sont obtenus par l'intermédiaire d'un dispositif appelé numériseur (ou scanner pour les anglophones), qui traduit des réalités analogiques (la lumière provenant de l'original et reconnue par des capteurs) en signaux numériques



Attention : Les archives filmées

Nous considérons dans cette catégorie les productions (films) contenant des images animées et des sons, avec un titre, un réalisateur, un générique et un visa d'exploitation pour certains cas: documentaires, courts et moyens et longs métrages, publicités filmées et films publicitaires...

Les techniques modernes permettent d'utiliser l'informatique pour créer, ou enregistrer ces images animées et sonores. Dans ce cas, on classera les films sur supports numériques dans les archives filmées et non pas dans les archives informatiques, ni dans les archives numérisées, car elles appartiennent à un secteur d'activité qui n'est pas celui de l'image fixe, mais bel et bien des **archives audiovisuelles**.

Bien sûr si quelques clips vidéo de quelques secondes ou quelques prises de vues documentaires se retrouvent dans un fonds d'archives raisonné, on les y laissera, sans faire de distinguo, pour ne pas attenter à la cohésion du fonds.

Dans le cours qui suit, nous allons étudier la création des **archives numérisées**, soit la génération de **copies numériques** de documents papier existant physiquement.

1.2 1.2. La numérisation: une technique aux bases anciennes

Un appareil permet de transformer l'image de l'original en signaux électriques : c'est le numériseur.

Le numériseur est une machine qui semble nouvelle, mais ses ancêtres datent du 19^e siècle.

Le principe de base de la numérisation est de transformer un document physique en impulsions électriques par l'intermédiaire d'un code de transmission

1.2.1 1.2.1. La transmission des documents écrits ou dessinés

Le premier appareil capable de transmettre des images en deux valeurs. le noir et le blanc. est le télégraphe
jeu 7

1. Introduction à la numérisation

il servait à transmettre des textes et images que l'on traçait sur un papier couché rendu partiellement conducteur du courant par ce dessin.

Le télégraphe autographique émettait des signaux binaires qui passaient par les lignes télégraphiques (à cette époque, le téléphone n'existait pas). Les signaux binaires sont:

1= passage du courant

0 = pas de courant.

Il est à noter que le codage binaire est le principe de fonctionnement des processeurs au cœur des ordinateurs

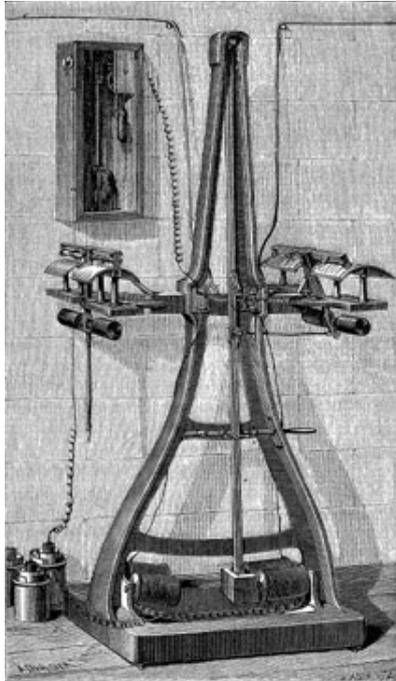


Image 3 Pantélégraphe autographique de Caselli

Il se retrouve comme procédé appliqué à la numérisation des images "au trait" que nous nommerons "bitonales" dans le cours qui suit. Le principe de Caselli est donc reconduit dans les machines actuelles.

La transformation du dessin en signaux électriques se faisant sur le pantélégraphe autographique de Caselli par un contact électrique direct, il manquait juste un perfectionnement : la cellule photoélectrique qui peut analyser la lumière réfléchie par n'importe quel original et la transformer en signaux électriques, sans contact physique

1.2.2 1.2.2. La transmission des photographies à distance

Dès le début du 20^e siècle, des savants avaient découvert les propriétés du sélénium, point de départ des recherches sur les cellules d'analyse photographique, ce qui permit au Français Édouard Belin d'achever la mise au point en 1925 d'un appareil permettant la transmission, aussi bien par le câble que par ondes hertziennes, d'images fixes en niveaux de gris.

1.2.3 1.2.3. L'héritage actuel

Dans notre numériseur informatique, l'analyse de l'image est encore faite par un procédé dit analogique: le courant varie en fonction de l'intensité du gris, pour devenir maximal pour le blanc, au niveau des capteurs de lumière de l'appareil.

Ces signaux sont ensuite interprétés en codage binaire avant d'être envoyés à l'ordinateur.

La "**discrétisation**" du signal va transformer ces variations continues en une série de niveaux prédéfinis qui donneront une illusion de continuité à la restitution du document.

Ces niveaux sont caractérisés par une valeur numérique (par exemple entre 0 et 256), d'où le terme de numérisation de l'image.

Alors que par le passé, les photographies en couleurs étaient décomposées en plusieurs films en niveau de gris pour l'expédition au bélinographe, les systèmes actuels reproduisent directement en un même fichier les couleurs des objets de la numérisation. Cependant le principe reste le même, les trois ou quatre couches de couleurs sont juste rassemblées dans un même fichier.

1.3 1.3. La bonne numérisation



Image 4 La bonne numérisation

Il faut bien avoir à l'esprit que le document numérique, comme le film ou l'enregistrement phonographique, est une reproduction de la réalité, et non la réalité.

Il peut être de **qualité médiocre**, telle une télécopie par exemple. Dans le domaine sonore, on pourrait comparer avec les nouvelles que l'on écoute sur un poste à transistor. La qualité sonore est médiocre, le message est cependant complet et compréhensible.

Il peut être de **haute qualité** au point de donner l'illusion de l'original, tout comme l'enregistrement d'un concert que l'on écoute dans un auditorium. La numérisation ne remplace pas le document d'archives, mais peut en garder une trace fidèle si on choisit bien son procédé.

On verra ci-dessous que :

- une bonne numérisation est une adéquation entre le document original, l'usage du fichier numérisé et la portabilité en réseau ;
- les objectifs de la numérisation et l'usage auquel on destine les documents numérisés sont déterminants pour le choix des procédés de numérisation.



Image 5 Notions fondamentales en numérisation

Un ensemble de connaissances est indispensable pour comprendre les mécanismes de numérisation de l'image. La connaissance de ces principes est essentielle pour garantir la qualité du produit fini et sa portabilité (taille du fichier à diffuser en ligne)

L'archiviste a tout pouvoir de faire les choix fondamentaux pour sa numérisation, et la réussite dépend non de la performance des machines mais du discernement de l'archiviste

2.1 2.1. Image en pixels



Image 6 Image en pixels

L'ordinateur étant une machine logique, il procède par un langage binaire 0 et 1. Il lui faudra donc décomposer l'image en éléments finis, appelés pixels (pour Picture Element) et affecter à chacun d'eux un code permettant de reproduire une couleur. Ce procédé est nommé discrétisation de l'image. Chaque pixel est un carré, qui se place dans une grille virtuelle définissant les limites de l'image numérique.

Les limites de l'image sont invariablement celles d'un carré ou d'un rectangle, permettant le placement de l'image dans des compositions numériques (pages de magazines, de livres, pages web, etc.)

Cependant, l'image peut apparaître avec des contours non géométriques. Cela s'appelle le détourage



Image 7 Image agrandie des pixels noirs et blancs, niveaux de gris et couleur.



Complément : Masque de détourage

Certaines images sont dotées d'un masque de détourage permettant de rendre transparentes certaines parties pour le placement sur un fond de couleur. Ce masque est un tracé vectoriel dans le monde de l'édition et est fabriqué dans les logiciels de retouche d'images professionnels. Dans la bureautique et l'imagerie "Web", on affecte à certains pixels la propriété de transparence, c'est la "couche alpha". Le format png va plus loin en offrant une véritable option de transparence graduelle. Il sera utilisé en archives pour l'option détourage des cartes postales.



Image 8 Personnage détourné, et converti au format Gif pour conserver cette propriété dans un navigateur web

2.2 2.2. Résolution

Le nombre de pixels contenu dans cette grille, pour une dimension physique de l'original, va déterminer la résolution (appelée aussi définition) de l'image. Cette **résolution** est issue des pratiques de l'imprimerie, en pixels par pouce : ppp ou en anglais dpi pour *Dot Per Inch*.

Plus il y aura de pixels par pouce :

plus la finesse de la reproduction sera élevée

plus on verra les détails

mais aussi plus le fichier sera lourd.

Lorsque l'on double la résolution, la taille du fichier en pixels (ouvert dans une application) est multipliée par 4.

Résolutions des images en tons continus

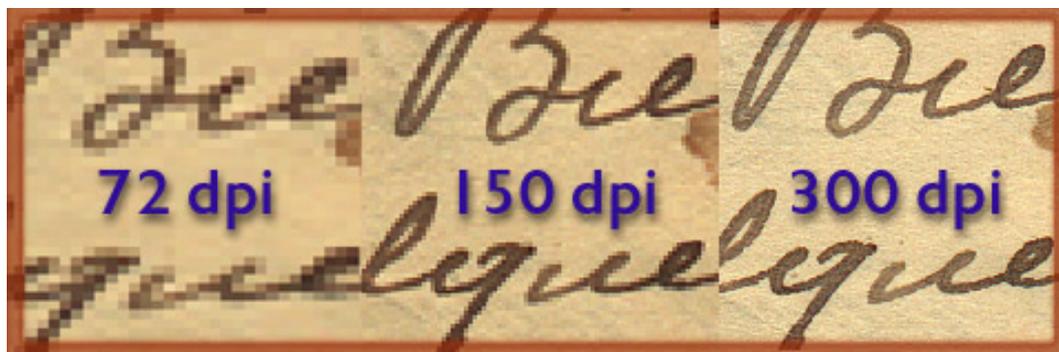


Image 9 Aspects des résolutions 72, 150, 300 dpi pour un document à tons continus (agrandissement 4x).

Les résolutions en tons continus proposées ci-dessus sont caractéristiques des usages les plus courants : 72 dpi pour la diffusion en ligne (Web), 150dpi pour l'impression domestique et 300 dpi pour la conservation et la restitution à l'échelle originale. Les mêmes résolutions s'appliquent aux documents en niveaux de gris.

Résolution des images bitonales (au trait)



Image 10 Aspects des résolutions 150, 300, 600 dpi pour un document bitonal (agrandissement 4x).

2. Notions fondamentales en numérisation

résolution va produire des effets très crénelés nuisibles à la reconnaissance des caractères et désagréables pour le lecteur. On pratiquera donc la bitonalisation dans des résolutions plus élevées que pour les tons continus.

La résolution 150 dpi analogue à celle d'un télécopieur, demeure lisible pour les textes dactylographiés ou imprimés. Pour les textes manuscrits, elle peut être insuffisante.

Les résolutions en mode bitonal proposées ci-dessus sont caractéristiques des usages les plus courants: 150dpi à 200 dpi pour diffusion en ligne et documents commerciaux, 300 dpi pour l'impression domestique et 600 dpi pour l'impression laser des dessins, des gravures... L'imprimerie utilise systématiquement des résolutions plus élevées: 1200 dpi ou 2400 dpi, pour s'aligner sur la finesse des flasheuses offset.

Choix de la résolution :

Les solutions de numérisation décrites dans ce cours sont basées sur la reproduction à l'échelle 1 des documents numérisés, ce qui est à la base du travail archivistique : restituer le document dans son intégrité, y compris le format. On verra qu'il y a une exception : les négatifs photographiques qui devront être reproduits avec un agrandissement.

C'est principalement le choix de la résolution du fichier numérique qui va déterminer sa qualité intrinsèque. Il est recommandé de numériser le document physique dans la résolution adaptée aux plus exigeant des usages que l'on en fera (par exemple celle de l'impression).

Par ailleurs, pour les fonds d'archives publiques, il faudra se conformer aux cahiers des charges standardisés, comme ceux préconisés par la Direction des Archives de France. Des liens contenus à la fin de ce cours au chapitre « Solutions de numérisation en tableaux ».

Génération de fichiers en résolution réduite

Pour la diffusion vers des réseaux publics et supports éducatifs, il suffira ensuite de générer des fichiers en plus basse résolution en utilisant les commandes de ré-échantillonnage des logiciels de traitement d'image.

Les documents multi-pages en fichier PDF peuvent aussi être sous-échantillonnés en pratiquant une nouvelle impression PDF du document élaboré en haute définition. Il faudra choisir une plus basse résolution dans le réglage «compression» de l'imprimante PDF en activant l'option «sous-échantillonnage».

Il existe des moyens de calcul des résolutions optimales en fonction de la finesse du trait pour la reproduction des gravures sur bois (estampes)



Complément : Résolutions et droits

La diffusion en ligne de fichiers image à haute résolution pose des problèmes de droits d'auteur des oeuvres. Si la résolution et la qualité sont élevées, des éditeurs peu scrupuleux pourraient utiliser le document pour l'éditer avec des profits commerciaux, et ceci complètement à l'insu du centre d'archives. Le problème devient critique parce qu'en publiant, les faussaires acquièrent des droits d'auteur sur l'ouvrage qu'ils ont réalisé. C'est pourquoi pour les iconographies modernes, comme les cartes postales, les affiches (surtout les affiches protégées par l'ADAGP), les photographies de personnalités...il faudra choisir des résolutions basses et surtout un nombre de pixels limité à 1000 sur le plus grand côté de l'image pour prévenir de ce piratage.

Cependant, l'archiviste n'appliquera pas forcément cette méthode aux documents très anciens appartenant à des collections publiques, des journaux anciens, des documents de techniques révolues qui ne présentent pas forcément autant d'enjeux médiatiques

2.3 2.3. Modes colorimétriques



Image 11 Modes colorimétriques

Pour la monochromie ou la couleur, les fichiers pixellisés peuvent être enregistrés dans différents modes. Le mode définit l'encodage que l'on va affecter à chaque pixel : il sera très simple pour un fichier représentant du noir pur et du blanc pur et complexe pour les fichiers représentant une image en couleurs: la tonalité, la luminance et la saturation de la couleur perçue par l'œil humain sont à traduire en valeurs numériques. Différents codages sont utilisés par les logiciels d'imagerie, mais la numérisation des documents d'archives fait principalement appel à trois d'entre eux :

1. le mode bitonal ou bitmap pour la reproduction des documents et textes illustrés de gravures,
2. Le mode niveau de gris essentiellement pour la reproduction des archives photographiques dites en « noir et

blanc »,

3. Le mode RVB (rouge vert bleu) pour la reproduction de tous documents dont on veut restituer la couleur

2. Notions fondamentales en numérisation

Remarquer la persistance de la tache en mode bitonal : les densités dépassant le seuil de conversion apparaissent.



Image 12 De gauche à droite un mode bitonal, un mode niveau de gris, un mode couleur appliqués à un même document.

La notion de couches

L'image bitonale et l'image en niveaux de gris sont formées d'une seule couche, c'est à dire une seule série d'informations disposées en cartographie sur un plan pour créer une série de pixels juxtaposés. Ces pixels forment une image

Dans l'image bitonale comme dans l'image en niveaux de gris, la présence d'une seule série de nuances permet de leur donner une couleur de référence. Par défaut, le fichier bitonal et le fichier en niveaux de gris sont des variations du noir.

Dans les fichiers en couleurs, il a fallu faire la décomposition de l'arc-en ciel, mais en se limitant strictement aux couleurs de base rouge-vert-bleu. Ce système est appelé additif, parce que le mélange avec 100% pour chacune des couleurs donne le blanc.

S'il y a trois couleurs, il y aura forcément trois couches, qui seront le rouge, le vert, le bleu. (RVB en Français, RGB en anglais). Chacune des couches sera nuancée en un certain nombre de niveaux, et c'est la combinaison de ces niveaux qui va donner les couleurs intermédiaires (le jaune, le vert, les pastels...) Le nombre de couleurs intermédiaires que l'on peut obtenir est facteur de la profondeur d'analyse allouée lors de la discrétisation du signal analogique.

Un autre système d'encodage de couleurs existe qui ne fait pas appel à une notion de couches, mais à un modèle mathématique élaboré : le LAB (Luminosité, couleur A et couleur B.) Ce modèle offre le spectre colorimétrique le plus étendu, mais il n'est pas usité comme mode colorimétrique en archives, son usage est plutôt scientifique, médical ou astronomique, et sert à faire fonctionner les logiciels de retouche d'images.

La notion de couleur

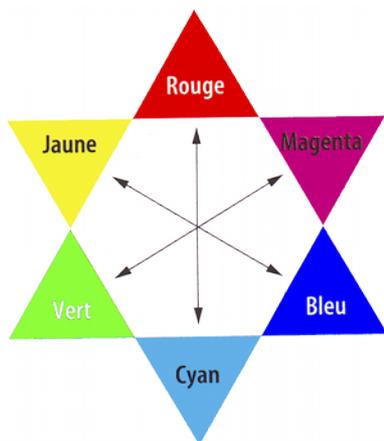


Image 13 roue chromatique

Les couleurs perceptibles sont celles issues du spectre de la lumière solaire, le fameux « arc en ciel » que l'on peut caractériser par trois couleurs de base, le rouge, le vert, le bleu. L'addition de ces trois couleurs en quantité égale donne le blanc pur. En faisant varier la proportion entre ces couleurs, on obtiendra une infinité de nuances, qui font toute la subtilité des couleurs de la vie. Ce mélange est appelé mélange additif. Plus on ajoute des couleurs, plus on va vers le blanc. Dans le cas de l'informatique, ce sera la lumière incidente provenant de l'écran.

Lorsque l'on veut reproduire sur un support non lumineux (le papier blanc) une illusion de l'arc-en-ciel, il faudra faire appel à un mélange soustractif, c'est à dire que plus on ajoute des couleurs, plus on va vers le noir. On utilise pour cela des couleurs complémentaires. Ces couleurs complémentaires sont le cyan, le magenta, le jaune. Ce sont les tubes de couleurs que tout peintre a dans sa trousse. Il a aussi un tube de noir pour représenter les ombres, et un tube de blanc pour obtenir les nuances pastel. Dans le monde de l'imprimerie, on utilisera les encres cyan magenta Jaune et du noir pour renforcer les contrastes dans les ombres, pour compenser la

transparence des encres. La nuance pastel est obtenue par le blanc du papier, entre les les points d'encre de chaque couleur de base.

La roue chromatique ci-dessous montre ces couleurs de base (spectre solaire) et les couleurs complémentaires. Alors que le rouge est une couleur de base, pour obtenir du rouge avec des couleurs complémentaires, il faudra mélanger le magenta et du jaune en proportions égales, du cyan et du jaune pour faire le vert, du cyan et du magenta pour faire le bleu.

2. Notions fondamentales en numérisation

Les spectres de restitution des couleurs

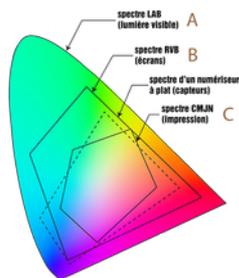


Image 14 Les spectres LAB zone A, RVB zone B, CMJN zone C

Comme on vient de voir précédemment, il y a deux principes de restitution des couleurs : par les couleurs de base, ou par les couleurs complémentaires. Dans l'image numérisée, ces deux familles de modes existent, mais ont des usages dédiés :

- par les couleurs de base, on fera tout ce qui est l'acquisition, le post-traitement, et le stockage de l'image,
- par les couleurs complémentaires, on ne fera que la reproduction de ces images sur le papier, selon le modèle CMJN (Cyan Magenta Jaune Noir).

Ce dernier mode en « quadrichromie » sera donc ignoré des archivistes, parce qu'on va le voir, son spectre colorimétrique est restrictif, il engendre une perte de données, et prend plus de place en enregistrement à cause des 4 couches au lieu de

3.

La figure ci-dessous montre les spectres de couleurs reproductibles par les modèles normalisés qui sont utilisés par les applications informatiques.

Le spectre le plus étendu identifié par A est celui du modèle LAB, (luminosité, couleurs A et B). Ce modèle est utilisé essentiellement en applications scientifiques, les numériseurs d'archives ne donnent pas de fichiers discrétisés dans ce mode. C'est aussi l'espace de travail de l'application Photoshop®, pour reculer les limites d'altération lors des transformations colorimétriques de tous les formats de fichiers.

Le spectre le plus répandu identifié par B est celui du modèle RVB (Rouge vert bleu). Il couvre la plupart des couleurs perceptibles par l'œil humain. C'est un spectre suffisant pour l'archivistique, bien que parfois certaines couleurs de papier ou de peintures très vives se trouvent à la frange de ce spectre.

Le spectre de l'imprimeur identifié par C est celui du modèle CMJN (Cyan Magenta Jaune Noir). Il apparaît nettement réduit par rapport au RVB, ceci apparaît nettement lors de la transformation d'une photo numérique contenant un ciel d'un bleu soutenu : il s'affadit brusquement. Le modèle CMJN est bridé de fait par la nature des colorants des encres d'imprimerie.

Les spectres des différents espaces colorimétriques

Cette figure est une représentation en RVB qui ne traduit pas la réalité des couleurs mais seulement le rapport entre les spectres couverts

La profondeur d'analyse des couleurs ou échantillonnage

La profondeur d'analyse de la couleur, parfois appelée « échantillonnage » va permettre d'obtenir des nuances de plus en plus subtiles lorsqu'on augmentera le nombre d'informations de description de la nuance sur la couche.

La profondeur d'analyse s'exprime en « bits », autrement dit le nombre d'informations 1 ou 0 qui définissent la couleur d'un pixel dans une couche.

L'échantillonnage le plus simple est celui du mode **bitonal** : 1 bit, autrement dit une information 0 pour l'absence de couleur et 1 pour la présence de couleur. (blanc=0, noir=1)

L'échantillonnage standard du **niveau de gris** se fait avec 8 bits, résultant en 256 niveaux de gris. Cela couvre les possibilités de perception de l'œil humain dans cette gamme.

Pour faire de la **couleur**, on affecte 8 bits à chaque couche R, V, B, ce qui en les combinant donne environ 16 millions de couleurs. L'œil humain peut en distinguer 60 000 environ

Mais il existe des **échantillonnages plus poussés**, de 14, 16, 32 bits par couche, ce qui donne 65536 nuances de gris en 16 bits ou des milliards de couleurs en RVB.

A quoi cela sert-il ? Dans le cadre de l'archivistique, l'acquisition des négatifs photo en niveaux de gris utilise le 16 bits pour pouvoir restituer une gamme complète de niveaux en récupérant toutes les nuances subtiles des images à bas contraste.

Pour la couleur, un échantillonnage de 14 à 16 bits par couche trouve son usage pour le « repêchage » des diapositives décolorées par le temps, cela permet de distinguer les nuances affaiblies par la dégradation des colorants.



Attention : L'échantillonnage reste dans l'espace couleur défini

Augmenter l'échantillonnage d'un fichier se traduit par « couper les cheveux en 4, puis en 8, puis en 16 etc. » c'est à dire que le cheveu restera le cheveu, si l'espace colorimétrique est limité, on restera dans les limites de cet espace. Pour saisir des couleurs supplémentaires, il faut changer d'espace colorimétrique. Cela peut être un RVB « maison » du fabricant de numériseurs, qui aura augmenté cet espace en fonction de ce que peuvent donner ses capteurs, ou un espace étendu universel comme Adobe RGB 1998 par exemple.

2. Notions fondamentales en numérisation



Complément : Profils colorimétriques des modes RVB

Il n'existe pas qu'une sorte de RVB, différents profils existent sur le « marché » Ils ont été développés soit par des scientifiques, soit par des développeurs d'applications. Des traducteurs assurent la conversion de profils à l'entrée dans l'application. Hormis en reproduction des photothèques en couleurs, ces profils ont peu d'influence visible sur les documents anciens écrits comme les registres, les livres, les manuscrits.

Diverses recommandations existent en France dans les ministères qui régissent la normalisation des formats d'archivages de leurs documents. Le profil RVB le plus rencontré est le sRGB IEC 1966-2.1 qui est un format passe-partout lisible sur toutes les plateformes. Cependant, pour la reproduction des originaux à grande richesse colorimétrique, la direction des archives de France recommande le profil Adobe RGB 1998 au gamut plus avantageux et permettant de contrôler la reproduction d'une plus large gamme de couleur en impression de haute qualité photographique (par exemple impression en hexachromie). L'utilisation de ce format nécessite l'utilisation d'écrans à gamut élargi pour contrôler les effets du post-traitement.

Ecrire un cahier des charges de numérisation.



Complément : Interprétation de la réalité par le numériseur

Lors de l'acquisition de l'image du document par le numériseur, les capteurs image vont dans tous les cas analyser une image en couleurs, et le processeur interne au numériseur va transformer ce signal analogique en signal binaire compréhensible par l'ordinateur (la discrétisation).

- il fera une interprétation des couleurs en deux niveaux (noir et blanc) pour les images bitonales autour d'un seuil de densité c'est à dire qu'il va transformer toutes les valeurs au dessous d'une certaine densité en blanc et toutes les valeurs au-dessus de cette densité en noir. C'est le réglage pertinent de ce seuil de densité qui fixera la qualité du fichier fini.
- il fera une transformation des couleurs en niveau de gris, c'est à dire qu'il va transformer toutes les couleurs en valeurs de gris, comme le ferait une pellicule noir et blanc à la prise de vue d'une scène de la vie. Là encore, comme on règle le diaphragme sur un appareil photo, il faudra régler la densité moyenne de l'image ou le point noir/point blanc pour en restituer toutes les nuances.
- Il fera une décomposition de l'image en trois ou quatre couleurs fondamentales pour générer le fichier en couleurs, et enverra à l'ordinateur les informations traduisant les densités de chaque couleur fondamentale dont la combinaison reconstituera la couleur de l'original.

L'on verra par la pratique que la restitution des couleurs n'est pas si simple qu'il paraît : l'original est éclairé avec une lumière qui n'est pas parfaite, les capteurs ont des déficiences colorimétriques... le pilote du numériseur va introduire des corrections de gamma, de densité, et élaborer de façon automatique ou fixée par l'opérateur des corrections colorimétriques et densitométriques. C'est bien de ce pilote que dépendra la qualité finale du document.



Complément : Les modes colorimétriques spécifiques

Il existe plusieurs modes colorimétriques applicables aux documents en couleurs, outre le bien connu RVB (rouge vert bleu). Ils sont utilisés principalement par les imprimeurs et les créateurs d'images numériques.

- Le mode **couleurs indexées** à nombre de couleurs réduites (256 couleurs ou moins) sert à l'affichage sur la toile (web). L'échantillonnage des couleurs peut être adapté à l'image (256 nuances créées à partir des couleurs de l'objet numérisé) ou appartenir à une palette prédéfinie: 256 couleurs Windows system ou 256 couleurs Mac OS system. L'usage de ces palettes accélère le téléchargement; c'est surtout utile pour les très petites images comme les icônes, les vignettes...
- Le mode **BICHROMIE**, (avec ses variantes trichrome, et quadrichrome) est une option de certains programmes professionnels permettant d'affecter directement des couleurs d'encre à un fichier à l'origine en niveau de gris. C'est ce mode que l'on utilise pour la simulation des images en sépia dans les publications. Les courbes d'encrage sont réglables, ce qui permet de contrôler les nuances et tonalités de l'image. Une image originale en niveaux de gris et son traitement en bichromie. Elle apparaît beaucoup plus attractive dans la publication, et une manipulation habile des courbes d'encrage permet d'améliorer le rendu des détails. Mais ce n'est pas un artifice utilisé par les archivistes, il intervient parfois lors de l'édition d'ouvrages patrimoniaux.



Une image originale en niveaux de gris et son traitement en bichromie. Elle apparaît beaucoup plus attractive dans la publication, et une manipulation habile des courbes d'encrage permet d'améliorer le rendu des détails.

Image 15 Une image originale en niveaux de gris et son traitement en bichromie

- Le mode TLS (Teinte-Luminosité-Saturation) est utilisés comme interface par des applications de bureautique, mais le fichier enregistré est un fichier RVB
- Le Web utilise des codes littéraux pour décrire les couleurs incluses ou hors de la "palette système". Dans les applications de retouche photo, ce code est indiqué dans la fenêtre de dialogue du choix de la couleur. On peut indiquer ce code directement dans l'éditeur html, par exemple pour appliquer à un fond de page la couleur de fond d'une image qui y est intégrée. Le spectre de ces couleurs est celui du RVB (écran)

2.4 2.4. Analyse et typologie des documents à numériser

Ce paragraphe analyse les types de documents que l'on peut rencontrer en vue de la numérisation.

A chacun de ces documents correspond un procédé physique de numérisation et un mode colorimétrique apte à traduire ses caractéristiques.

Pour déterminer ces procédés, il convient de faire une analyse générale de l'aspect du document et de déterminer sa catégorie ou typologie.

2.4.1 2.4.1. Analyse du document

La forme du document à numériser doit être analysée avant toute numérisation :

- son format (sa taille) déterminera le type de numériseur à utiliser,
- le nombre de pages déterminera l'automatisation du procédé,
- son aspect déterminera le mode colorimétrique à utiliser qui permettra de traduire ses caractéristiques.

2.4.2 2.4.2. Nature des documents déterminant la colorimétrie

Du point de vue de la numérisation et, par conséquent, de la **colorimétrie à utiliser**, la typologie des documents se décline en grandes familles:

- manuscrits en feuilles détachées, registres,
- textes imprimés, dupliqués ou dactylographiés,
- gravures au trait, illustrations au trait, dites "estampes" dans le monde des archives,
- simili gravures des ouvrages imprimés,
- héliogravures, cartes postales en phototypie
- lithographies et chromolithographies, affiches
- photographies, cartes postales photographiques
- diapositives, négatifs sur films ou plaques, microformes
- dessins originaux et peintures,
- plans et dessins techniques,
- documents imprimés rassemblant du texte, des gravures, et des similigravures, voire des lithographies ou chromolithographies.

Les manuscrits et textes imprimés sont relativement faciles à numériser, quel que soit le rendu que l'on ait choisi pour les restituer (trait, niveau de gris ou couleur).

On s'apercevra par la pratique que les fonds iconographiques (photos, négatifs, diapositives, plaques de verre, peintures, chromolithographies, héliographies, cartes postales, affiches, plans en couleurs,..) sont aussi relativement aisés à numériser, car ils font appel à des modes en demi-tons, faciles à retoucher et à adapter à l'usage à l'aide des logiciels de retouche photo. Les difficultés apparaissent dès lors que l'on a affaire à des imprimés. Les pages des ouvrages et catalogues contiennent souvent des typologies mélangées (texte, gravures au trait, photos en couleur ou niveau de gris); l'idéal consisterait à numériser chaque page dans le mode qui lui convient le mieux, puis d'assembler ces sous-fichiers en un seul pour reconstituer le document intégral. Mais ceci est fastidieux pour la reproduction d'une reproduction ! Les progrès des numériseurs, et surtout des moyens de transmission des fichiers lourds ont permis de procéder systématiquement dans le mode le plus favorable à la reproduction des images. Les artifices de compression et déramage font le reste pour obtenir un fichier portable et agréable à lire à l'écran

Les difficultés apparaissent dès lors que l'on a affaire à des procédés de reproduction imprimés comme la gravure au trait, la simili gravure et autres procédés photomécaniques d'imprimerie pour lesquels la réussite se décide lors de l'opération de numérisation. Les spécificités relatives à ce traitement nécessitent de la pratique et de la circonspection. Les retouches ultérieures sont malaisées, voire impossibles.

Enfin les plans au trait sont assez faciles à traiter, car les systèmes de numérisation bitonale utilisés actuellement dans l'architecture et l'industrie sont parfaitement adaptés. Le principal problème que l'on risque de rencontrer est celui de leur format, lorsqu'il dépasse la largeur de passage du numériseur, car les normes définissant les formats de plans ont été vraiment appliquées à partir des années 50. Mais attention, les beaux jours sont comptés, avec l'apparition du dessin assisté par ordinateur, l'utilité de numériser un plan disparaît et les numériseurs à plans disparaîtront aussi du marché dans un délai plus ou moins court...

Nous avons appris quelques notions essentielles, notamment combien le type de document à numériser influe sur le procédé de numérisation ou la colorimétrie .

Mais bien évidemment **il n'est pas question de numériser n'importe quel document**, encore faut-il **se demander pourquoi numériser**.

Pour répondre il faut tenir compte:

- de différents objectifs,
- de l'usage du fichier obtenu,
- de la portabilité du document.

3.1 3.1. Objectifs

Les objectifs de la numérisation peuvent être:

1. un moyen de sauvegarde d'originaux menacés de dégradation physique (manipulations d'originaux fragiles et processus de vieillissement naturel),
2. la diffusion imprimée en haute définition (édition du livre),
3. la diffusion en ligne ou dans un réseau interne d'un document par ailleurs correctement conservé,
4. la destruction des documents papier sans valeur archivistique ni graphique, pour réduire le volume d'archives physiques (numérisation de substitution à des fins de gestion d'archives courantes et intermédiaires).

Chacun de ces cas va mener à des approches différentes, influençant directement la **résolution et la fidélité** des documents numériques.

La résolution concerne la finesse d'analyse du document, c'est-à-dire la possibilité de montrer les détails les plus fins.

La fidélité concerne la colorimétrie et donc l'aspect du document. La dégradation de cette colorimétrie affecte la fidélité si on utilise des formats d'enregistrement avec pertes de données (réduction de la palette des couleurs ou si l'on choisit de reproduire en mode bitonal des documents comportant des couleurs).

On choisira:

- une basse ou moyenne résolution et on tolérera des pertes de données pour les objectifs des catégories 1, 2, 3 et W définies ci-après (voir 3.4),
- la haute résolution et une grande fidélité des couleurs pour l'objectif de la catégorie 4.

3.2 3.2. Usage du fichier final

L'utilisation du fichier final est déterminante pour choisir la qualité de la numérisation, mais là encore, le discernement de l'archiviste doit s'exercer en prévoyant les utilisations futures. On ne pourra pas transformer un document de basse résolution en haute résolution, tout comme on ne peut pas éditer un CD haute fidélité avec l'enregistrement d'un concert diffusé par un poste à transistor !

3.3 3.3. Portabilité

La bonne numérisation va être basée sur des formats d'enregistrement compressés dont la taille naturelle ne nécessitera pas de compression ultérieure, ni sur le disque de stockage, ni pour l'envoi en réseau.



Attention

Il est à noter que les fichiers image compressés sont parfaitement optimisés et incompressibles. Il est donc inutile de refaire une compression Zip sur des fichiers jpeg, PDF et tiff dont les options de compression ont été activées à l'enregistrement.

Il est aussi inutile de compresser les pièces jointes d'un courriel, car c'est l'application dédiée qui s'en charge avant l'envoi (compression Zip en général).

3.4 3.4. Répartition des documents numérisés selon l'usage

Dans ce qui suit, nous parlons de l'usage des fichiers, transformés ou non par un travail de post-production, c'est à dire tout ce qui est relatif à la préparation des fichiers maîtres et de leurs copies pour la distribution ou l'usage final. (voir chapitre 5-10)

Les catégories ci-après ne sont pas des standards internationaux, mais des familles rencontrées dans la pratique et nommées pour la compréhension de ce cours

Selon nos besoins et l'usage que nous voulons en faire, nous pouvons répartir les documents à numériser en différentes catégories, sachant que certains fichiers maîtres peuvent servir à générer des documents de plusieurs catégories à la fois:

- textes pouvant simplement être lus sans exigence graphique,
- textes et images pouvant être consultés sans exigence graphique,
- ouvrages ou documents en téléchargement,
- documents iconographiques numérisés pour la sauvegarde ou la reproduction imprimée,
- documents iconographiques pour la toile (Web).

Catégorie 1 : textes pouvant simplement être lus sans exigence graphique

C'est le cas typique des **documents administratifs ou d'affaires** que l'on met en ligne ou que l'on archive numériquement pour une conservation correspondant à la durée d'utilité administrative. Pour ces documents, le critère principal est la portabilité (petite taille) permettant un flux rapide dans les réseaux et la diffusion par Internet. L'impression de ces documents a l'aspect d'une télécopie en qualité normale. Il n'y a que du noir et du blanc, c'est une reproduction **au trait ou bitonale**. Idéalement, ces documents ont fait l'objet d'une indexation, c'est à dire une reconnaissance des caractères.

Catégorie 2 : textes et images pouvant être consultés, sans exigence graphique

Ce sont des **documents** que l'on met en ligne pour permettre une consultation rapide par des internautes ou utilisateurs d'un réseau interne (intranet). On accepte une dégradation des images par la basse définition et par la compression. Ce sont des documents qui sont acceptables à l'écran mais dont l'impression est médiocre. On notera comme exemple des aperçus d'ouvrages à consulter au centre d'archives, des comptes-rendus et bulletins de liaison en ligne. C'est par essence, de la documentation dans son acception la plus générale.

Entrent aussi dans cette catégorie les documents que vous composez avec un logiciel de publication d'amateur auxquels sont intégrées des images de documents d'archives à titre iconographique accessoire.

Catégorie 3 : documents ou ouvrages ou en téléchargement

Certains centres d'archives ou bibliothèques mettent en ligne des archives publiques ou des ouvrages anciens libres de droits, pour permettre un téléchargement et une impression par l'internaute; une perte de définition est admise sur les images, mais leur lisibilité est conservée. C'est une situation intermédiaire entre le simple texte et la qualité impression.

Cette même technique peut être utilisée pour la mise en consultation en réseau intranet de documents dont plusieurs services ont un usage courant. Il est possible de protéger individuellement ces documents par des clés et mots de passe avec certains formats d'enregistrement.

Catégorie 4 : documents iconographiques numérisés pour la sauvegarde ou la reproduction imprimée

Les documents dont on envisage de faire une copie de sauvegarde numérique, soit parce qu'ils sont menacés de destruction physique, soit parce qu'ils se dégradent, doivent être numérisés sans perte de données, et dans les meilleurs critères de qualité applicables à leur typologie.

Pour la copie de sauvegarde des iconographies sujettes à dégradation naturelle (photographies en couleurs), on choisira des modes colorimétriques avancés pour conserver les nuances d'origine. La reproduction des négatifs et diapositives devrait permettre au moins l'impression sur papier en 18x24 cm à 300dpi.

La Direction des Archives de France recommande la conservation dans une taille qui permet la restitution dans une taille 30x40 cm en 300dpi. La performance des systèmes de stockage actuels permet cette taille.

Catégorie W : documents iconographiques pour la toile (Web)

Cette catégorie englobe les images à basse résolution que l'on rend disponibles en ligne pour les internautes, l'enseignement etc...et les images utilisées comme icônes d'accès aux fichiers haute résolution d'une photothèque en ligne.

Les formats de fichier sont des formats « image » et ce ne sont pas des documents multi-pages : chacune des images est traitée à part pour l'intégration à une page internet. La résolution est basse et l'on accepte des dégradations importantes par la compression. Ces fichiers sont constitués à partir d'une copie en résolution **18**

réalisée à partir du fichier maître

3.5 3.5. Usages mixtes

Plusieurs de ces catégories peuvent être applicables simultanément à un document faisant l'objet d'usage multiples. C'est le cas pour une exposition à thème pour laquelle on utilisera une même image dans:

- une brochure en couleurs,
- des panneaux
- un site web avec téléchargement de documents et d'images.

La numérisation se fera en une seule fois dans **les plus hauts critères de qualité**, puis on générera des fichiers de moindre qualité pour les diffusions en ligne, par l'intermédiaire d'un logiciel de traitement des images. Cette opération couramment le sous-échantillonnage de l'image.



Attention

Il est cependant inutile de faire de la sur-qualité dans la numérisation des documents, car les espaces disque sont inutilement dévorés et le processus se trouve alourdi pour un résultat invisible. Chaque fois qu'on double la résolution, on multiplie la taille du fichier par quatre. Il est inutile de numériser, par exemple, les photos dans une résolution plus de deux fois supérieure à la taille du grain argentique.

4. Formats d'enregistrement des images numérisées

4

Les formats d'enregistrement et les compressions choisies vont déterminer la taille (ou poids) du fichier, sa portabilité en ligne et son universalité de reconnaissance par les applications .

Il existe des formats d'enregistrement sans réduction de qualité de l'image, (sans pertes de données) et d'autres avec réduction de qualité (avec perte de données).

Les pertes de données par compression résultent uniquement en une réduction du nombre de couleurs du fichier, mais jamais du nombre de pixels.

Les fichiers destinés à l'«archivage permanent»

Lors de l'acquisition avec le numériseur, l'image est entrée dans l'ordinateur avec la qualité maximale que l'on aura définie lors de cette acquisition. Il faut la conserver parce que reprendre un original est plus coûteux que le prix de l'espace disque.

Lorsque l'on acquiert une image il faut tout d'abord enregistrer le fichier d'acquisition dans un format n'entraînant pas de pertes de données. Ce fichier sera ensuite travaillé dans une phase dite de post-production comprenant le renommage, les éventuelles retouches colorimétrique et l'entrée des métadonnées. Le résultat sera de nouveau enregistré sans pertes de données pour donner le fichier maître destiné à l'archivage permanent.

Les copies dites de diffusion, ou d'exploitation

Du fichier **maître**, des **copies** de taille réduite seront générées de façon mécanique, pour le placement en ligne, afin de limiter les temps de téléchargement. Les métadonnées sont migrées elles aussi automatiquement depuis le fichier maître.

On pourra par la suite renouveler les copies pour la mise en ligne, en augmentant leur qualité au fur et à mesure de l'apparition de réseaux à haut débit.

4.1 A propos de ce chapitre des formats



Remarque

Ce chapitre peut sembler long et indigeste pour l'apprenant. C'est pourtant toute la réussite d'un plan de numérisation qui en dépend. Face à l'archiviste, des commerciaux peu scrupuleux peuvent proposer des prestations au résultat dégradant, et au final l'argent est dépensé pour un résultat médiocre. Revenez sur ce chapitre par la suite si dans un premier temps il vous paraît trop ardu, mais faites d'abord un petit passage dans le chapitre qui suit, avec ses définitions.

4.2 A. Les propriétés des formats d'enregistrement

Les propriétés des formats d'enregistrement

Le chapitre traite des formats d'enregistrement, ils ont un rôle essentiel dans le produit fini. Actuellement, il n'y a plus de machines de numérisation à performances médiocres, ce n'est que le format d'enregistrement qui décide de la conservation de la qualité du produit.



Définition

Dans les pages qui vont suivre, et déjà au début de ce chapitre, des termes nouveaux apparaissent et pourraient mettre le lecteur en difficulté. Mettons donc à plat ces termes pour bien comprendre le contenu des différents formats d'enregistrement.

Le format d'enregistrement

Le format d'enregistrement est l'ensemble des caractéristiques du fichier qui lui permettent d'être reconnus par les applications. La manière d'écrire le début et la fin de fichier selon certains codes déclenche le fonctionnement adéquat de l'application qui lit ou transforme ce fichier. Il existe des formats « propriétaires » dont les codes sont typiques à un éditeur, et des formats « universels » ou « libres » dont les codes sont divulgués publiquement. Nous les choisirons pour cette raison.

4. Formats d'enregistrement des images numérisées

Le fichier maître (destiné à l'archivage permanent)

C'est le premier fichier enregistré sans perte de données et qui contient tout ce qui est nécessaire à la conservation de l'archive numérique et à la génération des sous-fichiers de diffusion: image, corrections et métadonnées.

- **L'image** est l'objet informatique résultant de l'acquisition brute par le numériseur.
- **Les corrections** sont des réglages de densité de l'image, de colorimétrie par exemple dans le cas de la restauration de diapos aux couleurs dégradées par le temps. Les corrections sont temporaires pendant la durée de post-production, puis appliquées de façon permanente après contrôle par un expert.
- **Les métadonnées** sont les textes au format XML introduits dans le fichier image. Elles n'apparaissent pas sur l'image, mais dans les éditeurs de métadonnées ou les légendes automatiques de certains environnements. Dans le cas d'une image restaurée, ces métadonnées signalent le fait de la restauration.

Le fichier de diffusion ou d'exploitation

La caractéristique principale de ce fichier est la portabilité (taille réduite pour diffusion en réseau). L'on cherchera toujours le format le plus efficace pour cette fonction. Le PDF (portable document format) est très avantageux, il permet d'assembler plusieurs pages d'images avec un descriptif ne nécessitant pas d'application spécifique à la lecture des métadonnées

Le calque

Le calque est une image ou une correction se superposant à d'autres calques dans un même fichier image. Lorsque l'on veut obtenir une image « plate » définitive, tous les calques sont fusionnés (confondus en un seul). Certains formats enregistrent les calques, d'autres pas.

Calque de réglage

Un logiciel de retouche photo peut générer des calques de réglage d'image pour changer les caractéristiques de densité, de contraste, de colorimétrie... de l'image placée en-dessous. Un fichier enregistré avec ces calques permet le changement de ces valeurs à l'infini, sans altérer l'image originale qui est la à base du fichier.

La couche Alpha

La couche alpha est une série de pixels transparents juxtaposés à des pixels colorés d'opacité 100%. Dans le format GIF, cela permet de détacher des objets sur un fond coloré dans une page internet. La limite est brutale, et le contour présente un aspect crénelé.

La transparence

La transparence permet d'enregistrer des parties d'images dont l'opacité n'est pas 100%. Il y a peu d'application de la transparence en archives, mais lors du détournage d'une carte postale par exemple, cela permet de ne pas avoir des bords crénelés comme pour la couche alpha, le bord est progressivement transparent sur 2 ou 3 pixels.

La compression

C'est l'action qui consiste à réduire la taille du fichier enregistré par un codage spécifique à chaque algorithme de compression. La compression peut se faire avec ou sans pertes de données. Une compression sans pertes de données réduit la taille de 20 à 30% et jusqu'à 50% pour la plus performante. La compression avec perte de données peut réduire la taille de 95%.

L'algorithme

C'est le modèle mathématique qui permet à l'ordinateur de réaliser la réduction du nombre de données à enregistrer sur le disque. Tous les blancs et fins de ligne sont supprimés dans le fichier encodé qui a l'aspect d'un bloc en continu constitué de signes incompréhensibles. On n'y reconnaît même pas les textes des métadonnées...

La perte de données

Un terme un peu choquant prend place dans notre propos : la compression avec perte de données. Ce procédé consiste à réduire l'échantillonnage des couleurs, c'est à dire le nombre de couleurs contenues dans l'image, mais pas le nombre de pixels. Il en résulte une taille réduite du fichier enregistré et des artefacts visibles lors d'une compression importante. C'est un procédé lié à un format développé par le Jointed Photographs Expert Group (JPEG). Les métadonnées ne sont pas perdues lors de la compression du fichier.



Attention

En aucun cas les formats engendrant des pertes de données ne doivent être utilisés pour assurer la copie de sauvegarde d'archives en voie de détérioration physique.

Les métadonnées

Les métadonnées sont des textes qui sont insérés au fichier image, mais qui n'apparaissent pas visuellement sur celle-ci. Les métadonnées sont enregistrées au format d'échange XML et prennent les aspects que le lecteur de métadonnées leur donne. Les métadonnées sont distribuées dans des champs reconnus par des normes internationales dont le standard le plus connu est IPTC, créé à l'origine pour les photographes de presse internationale. Les appareils de numérisation complètent automatiquement les champs de métadonnées EXIF qui décrivent les conditions de prises de vues

4. Formats d'enregistrement des images numérisées

La production

C'est la phase où l'on capture l'image de l'original et qu'on l'enregistre une première fois en vue de la suite des opérations.

La post-production

Cette phase comprend toutes les opérations de traitement d'images et d'insertion de métadonnées. A l'issue de ces opérations, le **fichier maître** est constitué avec un enregistrement sans pertes de données.

Le stockage

Le fichier maître constituera une **archive permanente**. Mais au cours de son existence en centre d'archives, les supports physiques vont évoluer. Pour pallier leur obsolescence, il faudra régénérer les fichiers en les recopiant sur de nouveaux supports (Disques optiques, serveurs sécurisés, disques durs, bandes magnétiques...)

4.3 4.1. Les qualités des formats pour les besoins archivistiques

Il existe une grande variété de formats informatiques pour l'enregistrement des images numérisées. Le but de l'archiviste est de choisir un format :

- pérenne
- libre, ouvert, et décrit par une norme
- universel, lisible sur toutes les plateformes
- pouvant être lu par des logiciels anciens en fichier de diffusion (compatibilité descendante des copies)
- pouvant contenir des métadonnées

A ces préceptes, on pourrait presque ajouter : le format de diffusion devrait être différent de celui du fichier maître d'archivage, permettant de le reconnaître d'un seul coup d'œil dans les répertoires.



Remarque

Dans cette proposition, le format est celui qui correspond à l'objectif, on n'utilisera pas une « recette universelle » pour la comptabilité et pour les reproductions de tableaux, le résultat ne serait bon à rien mais mauvais à tout !

Le format pérenne

C'est presque une hérésie de parler de format pérenne dans un monde changeant comme celui de l'informatique. Cependant, il existe des « stabilisateurs » de format, qui sont le nombre des usagers ayant utilisé tel ou tel format sur des quantités importantes de fichiers.

Au nombre de ceux-ci, on reconnaît les formats Tiff, Jfif (improprement appelé jpeg), PDF et png

Le format du domaine public

Le format choisi ne doit pas être propriétaire, c'est à dire que ses clés de développement doivent être publiques et non soumises à droits d'exploitation limités. C'est le cas des formats dont nous allons parler.

Le format universel

Les formats cités dans ce cours ont tous « élu domicile » sur les plateformes Windows-Mac-Unix-linux, et sont lisibles parce que leurs formats ont peu changé avec le temps. Dans le cas du PDF, un format spécial « archives » a été créé et répond à la norme ISO 19005. Cette démarche vise à garantir une lecture dans le futur par une stabilisation du standard, qui cependant en est déjà à sa troisième version depuis sa création...

La compatibilité descendante

Les formats choisis doivent aussi pouvoir être lus par des logiciels datant de quelques années, ou des logiciels grégaires comme les lecteurs d'images intégrés à Windows. Cette compatibilité sert aussi à faire fonctionner des banques de données propriétaires ou libres qui se basent sur des standards anciens pour éviter le paiement de brevets.

Le contenu de métadonnées

Les formats devront pouvoir contenir des métadonnées au langage XML pour que lorsque l'image est détachée de son site de diffusion, l'utilisateur puisse toujours l'identifier complètement sans risque d'intervention lors de la copie (fastidieuse) de légendes en ligne. Les métadonnées sont une part essentielle du contenu du fichier maître. Le choix des champs complétés doit être judicieux pour conserver une compatibilité descendante.



Complément : Particularités des formats « photo »

L'avantage des formats d'images courants est qu'ils sont destinés à l'échange, donc en principe stables, compatibles et ouverts.

Une particularité existe cependant pour le format « Adobe Photoshop® » qui est un format propriétaire de la société Adobe®. Ce format mondialement reconnu offre une option de compatibilité descendante à l'enregistrement dans le logiciel Adobe Photoshop® (maximiser la compatibilité). Il est par ailleurs reconnu par nombre de logiciels l'il 22

4. Formats d'enregistrement des images numérisées

d'autres éditeurs à cause de son omniprésence dans le milieu de l'image. Il reste cependant la propriété de son créateur qui le fait évoluer en permanence.

On notera aussi une exception avec les nouveaux formats « raw » des appareils photo et chambres numériques, qui peuvent nécessiter un filtre de rendu propriétaire à chaque constructeur, livré avec l'appareil de prise de vues. Le format « raw » est un peu comme les anciens négatifs argentiques, il est nécessaire d'ajuster les valeurs lors du « tirage » de la photo. Les clichés réalisés dans ces formats devront donc être convertis en formats universels pour l'archivage permanent.

4.4 4.2. Formats sans perte de données

4.2.1 Aperçu des formats sans pertes de données

Les formats sans pertes de données sont à utiliser **pour constituer la banque de fichiers maîtres destinés à la conservation et/ou à la sauvegarde.**

Ils sont pourvus d'algorithmes de compression, qui réduisent la taille du fichier enregistré, mais le restituent sans perte de données, ni en couleurs, ni en pixels.

Les formats sans pertes de données peuvent être utilisés sans activer leurs moteurs de compression, mais c'est stupide car cela amène à une occupation importante d'espace disque et ralentit les échanges avec le serveur distant. Les formats modernes ont une activation automatique de la compression (cas du png).

La caractéristique typique de ces formats est bien celle des algorithmes de compression. Nous allons en faire le tour avant d'entrer dans les formats, car dans certains cas ils sont communs à plusieurs d'entre eux.



Acquisition de la carte postale sur fond noir

Résultat du détournement et enregistrement au format png

Exemple de traitement des cartes postales anciennes: l'acquisition se fait sur un fond noir mat, puis ce fond est éliminé avec un outil « baguette magique » qui se trouve dans la plupart des logiciels de retouche photo. La photo est enregistrée avec un fond transparent au format png, ce qui permet de conserver ces propriétés de transparence, et sans pertes de données pour le fichier maître. Cette CPA colorisée sur une similitude comprend des trames qui peuvent légèrement moirer à l'écran à certaines échelles.



Cartes postales au format png

4.2.2 Les algorithmes de compression

Il existe plusieurs algorithmes de compression qui peuvent coexister au sein d'un même format, dans ce cas, le choix du procédé sera à faire par l'utilisateur lors de l'enregistrement. De même, on retrouve les mêmes « compresseurs » dans plusieurs formats.

LZW Lemple-Zif-Welch du nom de ses inventeurs est utilisé dans les formats TIFF, PDF, GIF et est performant essentiellement lorsque les images contiennent de grandes zones monochromes. Mais comble de l'ironie, il augmente la taille du fichier lorsque l'échantillonnage est supérieur à 8 bits par couche, parce qu'il n'est pas prévu pour cet usage. Il appartient maintenant au domaine public.

4. Formats d'enregistrement des images numérisées

pour des images présentant de grandes zones monochromes, mais il est efficace sur les fichiers à plus large échantillonnage de nuances (16 et 32 bits).

CCITT groupe 4 (Comité Consultatif International Télégraphique et Téléphonique) est utilisé pour les images bitonales, pris en charge par les formats de fichier PDF et de langage PostScript.

Deflate-inflate est un algorithme inventé par Phil Katz pour le format de compression PKZIP en y associant le codage Huffman. Il est utilisé librement dans le format PNG, lui aussi développé par Phil Katz et dans le format Zip actuel (PKZIP pour Phil Katz ZIP). Les codes sources sont accessibles.

a) 4.2.3 Les formats d'enregistrement sans pertes de données

Le format TIFF

Le format **TIFF** (*Tagged Image File Format*) est plus particulièrement utilisé pour les documents mono-pages, les photothèques. C'est le plus connu et répandu avec son extension .tif.

Il s'avérera aussi idéal lors des phases de post-production pour l'enregistrement transitoire avant archivage définitif, par son aptitude à conserver les calques image et calques de réglage, avec leurs gradations. C'est le format utilisé en majorité par le monde de l'édition imprimée.

Une option multi-pages du format **TIFF** existe et est très usitée en Allemagne, mais ne peut pas être lue par tous les logiciels, aussi faudra-t-il en faire un usage mesuré.

PNG

Le format PNG (Portable Network Graphics) -prononcer peng- est un format libre et qui propose la plus grande universalité et un des meilleurs potentiels en termes de gain d'espace disque grâce à sa compression automatique « deflate ». Il reproduit les transparences de l'image et la transparence totale par couche alpha (pixels transparents), mais n'enregistre pas plusieurs calques. Il peut contenir des métadonnées qui ne sont pas toujours affichées par les éditeurs de logiciels propriétaires. Il n'est pas encore très répandu dans le monde des archives, mais des centres précurseurs l'introduisent dans leurs photothèques. Il est reconnu par un très grand nombre d'applications sur toutes les plateformes, et les systèmes libres (Linux).

JPF, JP2

Le format Jpeg 2000 étendu JPF est une version plus complète que l'ancien JP2 auquel il emprunte le moteur de compression avec ou sans perte de données, au choix de l'utilisateur. Le JPF montre les performances les plus intéressantes en matière d'épargne d'espace disque avec son option « sans pertes de donnée » activée. Le principal inconvénient est que son usage n'est pas encore très répandu bien que sa pérennité soit assurée, vu qu'il est normalisé. Le .jpf est encore illisible dans de nombreuses applications, vu que c'est un format récent qui n'est pas encore inclus par défaut aux filtres import-export des applications. Il est cependant très attractif pour l'archivage permanent des fichiers maîtres aplatis (en un seul calque) avec une taille pouvant atteindre la moitié de celle du png ou du Tiff Zip.

Ce format est étudié plus en détail dans la section « formats avec pertes de données » parce que c'était sa destination initiale.

PDF

Le format PDF (*Portable Document Format*) est plus particulièrement pertinent pour les documents multi-pages et les documents informatiques générés par les applications. L'enregistrement sans perte de données est à choisir par l'utilisateur dans les préférences avant de commencer les enregistrements. Avec le format PDF-A (PDF Archives) il offre une option de pérennité certaine. Il permet aussi d'enregistrer des pages textes et des pages images dans le même fichier.

Nota :Le PDF peut aussi être utilisé avec une compression engendrant des pertes de données, ce qui est intéressant pour la mise en ligne. Ce format est étudié plus en détail dans la section « formats avec pertes de données » parce que c'est sa destination première.

PSD

Le format PSD de Adobe Photoshop n'est pas un format d'archivage à proprement parler, mais on peut l'utiliser dans les copies de travail intermédiaires de post-production, lorsque des calques de réglage sont utilisés temporairement pour le rendu de l'image, ou pour d'autres opérations demandant des fonctions avancées (détourage, redressements, retouches non irréversibles...) Les formats photoshop les plus récents sont enregistrés avec une compression sans perte de données.

GIF

Le format GIF (Graphic Interchange Format) ne sera pas étudié ici, il a été avantageusement remplacé par le format PNG avec des fonctionnalités supplémentaires.

4. Formats d'enregistrement des images numérisées

4.2.4 propriétés intrinsèques des formats sans pertes de données

Potentiels de contenu des fichiers en phase d'enregistrement de la capture du numériseur					
Propriétés	TIFF	PNG	JPF	PDF	PSD
Modes Couleur	RVB, LAB	RVB	RVB	RVB	RVB, LAB
Mode Niveaux de gris	oui	oui	oui	oui	oui
Mode bitonal	oui	oui	non	oui	oui
Profondeur d'échantillonnage des couleurs par couche	1 à 32 bits	1 à 16 bits	8 à 16 bits	1 à 16 bits	1 à 32 bits
Type de compression sans perte de données et bits par couche	LZW 1 à 8 bits ou Zip 1 à 32 bits	« deflate » tous modes	Jpeg 2000 option «sans pertes»	Zip (gris-couleur) CCITT ou Zip en mode bitonal	Compression genre ZIP
Choix de la compression	utilisateur	automatique	utilisateur	utilisateur	automatique
Forme d'enregistrement classique	Enregistrement « plat », sans calques, à la sortie du numériseur				

Tableau 1 Potentiels de contenu des fichiers en phase d'enregistrement de la capture du numériseur

Potentiels de contenu et d'usage des fichiers en phase de post-production					
Propriétés	TIFF	PNG	JPF	PDF	PSD
Enregistre les calques de réglage	oui	non	non	non	oui
Enregistre plusieurs calques image	oui	non	non	Sur option	Oui
aplatit en calque image unique avec transparence		oui	Oui, transparence sur option	Sur option	
Enregistre les métadonnées IPTC, EXIF, XMP...	oui	oui	Oui sur option	Oui	oui
Options spéciales		Entrelacé pour affichage web	Compression avec perte de données possible dans ces formats	Conserve tracés vectoriels et caractères	
Forme d'enregistrement classique	Enregistrement avec les calques de réglage tant que dure la phase de traitement, puis aplatissage à la fin.				

Tableau 2 Potentiels de contenu et d'usage des fichiers en phase de post-production

4.5 4.3. Formats avec pertes de données

4.3.1 aperçu des formats

Certains formats d'enregistrement entraînent des pertes de données. Le moteur de compression utilisé par ces formats réduit le nombre de couleurs mais pas le nombre de pixels. La compression est destinée à réduire la taille des fichiers pour accélérer le téléchargement et la taille des documents finis.

Le format **JPEG** est destiné à la diffusion d'images à l'aide un taux de compression variable. On choisit un bon rapport qualité-taille pour l'usage que l'on veut faire du fichier. Il est capable de millions de couleurs.

Le format **PDF** avec perte de données est un format de diffusion acceptant des documents multi-pages pouvant contenir des images, des textes, des dessins. Lorsqu'il est utilisé à l'enregistrement d'images avec pertes de données, il contient les formats de la famille JPEG étudiés ci-après. Le choix du mode et du taux de compression est fait par l'utilisateur, et le programme prévoit des cas d'usage prédéfinis.

Le format **GIF** est typiquement conçu pour la toile, mais peut rendre des services pour le traitement de documents monochromes. Il est limité à 256 couleurs, ou à une image bitonale. La réduction du nombre de couleurs intervient lors de l'enregistrement, c'est pour cela qu'on le considère comme un format avec pertes de données bien que sa compression LZW n'en génère pas intrinsèquement. Il est progressivement abandonné au profit du jpeg.

Le format **JBIG** est destiné au codage des fichiers au mode bitonal émis par les télécopieurs. Son usage a pu être proposé par certains prestataires, mais il n'apporte rien par rapport aux TIFF et PDF avec compression CCITT groupe 4.

4.3.2 les formats avec pertes de données

JPEG -JFIF

Le format JPEG désigne en réalité plusieurs formats développés par le Joint Photographs Expert Group. Le plus répandu, que l'on peut trouver avec les extensions .JPEG, .jpg, .jpe, .jfif, .jfi. est un format jfif pour JPEG File Interchange Format qui a été créé pour permettre la diffusion sur toutes les plates-formes et en ligne des images numériques en couleur. Il répond à la norme ISO/CEI 10918-1 (JPEG baseline) ou UIT-T recommandation T81, et son code est ouvert et non propriétaire. Son algorithme permet une compression progressive en 10 à 12 niveaux selon les logiciels, avec une dégradation pouvant varier d'invisible (niveau 12) à importante (niveau 0).

Le format JPEG supporte les modes niveau de gris, couleurs RVB et CMJN. On peut y introduire des métadonnées, et un profil de couleur. Ce format ne supporte pas le mode bitonal (bitmap) ni les transparences.

En archivistique, son usage sera limité à la production des sous-fichiers à placer en ligne, et à la numérisation de cartes de grand format pour limiter la taille du fichier, bien que dans ce cas, le fichier JPEG 2000 étendu soit préférable par sa performance, pour l'établissement du fichier maître



Attention

Le premier enregistrement en JPEG, même en qualité maximale, réduit déjà imperceptiblement la gamme des couleurs de l'image.

Le ré-enregistrement avec le même niveau de qualité peut aussi réduire la qualité de l'image selon les applications, car seul le format du fichier compressé est normalisé, mais chaque développeur peut développer une « moulinette » différente pour y arriver. Hormis le réenregistrement dans une qualité supérieure, cette manœuvre est déconseillée, elle pourrait se traduire par une réduction importante de la qualité après plusieurs réenregistrements.

La compression par le format JPEG réduit le **nombre des couleurs de pixels**, mais ne **réduit pas le nombre des pixels** formant l'image. Les niveaux de qualité «maximale» n'altéreront pas visiblement la qualité de l'image ; cependant les nuances de couleurs seront réduites, ce qui peut être gênant pour le traitement ultérieur de l'image en **post-production** (travail des couleurs, des densités).

Les niveaux de qualité «supérieure» commencent à altérer visiblement l'image si on fait un agrandissement, il apparaît des «plaques» de même couleur très gênantes dans les dégradés subtils (ciel) et dans les reflets sur les objets lisses.

Les niveaux de qualité «moyenne» accentuent les défauts précédents et font apparaître du «sable» à la frange des zones de contraste. Les niveaux de qualité «basse» font apparaître un genre de tissage de panier (effet de carreau) en accentuant les défauts cités précédemment. Des zones de halo entourent les contrastes.

Le procédé de compression consiste à créer des plages de pixels de couleurs approchantes et d'en faire une moyenne.

La figure ci-dessus montre un agrandissement des pixels de l'image: dans les faits, ces phénomènes sont beaucoup plus petits et souvent peu discernables à l'écran, par contre ils transparaissent à l'impression.



Image 16 Niveaux de dégradation des images JPEG

Cet ultime niveau de compression (0) ne peut être utilisée que pour de l'affichage de page Internet.

JPEG 2000 (JP2, JPF)

JPEG 2000 est un procédé de compression d'images développé aussi par le groupe de travail Joint Photographic Experts Group. C'est devenu une norme commune à l'ISO, la CEI et l'UIT-T sous la dénomination ISO/CEI 15444-1. Le fichier JPEG 2000 peut être enregistré avec ou sans perte de données, en utilisant une « transformée en ondelettes » (méthode mathématique de codage du fichier natif).

Les performances de JPEG 2000 en compression avec perte de données sont supérieures à celles de la méth 26

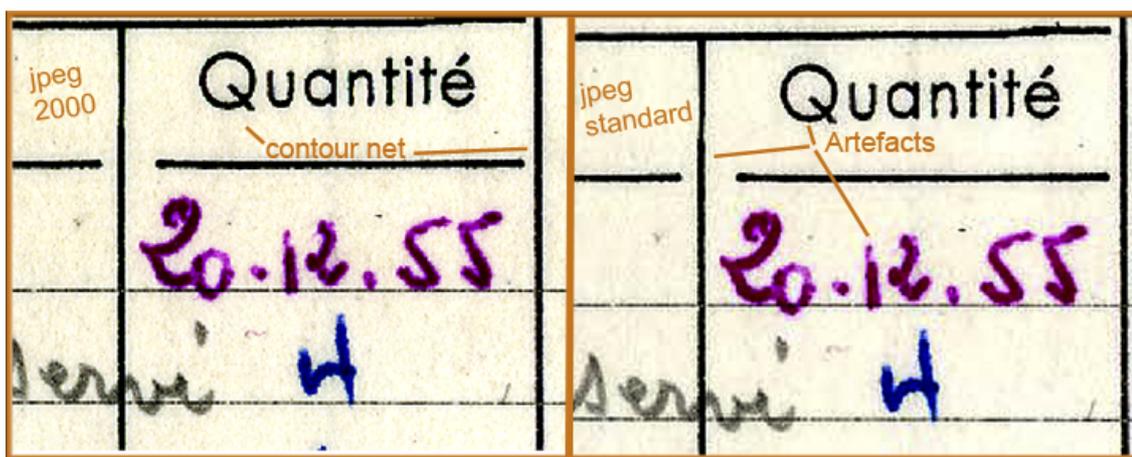
4. Formats d'enregistrement des images numérisées

compression JPEG standard. On obtient donc des fichiers d'un poids inférieur pour une qualité d'image égale. De plus, les contours nets et contrastés sont mieux rendus, le halo et l'effet de « sable » (artefacts) n'apparaissent plus. Un taux de compression de fichier de 60% en JPEG traditionnel (rapport de 0.4 : 1) est presque inévitablement générateur d'artefacts. Ils se traduisent de manière visible par la destruction de certains détails chromatiques et touchent la continuité des lignes obliques par effet de crénelage « aliasing ». Comparativement, un taux de compression supérieur à 80% en JPEG-2000 demeure visuellement excellent.

Le JEG 2000 contient plusieurs résolutions d'affichage possibles : on peut faire apparaître l'image dans un navigateur internet à différentes tailles en pixels, ce qui est un grand avantage pour moduler l'affichage en réseau en fonction du débit. Le webmestre n'installera qu'un fichier source et l'éditeur internet choisira la taille d'affichage nécessaire.

Les formats avec compression JPEG2000 supportent les modes niveau de gris, couleurs RVB et CMJN, et en plus par rapport au JPEG, le bitonal. On peut y introduire des métadonnées, et un profil de couleur.

Le format Jpeg 2000 est caractérisé par les extensions .JP2, .J2K, .JPC, .JPX



comparaison jpeg jpeg2000

Compression. A gauche le JPEG 2000 garde des contours sans artefacts, alors que le JPEG standard commence à montrer des artefacts autour des zones de grand contraste.

(agrandir pour bien voir les artefacts)

Dans des taux de compression très élevés (très basse qualité), les rendus tendent à être identiques. L'usage du JPEG 2000 à compression sera donc dédié à des fichiers pour lesquels on veut conserver une certaine qualité. Le JPEG 2000 a un très bon rendu sur les documents écrits, c'est pourquoi il est couramment proposé sur les copieurs numériques avec fonction numériser. Dans cet usage, le JPEG 2000 est encapsulé dans un PDF.

Le format Jpeg 2000 étendu

Le format JPEG 2000 étendu est une amélioration du précédent, en intégrant les outils indispensables à l'archiviste. Il répond à la norme ISO/CEI 15444-2:2004 et UIT-T Recommendation T.801 il permet en plus :

- l'intégration des métadonnées d'image XMP, EXIF, IPTC, sur option par l'utilisateur,
- la conservation d'un calque unique avec sa transparence (sans crénelage des bords francs) sur option de l'utilisateur (à défaut le calque est aplati sur fond blanc)
- l'insertion d'un profil de couleur ICC sur option par l'utilisateur
- l'utilisateur peut régler la taille du carreau sur lequel se fait la compression, pour optimiser la taille (plus le carreau est grand, plus la compression est efficace, parce qu'il y a plus de couleurs de pixels à mettre en commun).

Le format Jpeg 2000 étendu est caractérisé par l'extension .JPF

A propos des formats Jpeg 2000

L'usage « direct » du format JPEG 2000 peut poser certains problèmes actuellement parce que les logiciels capables de lire ce format ne sont pas encore très répandus. A part dans Mac OSX, il faut faire appel à des logiciels de conversion, libres ou propriétaires. Pour assurer la lisibilité sur toutes les plateformes, l'idéal est d'encapsuler l'image JPEG 2000 dans un fichier PDF, ce qui en assure la lisibilité sur toutes les plateformes et dans tous les environnements.

Le format JPEG 2000 natif est l'allié idéal du webmestre, parce qu'il contient plusieurs versions de l'image dans des tailles différentes, et une seule source est à placer sur le serveur. C'est la page internet qui fait appel à l'image à la taille désirée.



Attention

Le premier enregistrement en JP2 lorsque l'option « sans pertes » n'est pas cochée réduit déjà imperceptiblement la qualité de l'image.

Le réenregistrement avec le même niveau de qualité peut dans certaines applications chercher à appliquer de nouveau le même rapport de compression à partir du fichier qui avait déjà été compressé. Le résultat est ravageur. (par exemple si on avait voulu gagner 30% la première fois, en le faisant une deuxième fois, le fichier sera alors réduit de 51% de sa taille originale et la qualité s'en ressentira)



Complément : Amélioration des rendus du JPEG

Il existe des extensions (plug-ins) à installer dans vos logiciels de retouche d'image qui rendent lisses et sans «sable» ni «paniers» les compressions JPEG, dans une taille inférieure à la compression classique (par exemple: proJPEG de BoxTop software). Leur fenêtre de dialogue montre en détail la texture de l'image et l'utilisateur peut pousser son taux de compression en fonction du résultat attendu. Cette méthode peut être appliquée pour la conservation de grands fonds iconographiques pour gagner un espace disque conséquent par rapport au TIFF ou à une compression à taux fixe prédéterminé. Cependant cette solution ne peut être mise en œuvre que par un opérateur très averti et distinguant correctement les couleurs. Il peut être nécessaire de pratiquer un test de détection du daltonisme.

PDF en usage de diffusion

Il est ici nécessaire de parler du format PDF avec des pertes de données, car c'est sous cet aspect qu'il est majoritairement utilisé dans la diffusion des documents d'archives.

Le format PDF (portable document format) est un langage de description de page c'est à dire qu'il peut positionner divers éléments sur une page qui a une échelle de restitution connue en centimètres. Un des principaux avantages de ce format est que les fichiers au format PDF sont fidèles aux documents originaux : les polices, les images, les objets graphiques et la mise en forme du fichier source sont préservés, quelles que soient l'application et la plate-forme utilisées pour le créer. Il a été créé à l'origine pour le monde de l'édition imprimée, afin d'envoyer à l'imprimeur un fichier à restituer sans surprise. Bien sûr, dans ce cas d'usage, toutes les images sont enregistrées sans pertes de données.

Ce format peut contenir ces éléments divers :

- des images bitonales
- des images en tons continus (gris ou couleur)
- des dessins vectoriels
- du texte
- mais aussi des signets, hyperliens, des sons, des séquences vidéo.

A l'origine, la page est décrite selon le langage PostScript® Apple, puis les éléments sont pris un à un et compressés à l'aide des algorithmes les plus performants pour chacun d'entre eux.

Ainsi, les images peuvent-elles être encodées en JPEG-JFIF ou en JPEG 2000. Dans les applications générant un fichier PDF sur un ordinateur à partir du fichier source d'une application, il existe des fenêtres permettant de régler le taux de compression de l'image, mais aussi de réduire la résolution (taille en pixels) des images.

La taille en pixels des images se règle par un filtre passe-bas³ incorporé à l'application, avec une réduction de la résolution des images se trouvant au-dessus d'un nombre prédéfini de pixels par pouce à la taille de restitution. Une imprimante virtuelle PDF permet de produire cet artefact à partir d'un fichier PDF à haute résolution/sans perte de données et de générer des copies légères à basse résolution/avec pertes de données pour la mise en ligne.

Pour les acquisitions de séries de documents en mode bitonal, c'est la machine de numérisation qui donne accès aux réglages de résolution en clavier de commande ; la bitonalisation et la compression se font directement dans la machine et les pages sont livrées assemblées en un seul fichier PDF reproduisant le lot de chargement de l'appareil

4.3.3 Le format PDF-A

Le format PDF-A est un format spécialement dédié à l'archivage issu du format Adobe PDF et qui a été normalisé pour assurer la pérennité à long terme. Etant normalisé, ce format échappe à la propriété de son éditeur, et les codes sources deviennent accessibles et publics.

La spécification PDF/A a été publiée par l'ISO et est utilisée par les organismes d'archivage du monde entier pour garantir la sécurité et la fiabilité de la diffusion et des échanges de documents électroniques.

La principale différence avec le PDF de ressource Adobe est la notion de fichier autoporteur uniquement dédié à la conservation de documents « pages ». Il n'y a pas de possibilité d'introduire des URL, des liens vers des fichiers externes, des sons, des images animées comme dans le PDF classique. Le PDF-A est conçu pour la restitution identique et permanente du document sur toutes les plateformes et à tout moment de son évolution.

Le contenu de la norme ISO 19005 optimise l'indépendance matérielle et logicielle ainsi que l'auto-documentation. Il comprend la définition du format PDF/A, mais aussi la façon de développer un outil de visualisation de fichier conforme à ce format. Cela garantit la possibilité future de disposer d'un outil de visualisation. La norme ISO 19005 contient également le document « PDF reference manual » d'Adobe® systems

Par ailleurs, l'intégralité du format PDF dans la version 1.7 a été normalisée par l'ISO en juillet 2008 sous la référence ISO 32000. Cette version sert de base au PDF-A le plus récent

Évolutions du format PDF-A

On croyait lors de la création de la norme avoir fixé à jamais le PDF dans le marbre, mais en fait, le format PDF archives suit les évolutions du format PDF créé par Adobe, pour profiter des dernières ressources dynamiques, ceci afin de satisfaire aux exigences du monde des affaires.

La norme ISO 19005 évolue par adjonction de parties, qui définissent à chaque fois des compléments au format de base.

4. Formats d'enregistrement des images numérisées

Documents -- Format de fichier des documents électroniques pour une conservation					
Partie	Nom	Contenu	Date d'éditi	Norme ISO	Version PDF de base
Part 1	PDF/A-1	Utilisation du PDF 1.4 créé	2005	ISO 19005-1	PDF 1.4 (Adob référence PDF 3e édition, 2001)
Part 2	PDF/A-2	Utilisation du PDF	2011	ISO 19005-2	PDF 1.7 ISO 32000-1: 2008
Part 3	PDF/A-3	Utilisation de l'	2012	ISO 19005-3	PDF 1.7 ISO 32000-1: 2008

Tableau 3 ISO 19005 - Gestion de documents

On peut juste regretter que le normalisateur n'ait pas créé des champs de métadonnées avec des libellés adaptés au monde des archives : cote, dates extrêmes... mais l'évolution est toujours possible.

Devenu normalisé, le format PDF est omniprésent, et a un bel avenir devant lui. Chaque internaute y a accès avec le logiciel Adobe Reader® distribué en ligne, et disponible pour toutes les plateformes et versions de systèmes d'exploitation.

Au sujet de la protection

Le format PDF-A offre automatiquement à l'enregistrement une protection du document contre les modifications.

Pour le PDF standard, il existe des options permettant de protéger le fichier PDF contre la copie du contenu, voire l'impression, mais certains éditeurs de logiciels contournent ces protections. Le seul procédé résistant bien est l'encryptage et le mot de passe à l'ouverture, mais c'est peu utilisable dans un flux intense de documents en ligne.

On préférera réduire la qualité graphique du contenu si on veut empêcher l'utilisation des images en ouvrages imprimés.

4.3.4 propriétés intrinsèques des formats avec pertes de données

Le tableau ci-dessous dresse un inventaire des formats avec pertes de données, et de leurs propriétés. Le format PDF peut contenir les formats JFIF et/ou JPEG 2000, générés par les machines ou les applications.

Dans tous les cas, les fichiers à mode bitonal sont enregistrés et compressés sans pertes de données.

4. Formats d'enregistrement des images numérisées

Potentiels de contenu des fichiers avec compression à perte de données					
Propriétés	JPEG-jfif	JPEG 2000 – JP2	jpeg 2000 étendu -JPF	PDF , PDF/A	GIF (pour mémoire)
Modes Couleur	RVB, CMJN	RVB, LAB, CMJN	RVB, LAB, CMJN	Celui du jpeg choisi	RVB
Mode Niveaux de gris	oui	oui	oui	oui	oui
Mode bitonal (mode enregistré sans pertes)	non	oui	oui	oui	oui
Profondeur d'échantillonnage des couleurs par couche	8 bits	1 à 16 bits	1 à 16 bits, capacité jusqu'à 38 bits	1 à 16 bits	4 bits
Type de compression avec perte de données	Jpeg 1991 (IBM) 6 à 12 niveaux selon l'application	Transformée en ondelettes taux continu de compression de 0 à 99%	Celui du jpeg choisi	perdes par le sous-échantillonnage 4 bits, puis codage ZIP ou CCITT dans fichier	
Choix de la compression	utilisateur	utilisateur	utilisateur	Utilisateur via fenêtre de compression image	automatique
Forme d'enregistrement classique	Enregistrement « plat », plat		Plat ou un calque avec transparence	Selon le jpeg choisi	plat
Compatibilité en décompression (lecture)	Totale, avec petites différences imperceptibles selon l'encodage et les marqueurs du fichier			Universel le ISO 32000	Totale (libre)

Tableau 4 Potentiels de contenu des fichiers avec compression à perte de données



Complément : Utilité d'un calque avec transparence ?

Il peut sembler incongru de citer un mode permettant d'enregistrer et de présenter des images avec des transparences, alors que les archivistes travaillent sur des documents bien opaques. En fait, c'est une utilité essentielle pour détourner les cartes postales par exemple, qui seront présentées avec leur contour physique, mais pas le dos du numériseur. L'élimination du fond se fait lors du post-traitement et parfois automatiquement à la prise de vues.



Complément : Le format GIF

Le format **GIF** est réservé à l'usage sur le web. Il sous-échantillonne les couleurs à 256 niveaux, mais peut produire des fichiers plus lourds que le jpeg pour une même image. En revanche, il présente la possibilité de définir des zones transparentes, ce qui est utile pour le détourage d'objets présentés en pages web. Il n'a pas d'intérêt direct pour les activités de numérisation des archives, mais seulement pour le webmestre chargé de leur diffusion Internet. L'enregistrement des numérisations initiales se fera dans les formats tiff ou jpeg, et une exportation sera faite vers le gif. Des «plug-ins» d'optimisation existent aussi pour ce format.



A noter : cette figure contrairement à celle du format jpeg n'a pas été agrandie, car son usage est réservé au navigateur web, et dans ce cas, c'est l'illusion qui compte !

Image 17 Aspects des options de format Gif en 256 couleurs

L'image au format gif avec la palette de couleur «système» accélère le téléchargement au détriment de la justesse de l'image. Avec la palette de couleurs «adaptive» l'image est à peine plus longue à télécharger, mais plus précise. Le Gif a un rendu assez inattendu sur des documents monotones comme les reproductions de manuscrits. L'exemple ci-dessus est flatteur car c'est une prise de vues en haute définition à 300dpi, et les grandes plages d'une même couleur sont inexistantes. Le rendu du papier reste cependant très approximatif.

Le GIF permet de créer de petites animations où différentes images contenues dans le fichier s'affichent successivement selon le rythme défini par leur créateur.

4.6 4.4. Choix des formats d'enregistrement selon l'usage

Nous avons vu ci-avant combien le type de document a numériser influe sur le procédé de numérisation ou colorimétrie .

Nous avons dit que pour choisir le procédé adéquat il faut tenir compte de l'usage que l'on veut faire du fichier .

Ce chapitre fait un inventaire des **règles de base à respecter pour réaliser une photothèque numérique** (caractérisée par des **fichiers individuels**) et **une bibliothèque numérique** (caractérisée par des **fichiers à pages multiples**).

- **Définition de la photothèque numérique** : dans notre usage, on appellera **photothèque toute collection de documents à page unique**, soit
 - les photographies sous toutes leurs formes,
 - les cartes postales,
 - les dessins, tableaux, gravures,
 - les registres ou documents qui sont décomposés en pages uniques servies par une base de données,
 - les plans et dessins techniques.

La numérisation se fera dans un format image mono-page pour chacun des documents et individuellement.

- dans certains cas, les séries de cartes postales (dont l'original est monté en accordéon) seront restitués dans un format multi-pages, mais avec une qualité photographique.
- **Définition de la bibliothèque numérique** : la bibliothèque numérique comprendra les documents multi-pages que l'on téléchargera par sections ou entiers, soit
 - les ouvrages et catalogues,
 - les actes,
 - les registres traités en sections,
 - les thèses, rapports et autres documents multi-pages,
 - certains plans techniques multi-folios,

La numérisation se fera dans un format permettant l'assemblage des pages comme le PDF.

4.6.1 4.4.1. Création d'une photothèque numérique

La photothèque a une double identité : il existe la photothèque de fichiers maîtres, dont la préservation doit suivre les règles de l'archivage permanent, et une photothèque pour la mise en ligne, constituée généralement de copies à basse résolution. Par mesure de prudence, seule cette dernière doit être accessible depuis un réseau connecté à internet.

Dans une photothèque en ligne, il y aura un fichier par image, enregistré dans un format reconnu par tous les systèmes d'exploitation Apple, Windows et systèmes des tablettes et smartphones

L'accent sera mis sur la portabilité

a) Création de la photothèque de sauvegarde ou "source"

Pour cette raison, il est conseillé de constituer une **photothèque de base** dont les éléments sont enregistrés dans un format **sans perte de données**.

Cette photothèque n'est pas obligatoirement en ligne; elle peut être stockée sur un serveur, un disque dur, des DVD; elle peut être mise à disposition seulement **à titre payant** pour les éditeurs qui en tirent un profit pécuniaire. Les fichiers maîtres doivent contenir des indications sur les droits de reproduction pour l'édition imprimée (introduire ces informations dans les métadonnées de l'image.

L'entretien d'une photothèque source, comme tous les stockages d'archives, nécessite une surveillance et a un coût.

- les supports optiques (DVD) doivent être recopiés périodiquement
- le contenu du serveur doit être sauvegardé régulièrement sur un jeu alternatif de bandes, ou sur un serveur distant en réseau sécurisé (cloud☁).

b) Création de la photothèque en ligne

A partir de cette photothèque originale, on va générer des **copies à basse résolution**, auxquelles on appliquera des **compressions importantes**, pour accélérer le téléchargement. Les formats seront le Jpeg, , voire le PDF si l'on veut insérer des légendes visibles ou diffuser des séries assemblées. La **qualité** de ces **copies** doit être suffisamment **médiocre** pour empêcher leur utilisation dans le monde de l'édition imprimée.

irrécupérable à la réouverture du fichier. On peut intégrer à ces images le texte du copyright, en indiquant la provenance, l'auteur, le nom du centre d'archives, la cote de l'original etc. Dans cette catégorie d'images, on générera aussi une série d'icônes ou vignettes à placer dans la page web qui donne accès aux images en basse définition et/ou au bon de commande pour une image haute résolution. Le format JPEG 2000 peut être utilisé pour contenir en un même fichier diverses résolutions.

4.6.2 2.4.4.2. Création d'une bibliothèque de documents à pagination multiple

Une bibliothèque rassemblera une série de documents manuscrits, registres ou imprimés

que l'on décide de diffuser en ligne. Mises à part les photographies ou iconographies faites pour un produit fini (l'imprimé) qui doivent être d'une haute résolution, la numérisation des documents d'archives sériés sert généralement simplement à une consultation à distance.

La numérisation peut se faire dans ce cas dans des formats avec pertes de données puisque l'aspect n'est pas l'objectif premier et que le but est d'accélérer le téléchargement.

Dans un système informatique, les formats courants d'enregistrement d'images permettent d'éditer une seule image par fichier. Comme notre numérisation source sera faite d'un fichier image par page numérisée, il faudrait diffuser une suite de 250 fichiers pour un ouvrage de 250 pages, ce qui est en pratique inutilisable. (cela n'a pas empêché certains prestataires de le faire au temps des réseaux indigents, mais c'est à présent stupide)

Pour rendre la consultation possible et aisée, il est **indispensable de rassembler ces images primaires**.

Il existe deux solutions:

- Créer une base de données capable d'assembler les pages et de les fournir en un bloc à l'utilisateur.
Il existe une multitude de solutions de bases de données, mais elles sont pour la plupart prisonnières de l'environnement du centre d'archives ; certaines nécessitent même l'achat ou le téléchargement d'un progiciel de lecture pour l'exploitation à distance. Certaines de ces bases de données mettent les fichiers natifs dans un "paquet" (exemple le format ZIP) que l'on doit "déballer" à la réception. Au final, le chercheur aura encore une foule de fichiers non rattachés qu'il devra ouvrir un à un.
- Diffuser directement un fichier ou des groupes de pages au format PDF-A
Certains logiciels de base de données sont capables de faire cet assemblage en partant de pages individuelles ou de chapitres au format PDF, de sorte que l'internaute reçoit un fichier monolithique contenant toutes les pages qu'il a demandées. (exemple *Gallica*¹) La diffusion en PDF est donc la plus adéquate pour les documents à pages multiples.

Les trois paragraphes qui suivent décrivent la méthodologie pour constituer un document multi-page à diffuser en ligne:

- *l'acquisition des pages,*
- *l'assemblage de ces pages,*
- *l'habillage du document et sa protection,*
- *la recopie du document en basse définition pour sa diffusion en extranet,*
- *notion sur les robots d'assemblage.*

4.4.2.1. Numérisation des éléments à assembler

Lorsque les pages d'un document sont à assembler en un fichier PDF, il est pratique de les numériser directement dans ce format et d'en régler les propriétés sur la machine si aucun post-traitement n'est nécessaire, cela permet de les assembler éventuellement avec des logiciels.

Pour les documents numérisés en mode bitonal, il ne sera pas fait de réduction de résolution ultérieure, les documents seront numérisés directement dans la résolution utile.

Pour les documents numérisés en couleurs (les manuscrits, les chartes...), dont on envisage une diffusion sur plusieurs années, un fichier maître avec la meilleure qualité envisagée pour le projet sera constitué à la capture, dans le but d'anticiper les possibilités des futurs réseaux. Ceux-ci vont permettre dans moins de 5 ans de diffuser des fichiers de 100 MO comme on le fait d'une page internet de 100kO de nos jours. En attendant, des copies de plus basse qualité seront à mettre en ligne.

Lors de la numérisation en mode « livre ouvert » les machines actuelles permettent de redresser l'image, effacer l'ombre de l'espace de reliure, d'effacer les bords...

4.4.2.2. Assemblage des fichiers au format PDF

L'assemblage constitue la post-production. Alors que la numérisation est machinale, l'assemblage nécessitera des formes de discernement propres à l'archiviste. Si ce travail n'est pas effectué par le centre d'archives, il aura fallu au préalable installer dans le document physique des balises matérialisées par des pages qui contiennent toutes les informations nécessaires à un travail effectué de façon mécanique et aveugle :

- le titre du bloc
- les métadonnées à entrer
- éventuellement la page de garde à assembler...

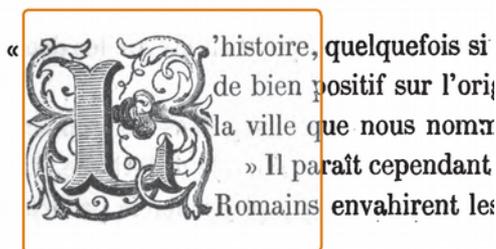
4. Formats d'enregistrement des images numérisées

- le mode colorimétrique de numérisation
- l'emplacement des signets

Le format PDF va permettre d'assembler des pages qui auront été numérisées dans de différents modes, des pages de tailles différentes, des pages de sources différentes : pages numérisées, pages de description au mode texte ajoutées par le centre d'archives...

Par exemple les pages de texte pur peuvent avoir été traitées en mode bitonal et celles imagées en mode couleur ou niveaux de gris. Une ségrégation peut être pratiquée à la numérisation, mais selon le coût de prestation, il peut être moins coûteux de tout numériser en couleurs et de « bitonaliser » le texte par la suite.

Beaucoup de prestataires proposent des numérisations qui sont sensées reconnaître les plages de texte et d'images (tuilage de l'image) mais cela ne donne pas toujours de bon résultats. Par exemple, les gravures sont reconnues comme des images en niveaux de gris lorsque la résolution insuffisante alors que ce sont des modèles parfaits de fichier bitonal. On se gardera donc d'utiliser ce procédé pour les ouvrages anciens, en préférant un fichier bitonal basé sur un test pilote pour déterminer la bonne résolution.



Cette zone a été traduite en niveaux de gris et le reste de la page (texte) en mode bitonal

Page tuillée avec reconnaissance imparfaite des gravures (typique de Google Books)

L'assemblage se fera selon les circonstances de diffusion, en reliant un nombre de pages qui correspond à la capacité de diffusion et/ou au centre d'intérêt de création d'un bloc, par exemple :

une année d'un registre paroissial en bitonal,

Un chapitre d'un livre très imagé,

Tout un ouvrage s'il n'y a que du texte (bitonal),

Toutes les parties intéressant un sujet et extraites de divers livres, magazines, archives, publications électroniques...

L'assemblage est la partie où le rôle de l'archiviste se fait sentir : il est facteur du contenu, et non de l'aspect trivial d'un emplacement ou d'une date de dépôt...

Le point de vue de l'éditeur Adobe l'archivage au format pdf

4.4.2.3 L'indexation ou reconnaissance de caractères

Le format PDF est typiquement celui dédié aux représentations imagées des anciens imprimés, qui ne sont pas des documents texte mais des images.

Qu'est ce qu'une indexation ?

C'est un moyen informatique de reconnaître le texte représenté sur l'image numérisée.

Le procédé s'appelle ROC en Français pour Reconnaissance Optique des Caractères et OCR en anglais (Optical Recognition of Characters)

Lorsque ce texte est reconnu dans la page, dans le document, il est possible d'utiliser un moteur de recherche de texte et trouver directement l'objet de sa recherche.

Les lecteurs de documents PDF ont un moteur de recherche intégré, et les disques de stockage indexés permettent l'identification directe de tout fichier contenant les mots recherchés.

On voit là directement l'intérêt d'indexer les pages de texte numérisées et aussi d'intégrer des métadonnées en clair dans les images.

L'indexation de l'image d'un texte

On peut indexer un fichier numérisé en conservant son aspect original, en couleurs, niveaux de gris ou bitonal, autrement dit en ayant le texte sous-jacent à l'image réelle du texte, exactement à sa place physique. C'est le mode préférable pour l'archiviste, parce que :

- Les caractères mal formés sont toujours reconnaissables à l'œil, alors qu'ils feront chuter le logiciel de reconnaissance optique des caractères
- L'aspect du document original sera préservé, ce qui est la prétention d'un ouvrage d'archiviste.
- Le logiciel ROC reconnaît tout ce qui lui semble un caractère, y compris les taches, traits, petits dessins, mais il n'interprète pas les aplats, les images.

Ce procédé est promis à un grand avenir et est appliqué à des bibliothèques en ligne comme Gallica ou Google books. Mais il a aussi ses limites, pour les raisons suivantes :

- les textes manuscrits ne sont pas reconnaissables, sauf pour les calligraphies et écritures de dessin

4. Formats d'enregistrement des images numérisées

- La limite du ROC apparaît vite lorsque les caractères sont mal formés, peu contrastés, trop petits, d'une police trop ornée...
- Le fonctionnement du ROC nécessite une résolution assez élevée, que l'on peut estimer à 300 dpi pour les documents modernes très contrastés ou avec de grands caractères de 10-12 points, et de 400 à 600 dpi pour les caractères plus petits ou peu contrastés.

Certaines machines intègrent dès la capture un post-traitement d'indexation du texte, sous-jacent à l'image du texte en image numérisée.

Diffusion des fichiers PDF indexés

Lors d'une demande de prestation, l'attention est à porter sur la mise en ligne des documents indexés. Certains prestataires indexent le texte, mais les textes indexés restent dans le site internet et ne migrent pas dans le fichier PDF téléchargé par l'internaute. Il doit donc faire un téléchargement du texte seul (quand c'est possible). Le texte se trouve alors dissocié de son image dans le document original. C'est parfois malcommode... à part pour les plagiaires ! Un bon document PDF doit contenir son texte indexé sous-jacent.

Dans certains lecteurs de fichiers PDF, il est possible d'extraire en une seule fois tout le texte contenu dans le document par une commande d'export ou « d'enregistrer sous ». Il sort avec ses imperfections de reconnaissance, s'il n'a pas été vérifié manuellement (recherche des suspects).

L'enregistrement final

Par simplification, et pour assurer la plus grande pérennité possible, l'archiviste devra utiliser le format PDF-A pour la diffusion hors des espaces de travail du centre et pour le stockage avec l'option « sans perte de données ».



Complément : Liens dynamiques et protection individuelle

Liens dynamiques du document PDF

Le lien dynamique le plus usité dans un document PDF est le signet. Dans une fenêtre juxtaposant le document, l'archiviste peut créer des signets avec le titre de la partie ainsi repérée.

Un éditeur PDF permet aussi d'ajouter des liens dynamiques (hypertexte) vers les pages du document lui-même, à partir de son sommaire.

Des liens peuvent aussi être créés vers un site internet, ou vers un autre document placé sur un serveur local ou distant. Mais dans ce cas, il faut bien être sûr de la stabilité de l'emplacement des sources. Le format PDF-A interdit l'insertion de ces liens qui ne sont pas pérennes par définition.

Protection du document PDF

Le format PDF 1.X permet d'inclure au document des protections partielles ou totales. La copie du document hors de son serveur d'origine ne déverrouille pas ces protections, le document garde donc à tout moment et en tout lieu sa protection originale, tant qu'on utilise Adobe Reader® ou un logiciel approuvé. Malheureusement, certains logiciels de lecture ne respectent pas la protection ou transforment les pages en simples images, sans fonctionnalités ni de protection, ni d'indexage.

La mise en service d'un mot de passe peut verrouiller l'accès complet au document, et un autre mot de passe peut verrouiller individuellement les fonctions de la liste ci dessous :

- l'impression
- la copie d'éléments texte et images
- la modification des pages ou du contenu texte
- l'adjonction d'annotations.

La mise en place du mot de passe déclenche le cryptage en 40 ou 128bits RCA du document, mais il n'est pleinement efficace que s'il conditionne l'ouverture. Cette disposition est intéressante pour **protéger les fonds d'archives à un niveau juste nécessaire**, sans pénaliser les utilisateurs, ni favoriser les plagiaires.

4.4.2.4. Recopie des documents en basse résolution

C'est avec imprimante virtuelle PDF que l'on peut générer en une seule opération un document à basse résolution pour le web en faisant une réimpression PDF du document maître en haute résolution: il suffira de régler les propriétés de l'imprimante PDF pour la basse résolution désirée (livre électronique ou écran) pour obtenir instantanément un document léger et transportable.

Si l'on veut garder **quelques pages** seulement de ce document en **haute résolution**, il suffira de les copier du document source et de les réintégrer dans le document allégé par la commande de remplacement de pages.

Vérifier si la réimpression en PDF ne détruit pas les liens hypertexte que l'on a pu placer dans l'original, si on utilise des logiciels pour la conversion.

4.4.2.5. Assemblage par un robot en ligne

Le découpage et l'assemblage d'un document peut se faire sur demande de l'utilisateur, de la page n à la page n+x, par l'intermédiaire d'un robot d'assemblage (exemple: Bibliothèque Nationale de France - *Gallica*²).

Le développement de cette solution nécessite des compétences informatiques assez importantes. On peut créer un assemblage de fichiers image à travers une banque de données et déclencher l'impression finale à travers une simple imprimante PDF en ligne.

Le meilleur robot sera celui qui assemble directement des pages PDF pour constituer le document particulier. Pour

créer ce robot, il faudra utiliser les codes source PDF qui sont publics.

4.7 4.5. Formats libres et formats propriétaires

Les formats propriétaires proposés par des distributeurs de matériel de numérisation existent encore, mais tendent à se raréfier. La plupart des machines enregistrent de nos jours en Jpeg, tiff et JPEG 2000, et en PDF contenant l'un de ces formats.

Il existe cependant une exception à cette règle, c'est le PIX2 PDF proposé par un distributeur français, dont le but principal est de reproduire les documents imprimés et écrits modernes, en couleurs, et dans une taille très petite correspondant à un fichier bitonal. L'astuce est simple (en apparence). Toutes les parties en fort contraste de l'image sont bitonalisées puis colorisées dans leur ton d'origine, et placées en premier plan alors que le fond de page est entièrement flouté pour constituer une image très légère à l'enregistrement. Le résultat est bluffant, mais il y a plusieurs inconvénients majeurs :

- A chaque image numérisée, il faut payer une redevance à l'éditeur, quand le crédit est fini, il faut racheter des unités...
- Il est impropre au traitement des images, l'objectif est le formulaire imprimé moderne
- Les fichiers ne sont pas toujours imprimables sur les imprimantes à langage PCL, l'aplatissement de l'image ne se fait pas toujours, et seul le fond flou apparaît.
- Il s'agit d'un artefact, plutôt destiné à l'usage commercial qu'aux archives durables.
- Enfin, étant propriétaire, il n'a pas de pérennité certaine et est hors normes internationales de facto.

4.5.1. Formats propriétaires

Les formats «propriétaires» sont ceux développés par une marque et dont les codes source ne sont pas publics. En principe, il n'est pas possible de modifier ou d'adapter soi-même ces formats pour son usage propre.

Bien souvent ces formats ne sont lisibles que par la seule application d'origine. Certains formats propriétaires sont cependant lisibles par des applications connexes ou faisant partie du package de l'éditeur original. Le danger de la disparition de ces éditeurs fait que l'on doit absolument éviter d'utiliser ces formats et n'archiver qu'avec des formats ouverts.

4.8 4.6. Microfilm et numérisation : des champs d'application connexes

Aussi incroyable que cela paraisse, la volatilité des supports de stockage informatique ont donné une nouvelle vie aux procédés de micrographie (microfilmage-microfichage). Les microfilms ont une durée de vie supérieure (et prouvée) aux supports informatiques.

Quand la microforme constitue une archive informatique

Plus on dématérialise les documents, plus le flux d'information virtuelle s'intensifie, plus on ressent le besoin de sécuriser les informations importantes.

La micrographie est à la conquête d'une nouvelle application : **la sauvegarde informatique.**

Le microfilm "numérique" (entendre par là un microfilm argentique qui a été impressionné par un laser numérique) s'impose désormais comme le moyen le plus sûr d'archiver à long terme des informations provenant de systèmes informatiques (les systèmes de banque, de gestion de caisse...). La matérialisation de ces informations les met à l'abri de l'effacement par inadvertance, de l'incompatibilité informatique et de la fragilité des supports d'enregistrement informatiques.

Ce champ d'application couvre les documents juridiques et légaux, les formulaires, documents sociaux, plans et documents au trait, mais ne s'applique pas bien aux documents contenant des photographies, de par la nature même de la microfiche.

L'information sur un microfilm est garantie pour un siècle, les laboratoires de R&D KODAK travaillent actuellement sur un support argentique de 500 ans d'archivage.

Ces sauvegardes sont effectuées à travers des traceurs micrographiques écrivant sur les microformes normalisées. Les nouvelles générations de films argentiques mis au point et identifiés LE 500 par les normes ANSI et ISO, préparent à une durée de vie minimum de 500 ans

On assiste donc à un enregistrement simultané sur deux supports de conservation:

- le support informatique pour la diffusion, l'accès rapide, que l'on pourrait assimiler à une archive tampon,
- le support d'archivage permanent (de sauvegarde) sur des microfiches argentiques.



Attention

Les documents informatiques stockés en microfiches perdent leur nature même de document numérique, car ce sont seulement des images du document qui sont enregistrées sur la microfiche. La re-numérisation de la microfiche ne

optique des caractères!

Numérisation des microfiches

Inversement, les archives permanentes enregistrées sur microfiches, sont maintenant numérisées pour permettre leur exploitation en ligne. Les plans techniques, les registres, les formulaires ayant fait l'objet d'un microfilmage "remontent à la surface" par le biais de ce procédé.

Le point de vue de l'archiviste

Il faut respecter scrupuleusement les normes de microfilmage lors de la sauvegarde de documents informatiques sur les films, en particulier respecter la taille des caractères pour assurer la lisibilité. Alors que certains documents ont été composés avec des caractères ne satisfaisant pas les normes de microfilmage (le caractère *Times* par exemple, installé par défaut par *Microsoft*®), il sera peut-être nécessaire de faire un pré-traitement de mise en conformité-conversion des textes en polices conformes aux prescriptions de microfilmage (*Helvetica* ou *Arial* par exemple), ou changer de format de microfiche pour assurer la lisibilité.

Les numériseurs vont faire l'objet de notre attention dans ce chapitre. Dans les pratiques actuelles, un centre d'archives national ou départemental est équipé de grands moyens, mais les services d'archives locaux ne possèdent que quelques numériseurs pour le traitement de petites séries ou de documents isolés.

Les prestataires de numérisation s'équipent de machines performantes et automatisées. Il existe même des numériseurs de livres qui tournent automatiquement les pages. C'est le cas de google books, inconvénient : les planches d'illustrations ne sont pas ouvertes !

Le numériseur lui-même ne peut pas dans tous les cas être utilisé seul, il faut lui adjoindre un ordinateur pour l'enregistrement et traitement conditionné des images. Ce traitement qui se fait avant enregistrement corrige les courbures de pages, élimine les bords, redresse les vues...

A l'issue de ces opérations, les fichiers sont enregistrés dans un des formats définis ci-avant.

5.1 5.1. Phases des opérations de numérisation

C'est la nature physique des originaux à numériser qui va déterminer les machines à utiliser. Une numérisation de masse fera appel à des prestataires qui disposent de moyens automatisés de production, et d'une unité de fabrication travaillant avec un manuel d'assurance de la qualité, suivant ISO 9000.

Les phases de production s'ordonneront de cette manière :

- Sélection et ségrégation des originaux par formats, par types, et par niveau de fragilité
- Création des fiches d'inventaire des archives ordonnées
- Emballage pour un transport et un stockage optimal
- Prise de mesures assurant la sécurité physique des originaux pendant le transport et le stockage intermédiaire chez le prestataire
- Pratique de la numérisation
- Enregistrement dans le format de fichier maître,
- le contrôle qualité des enregistrements
- récolement des fichiers obtenus en regard des fiches d'inventaire
- Éventuellement le nommage chez le prestataire
- La restitution, le récolement et le reclassement des originaux

A l'issue des opérations de numérisation, le nommage et la post-production sont à effectuer sur les fichiers maîtres.

Pour un tel traitement, il s'agira d'une gestion de projet, et l'archiviste du centre sera plus un préparateur et un donneur d'ordre qu'un praticien.

Cet aspect de projet de numérisation est bien développé dans le chapitre « cahier des charges de numérisation » du site internet de la Direction des Archives de France.

L'annexe 1 de ce cahier des charges fixe les résolutions et formats approuvés par la DAF, et il conviendra de s'y conformer, même lorsqu'on numérise avec les propres moyens du centre d'archives.



Complément

Écrire un cahier des charges de numérisation du patrimoine : guide

<http://www.archivesdefrance.culture.gouv.fr/static/4132>³

Annexe 1 : recommandations sur les caractéristiques des fichiers image

<http://www.archivesdefrance.culture.gouv.fr/static/1309>⁴

5.2 5.2. Le choix du numériseur

La gamme des appareils d'utilisation courante existant sur le marché est adaptée à la numérisation des originaux en feuilles détachées actuels. Ces appareils permettent de numériser jusqu'au format A3, dans tous les modes colorimétriques.

Pour les registres, les livres, les plans de moyen format, les archivistes font appel à des numériseurs spéciaux, qui sont bâtis comme les bancs de microfilmage. La caméra statique ou dynamique selon les appareils, scrute l'original disposé en-dessous, dans un arrangement que l'on nomme « à livre ouvert ». L'original est éclairé par le haut et la numérisation se fait par l'image réfléchie.

Enfin, pour la numérisation des plaques de verre, des négatifs, diapositifs et ektachromes, des fabricants de matériel photo proposent des numériseurs à plat avec un couvercle lumineux et des porte-films. D'autres appareils sont dédiés à la numérisation des films en bandes. Dans ces deux derniers cas, La numérisation se fait par transparence

5.3 5.3. La disparition de certains types de numériseurs

Aussi incroyable que cela paraisse, les numériseurs photographiques à plat et les numériseurs à films sont condamnés à court terme. Ils ont été l'objet d'un engouement passager lorsque le grand public voulait transférer ses images familiales, diapos, négatifs et tirages sur un support numérique. Depuis 2010, le marché se réduit à peau de chagrin parce que l'image est directement acquise en numérique avec les appareils photo et les téléphones multifonctions. Les fabricants abandonnent ces fabrications non rentables, et les appareils encore abordables actuellement, vont devenir dans l'avenir des machines rares et chères réservées exclusivement à l'usage des professionnels du patrimoine et des archivistes.

Les numériseurs de bureautique échapperont à cette disparition, parce qu'ils sont seulement une partie d'un appareil combiné photocopieur – numériseur – imprimante – Téléfax. Mais leurs qualités de numérisation sont insuffisantes au traitement des images de qualité

5.4 5.4. Analyse des caractéristiques des numériseurs

Les principales performances des numériseurs sont exprimées par:

- La résolution optique, caractéristique du nombre de cellules du capteur CCD (et non de sa résolution logicielle qui est le résultat d'une extrapolation). La résolution optique est indiquée par le constructeur. La résolution optique devrait être 3 fois l'épaisseur des traits les plus fins des gravures (estampes) à analyser, soit en pratique au minimum 1200 dpi par 1200 dpi pour les documents opaques.
- La résolution optique des numériseurs à négatifs ou des numériseurs à plat que l'on utiliserait pour des négatifs de moyen format doit être plus élevée, car, dans tous les cas, on fera un agrandissement de l'image de l'ordre de 7 à 10 fois pour l'utilisation. La définition optique minimale devrait être de 2400 dpi x 2400 dpi pour agrandir 8 fois. Les appareils du marché actuels proposent 9600 x 9600 dpi, c'est amplement suffisant, car au-delà, on va juste analyser les grains de l'émulsion photographique.
- La capacité à **distinguer et restituer les couleurs**. On trouve des numériseurs de 24 à 48 bits, soit de 16 millions à des milliards de couleurs possibles pour chaque pixel. Une profondeur de 24 bits, convient à la reproduction des documents d'archives opaques. Il est nécessaire de numériser à 48 bits uniquement pour les diapositives ou les tableaux qui contiennent des couleurs subtiles ou a la limite du modèle RVB.
Mais la véritable distinction des couleurs n'est pas quantifiée dans les notices, c'est le logiciel de conversion qui fait la différence (voir ci-dessous).
- La capacité à restituer les **différences de densité**, par exemple lors de l'analyse des diapositives qui ont de grands contrastes entre les zones les plus claires et les zones les plus foncées. Le constructeur annonce ce rapport par un nombre suivi de D (par exemple 4,2 D pour un numériseur de haut de gamme). Cette caractéristique est essentielle en archives si on veut explorer les détails des plaques photo en verre qui peuvent présenter d'importants contrastes ou inversement une faiblesse de tons nécessitant une analyse très sensible.



Complément : Le choix raisonné

Pour un usage en centre d'archives, ce n'est pas la vitesse de numérisation qui prime, mais plutôt la qualité. Le pilote de l'appareil peut être programmé par un tiers ou par le fabricant de l'appareil. Si l'on exécute simultanément des numérisations sur plusieurs appareils de marques différentes, l'usage d'un pilote commun peut donner des résultats plus constants et simplifier l'approche par l'utilisateur.

5.5 5.5. Inventaire des fonctionnalités

Les fonctionnalités des numériseurs destinés à la photographie sont communes à tous ces appareils, on y retrouve :

- un aperçu rapide et des fenêtres de cadrage que l'on peut mémoriser,
- des contrôles de densité, point blanc/point noir, gamma, correction des couleurs,
- des contrôles de rapport de restitution (réglage programmé du rapport entre la taille d'image et la taille dans le fichier enregistré ; en archives on choisira 100% pour les documents opaques),
- un contrôle de détramage pour la reproduction des imprimés,
- un choix du fichier cible, et une numérotation automatique des clichés de même nom,
- Un choix de profil colorimétrique et de format d'enregistrement, parfois sans compression pour les fichiers sans perte de données ; une post-production sera nécessaire pour effectuer les compressions

Fonctionnement technique d'un numériseur

Une lampe fluorescente ou de diodes éclaire l'original opaque et l'image est captée à travers un système optique par des capteurs CCD pour Charge-Coupled Device, (ou en français « dispositif à transfert de charge » DTC), délivrant un signal en mode analogique. L'image est ensuite discrétisée par une conversion des signaux analogiques en valeurs numériques. Et c'est dans cette opération que se décide la qualité du résultat.

C'est le pilote du numériseur qui réalise cette conversion.

Lorsque le numériseur travaille par transparence, la lumière traverse l'original. Il apparaît à ce moment les mêmes sujétions que pour le tirage photo en agrandisseur : les poussières apparaissent sur le tirage numérique.

Les numériseurs à livre ouvert, ou caméra dynamique, éclairent l'original à distance, l'ambiance du laboratoire de prises de vue a alors une importance, la lumière ambiante doit être diffuse.

Effets de la lumière du numériseur

En règle générale, les éclairages produits par les numériseurs modernes sont peu agressifs pour les originaux, il n'y a pas d'échauffement, et peu d'ultra-violets et surtout l'exposition est d'une durée très courte. Les lampes à filaments utilisées auparavant pour la photo « au statif » étaient bien plus agressives, surtout par la chaleur qui séchait inconsidérément une seule face du document. Cependant, certains conservateurs restent frileux quant à l'exposition d'un original à une lumière électrique. Ils préfèrent le laisser en dépôt ce qui a pour résultat de manquer d'en faire une sauvegarde numérique qui lui serait plus salvatrice que nuisible.



Complément : Fonction avancée de détramage

La plupart des constructeurs proposent une fonction détramage, mais il existe deux moyens pour parvenir à ce résultat :

- La correction logicielle qui applique un flou sur l'ensemble de l'image. Ce moyen est celui des numériseurs de supermarché, et donne des résultats décevants.
- Le détramage par numérisation séquentielle qui a pour principe de numériser plusieurs fois une zone de l'original, en déplaçant la grille virtuelle des pixels de l'image et en en faisant une moyenne. On reconnaît aisément cette fonction à l'usage, car la numérisation dure plus longtemps qu'en mode normal, et quelquefois le numériseur « bourdonne » pendant l'opération (l'attelage optique mobile avance et recule d'un demi-pixel pour effectuer ses deux passes d'échantillonnage).

5.6 5.6. Numériseurs de petit format

Cette famille de numériseurs sera essentiellement compétente pour ce que nous avons appelé la "photothèque" constituée de documents plats et mono-page. Ils sont dédiés à des traitements fins et circonstanciés, généralement à la demande. La qualité plutôt que la rapidité demeure le premier critère. Ces appareils peuvent aussi traiter des originaux transparents (plaques, films).

5. Numériseurs

Numériseurs à plat à vocation photographique

Principales caractéristiques des numériseurs à plat	
Format	jusqu'à A3, A2 dans certains cas
Originaux	opaques ou transparents avec le dos lumineux, exposition sur vitre
Vocation	photographies sur papier, films en châssis, diapositives, plaques de verre.
Vocation supplémentaire	petites estampes, lithographies, cartes postales, lettres, actes...mono-page
Profondeur de champ	1 cm environ pour les appareils à miroir et optique
Temps de numérisation	20-30 secondes à plusieurs minutes selon les tailles d'originaux
Automatisation du passage	non
Automatisation des prises de vues de diapos et films	oui la plupart du temps
Réglage du gamma et des couleurs	Par l'opérateur, pré-réglages enregistrés possibles
Restauration automatique des couleurs	sur option
Résultat	très bon fichier pouvant servir de sauvegarde des originaux en voie de dégradation

Tableau 5 numériseurs à plat

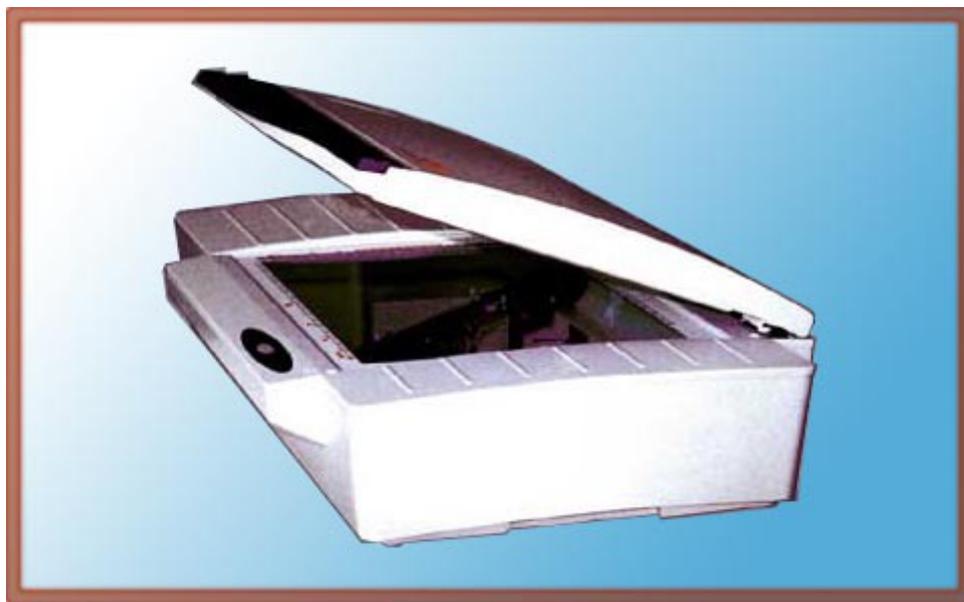


Image 18 Numériseur A3 à plat avec zoom optique interne

Le numériseur à plat à vocation documentaire

C'est par excellence le copieur multi-fonction utilisé couramment dans l'industrie et le commerce, mais cela peut aussi être un numériseur ressemblant à la partie supérieure d'un copieur (voir illustration du numériseur à chargement automatique de document ci-après). Il ne traite que les documents opaques. Il est connectable à un réseau par lequel le document numérisé va rejoindre son lieu de post-traitement, voire de stockage si on le choisit ainsi.

Les appareils de cette famille sont universels et très rapides. Les plus récents offrent une qualité de numérisation telle que l'on peut les utiliser pour traiter soi-même des séries à la demande. Un opérateur peut numériser plusieurs centaines de photographies ou documents dans une journée. Les fichiers sont issus en modes bitonal, bitonal à diffusion de points, niveaux de gris et couleur. Des systèmes experts avec débrayage des images sont installés et donnent les meilleures performances pour la reproduction des imprimés. Différents formats et niveaux de compression sont réglables sur l'appareil ou à distance. On peut pré-enregistrer des programmes de configura

5. Numériseurs

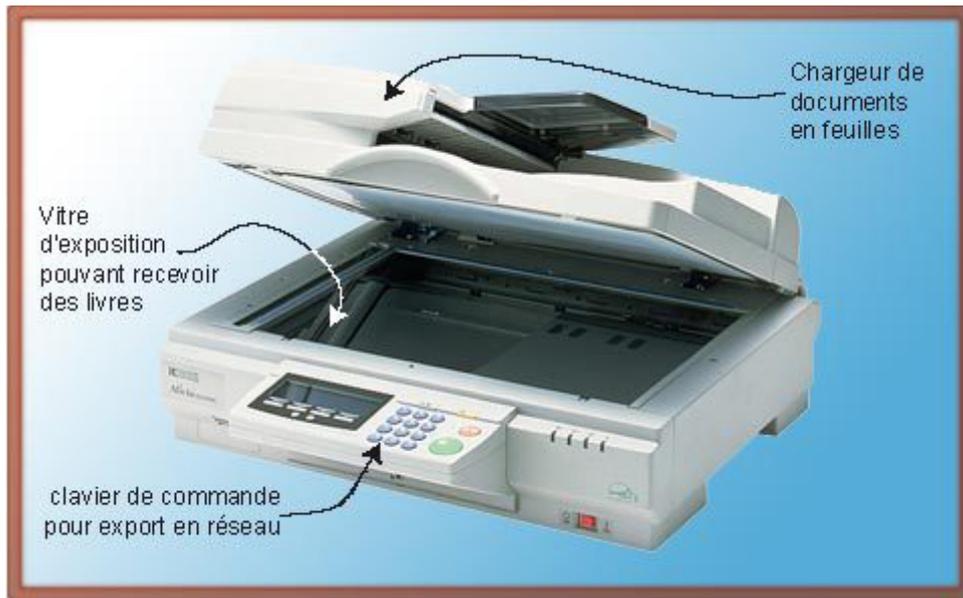
qui permet de confier les tâches à des personnels non techniques.

Caractéristiques de la numérisation avec un copieur multifonctions connecté	
Format	jusqu'à A3
Originaux	Opaques, exposition sur vitre
Vocation de base	Documents manuscrits ou imprimés, affichettes, petites estampes, lithographies, lettres, actes...
Vocation supplémentaire	photographies sur papier, pages d'albums photo, cartes postales...
Profondeur de champ	1 cm environ pour la plupart des appareils, 0,2mm pour les petits appareils à éclairage par diodes
Temps de numérisation	2 à 5 secondes selon les tailles d'originaux et la résolution
Automatisation du passage des originaux	Oui avec le chargeur automatique feuille à feuille
Possibilité de faire des fichiers PDF multi-pages	Oui, c'est la vocation première. PDF sans pertes ou haute qualité possible selon les appareils, Tiff multi-pages
Réglage des couleurs et du gamma	Non, seulement le mode et la densité
Résultat	très bon fichier pouvant servir à la diffusion, et à la copie numérique de sauvegarde simple

Tableau 6 Le numériseur à plat à vocation documentaire



Copieur numérique connecté



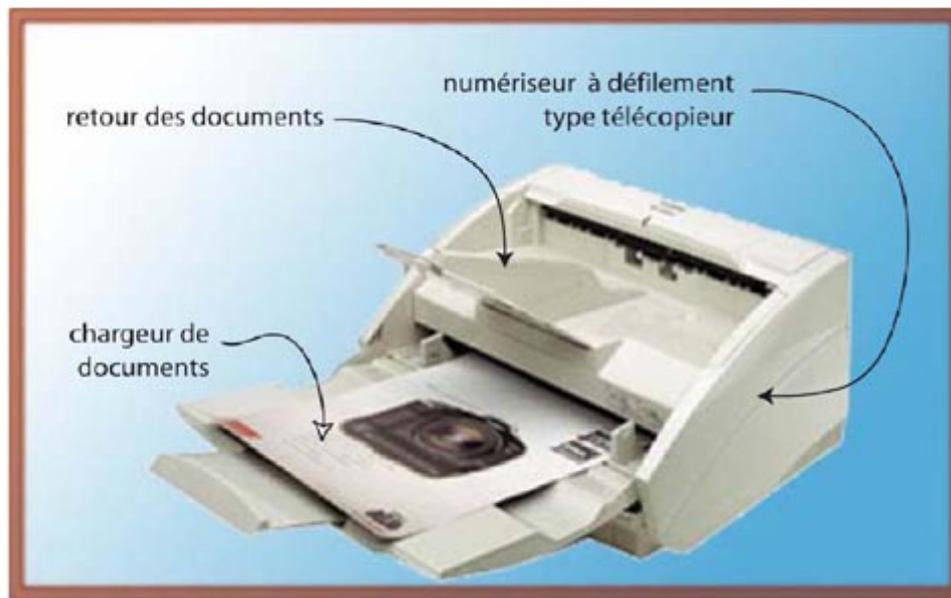
Numériseur doté d'un chargeur automatique de documents. (appareil de bureautique)

Le numériseur à plat à vocation documentaire pour la série

Dans cette catégorie des numériseurs à plat de bureau, il existe actuellement des numériseurs dédiés à la dématérialisation des documents commerciaux et industriels, intégrés dans le flux de gestion des GED (gestion Electronique des Documents). Ces numériseurs ont les caractéristiques informatiques des copieurs connectés, mais ils sont limités à la capture en série de documents de format A4, généralement au format bitonal où ils excellent. Le pilote peut comprendre un système automatique de reconnaissance des caractères (ROC) pour indexer les documents. Leur usage s'étend à la numérisation en couleurs et niveaux de gris pour la capture des documents imagés comme les rapports d'expertise, magazines...

Caractéristiques de numérisation avec un numériseur à vocation documentaire	
Format (taille maximale)	A4 (210x297) et/ou légal américain (215 x355 mm)
Originaux	Opaques, exposition automatique au défilement
Vocation de base	Documents feuille à feuille, couleur ou monochromes
Vocation supplémentaire	aucune
Temps de numérisation	1 seconde par page ou moins
Automatisation du passage des originaux	toujours avec le chargeur automatique feuille à feuille
Possibilité de faire des fichiers PDF multi-pages	Oui, c'est la vocation première.
Réglage des couleurs et du gamma	Non, seulement le mode
Résultat	Fichier pour la documentation ou l'archivage non historique

Tableau 7 Le numériseur à plat à vocation documentaire pour la série



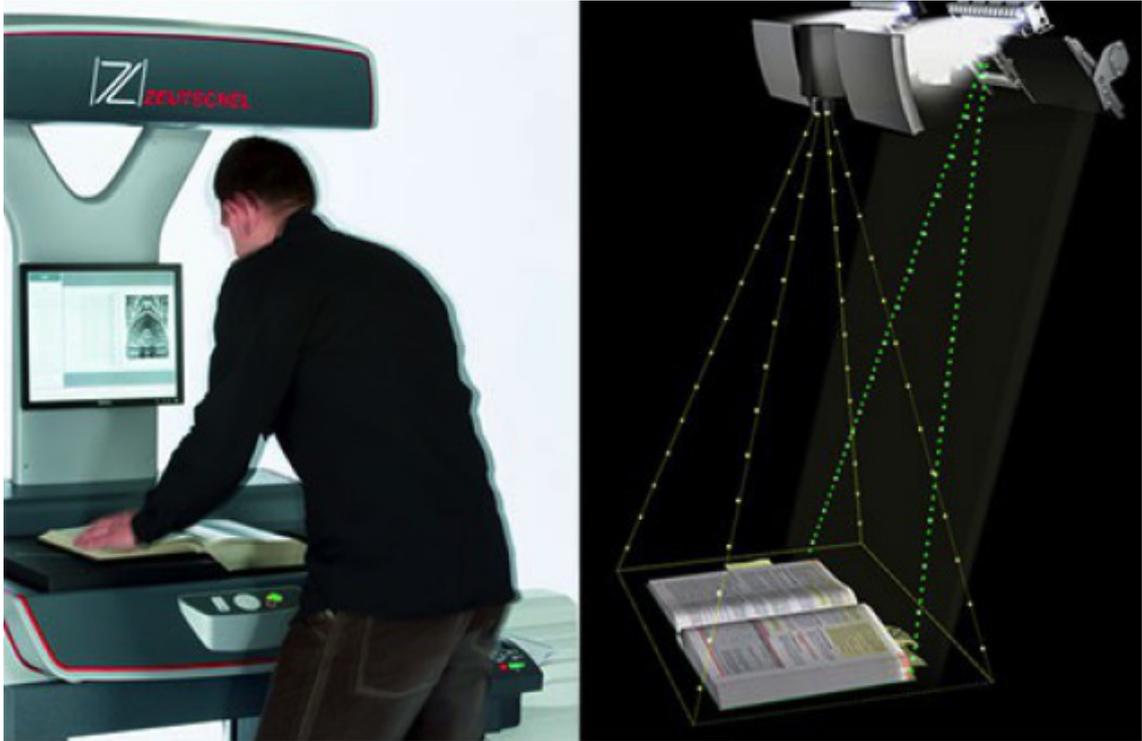
Numériseur multi-pages pour les documents d'affaires.

5.7 5.7. Numériseurs de grand format

Numériseur de livres ouverts

Ces numériseurs appartiennent typiquement au monde des archives et des bibliothèques et sont construits de façon analogue à un banc de microfilmage de registres: le livre ou l'original de grande dimension (A2 généralement, A1 maximum) est posé sur un plateau compensé. La caméra est placée au-dessus et une source de lumière disposée sur la colonne éclaire l'objet à numériser. L'épaisseur de l'objet peut atteindre 200 mm. L'acquisition de l'image se fait par une rangée de cellules disposées de façon linéaire, et une barrette de lumière scrute progressivement l'original. Ce mouvement étant fait à partir d'un axe de rotation, il y a des compensations de mise au point et géométriques.

Le logiciel d'acquisition comporte un effacement des bords, un redressement et une séparation des pages, une linéarisation des parties se trouvant dans la courbure due à la reliure, l'effacement de l'ombre de la reliure, et des algorithmes de bitonalisation agissant par carreaux pour tenir compte des différences d'éclairément. Certains ont une vitre que l'on peut appliquer sur l'original pour corriger sa planéité, mais cela peut être nuisible aux reliures fragiles.



Numériseur à plateaux compensés et à barrette lumineuse

Cette illustration montre l'opérateur disposant un livre sur les plateaux compensés et le principe d'éclairage par une barrette lumineuse.

Il existe des versions allégées de ces appareils pour l'acquisition des livres in-quarto, destinés aux bibliothèques et centres de documentation. Leur prix est plus abordable, et ils permettent de numériser les petits livres sans les presser de façon triviale sur un numériseur à plat !



Le numériseur pour petits livres



Plus jamais cela ! la reliure crie de douleur !

Numériseur à caméra dynamique

Les numériseurs de grand format sont initialement dévolus à la numérisation à livre ouvert. Leur forme de construction est directement issue des machines de microfilmage des registres d'état civil, avec plateaux compensés pour placer les pages dans un plan. Tout comme dans le précédent appareil, la caméra surplombe l'original à numériser, et est dynamique (la caméra se déplace) pour couvrir les très grands formats. Ces appareils conçus pour les grands registres sont aussi très utilisés pour les affiches, les cartes, mais ne se rencontrent guère que dans des centres d'archives nationaux ou chez les prestataires.

Les appareils actuels numérisent en couleurs. Ce sont des chambres de reproduction améliorées par le mouvement de la caméra, permettant de compenser la petite taille du capteur par des prises de vues en carreaux complémentaires. Ces appareils n'ont pas la finesse d'exploration d'un numériseur à plat, du fait même de leur conception qui est plutôt celle d'un appareil photo.

Caractéristiques de numérisation avec un numériseur a caméra dynamique	
Format	jusqu'à A0 environ
Originaux	Opaques, posés sur table ou plateaux compensés
Vocation de base	Registres, livres
Vocation supplémentaire	Cartes, plans, affiches
Profondeur de champ	Plusieurs centimètres
Temps de numérisation	30 secondes à une ou 2 minutes selon le format et la résolution
Automatisation du passage des originaux	Non, sauf cas particulier d'utilisation d'un tourne-page automatique pour la numérisation des livres
Réglage des couleurs et du gamma	Sur certains pilotes
Résultat	très bon fichier pouvant servir à la diffusion, et à la copie numérique de sauvegarde simple

Tableau 8 Numériseur à caméra dynamique



Numériseur à caméra dynamique

Numériseurs à plans

Les numériseurs à plans font généralement partie d'une chaîne de reproduction des plans «au format» et sont complétés par une imprimante de grand format A0. C'est par essence l'outil du bureau d'architecture. Dans le cas des archives, l'imprimante ne sera pas utile, et la numériseur, présenté ici sur un pied support sera plutôt utilisé sur une table, pour ne pas endommager les originaux fragiles. Les appareils actuels fonctionnent dans tous les modes colorimétriques et ont une résolution optique de 1200 dpi, ce qui est utile au format bitonal.

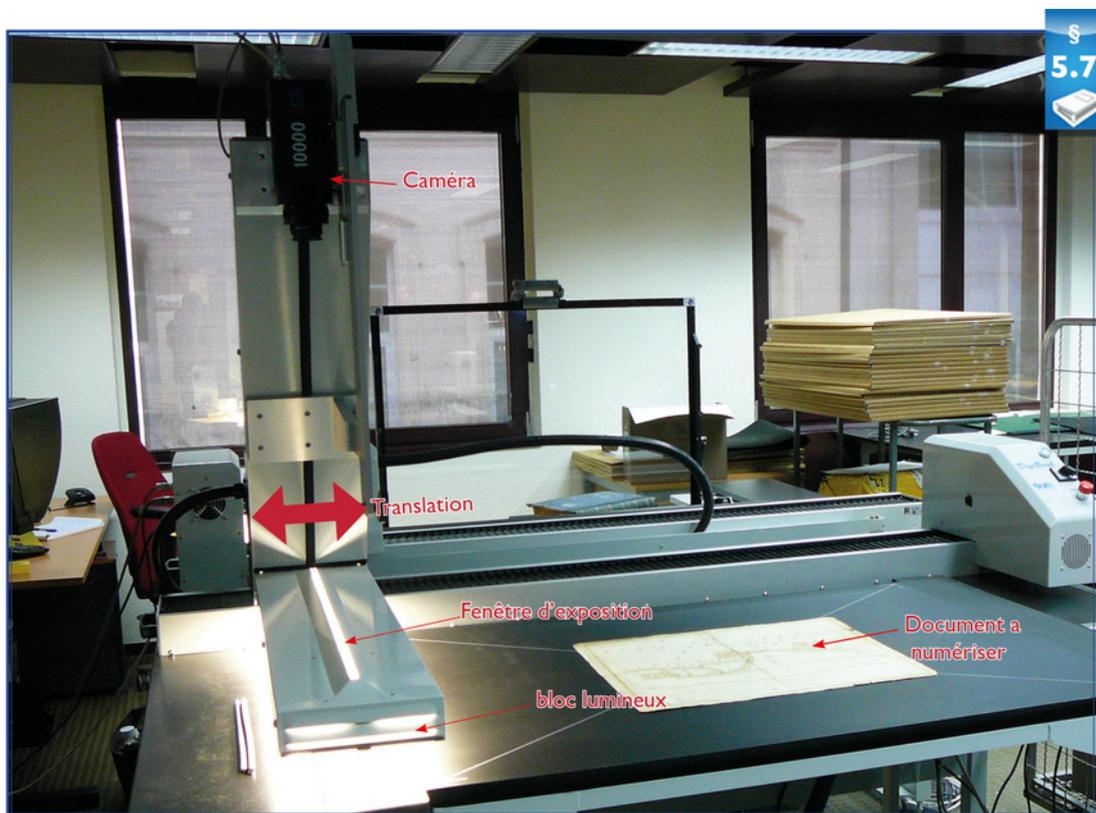
Les numériseurs à plans acceptent seulement des originaux souples qui défilent devant une rangée de capteurs. Il n'est pas possible d'exposer des livres, des originaux rigides ou cartonnés. Les calques et les papiers peuvent passer dans ce transporteur. Les originaux mécaniquement très fragiles sont à placer dans des chemises transparentes en Polyester dit Mylar® . La largeur admissible du plan est généralement de 914mm mais peut aller jusqu'à 1120 mm, alors que la longueur est seulement limitée par la fin de défilement du document.



Numériseur à plans à défilement sur un pied-support

Caractéristiques de numérisation avec un numériseur à défilement pour les plans	
Format	Largeur utile 673, 914 et 1120 mm, longueur illimitée (défilement)
Originaux	Opaques, souples, épaisseur 2mm maximum avec protecteur
Vocation de base	Plans, tirages de plans
Vocation supplémentaire	Cartes, affiches, lithographies, estampes...
Temps de numérisation	Environ 5 cm/s en couleurs et 50 cm/s en monochrome à 200dpi
Principe d'acquisition	Technologie CIS, éclairage successif de l'original par des diodes verts, bleues, rouges
Réglage des couleurs et du gamma	En post traitement
Résultat	Très bon fichier ayant une résolution supérieure au processus photo numérique, ouvrant la porte à la reproduction des estampes au (bon) mode bitonal
Sorties	Bitonal, niveau de gris, couleur, Tiff, Jpeg, PDF, PDF-A

Tableau 9 propriétés des numériseurs à défilement



Numérisation d'un plan sur un numériseur de grand format «suprascan» à caméra dynamique. Le bloc lumineux est placé à proximité de l'original pour un éclairage symétrique sans effet d'ombre. La caméra équipée d'un capteur linéaire scrute le document à travers la fenêtre d'exposition. L'ensemble du bâti caméra et source de lumière se déplace de gauche à droite pour balayer tout le document (Aux Archives de l'Etat en Belgique - Bruxelles photo P.Perrot)

PIAF
PORTAL INTERNATIONAL
ARCHIVISTIQUE
FRANCOPHONE

Numériseurs grand format



Exemple de poste de travail sur un numériseur de grand format. Le registre est placé sur les plateaux compensés. On peut rabattre la vitre d'exposition pour obtenir une meilleure planéité de l'original s'il n'est pas fragile. Sur le moniteur à droite, le document numérisé. Un petit écran au pied de la colonne renseigne l'opérateur sur le nombre de vues et les modes de fonctionnement sélectionnés. (Aux Archives de l'Etat en Belgique - Bruxelles photo P. Perrot)



Numériseurs grand format

5.8 5.8. Numériseurs de films et microfilms

Ces appareils sont conçus pour la numérisation des films photographiques en bandes ou des microfilms. Mais même s'ils sont conçus sur les mêmes principes, un seul appareil ne peut pas remplir les deux fonctions. Alors que les numériseurs photo sont conçus pour la couleur avec des pilotes spéciaux qui éliminent la couche orangée des négatifs, les numériseurs de microfilms sont dédiés à l'image monochrome et sont automatisés pour l'avance vue par vue sur toute la longueur d'une bobine ou les vues d'une microfiche.

Numériseurs de films photo

Le numériseur de film photo comporte une source lumineuse diffuse qui traverse le film. L'image ainsi produite est analysée à travers un système optique, puis discrétisée avec application de filtres, corrections de gamma et de niveaux, qui sont ajustables sur une prévisualisation. Ces appareils acceptent le film 35 mm et les diapositives 5X5. Quelques appareils acceptent le film en rouleaux 6cm -formats 620, 120 et 220- restituant des images 6x9 cm, 6x6 cm 6x4,5 cm). Les planfilms ne peuvent pas être numérisés avec ces appareils, il faut utiliser un numériseur à plat à dos lumineux.



Exemple de numériseur de film photo acceptant le film 35 mm et le film en rouleaux (6cm)

Numériseurs de microformes

Parmi les numériseurs de films **il y a des numériseurs spécifiques pour les microformes.**

5. Numériseurs

pour les transférer sur un support dont la pérennisation est encore problématique (compatibilité des encodages, durée de vie limitée des CD-ROM).

Ce procédé se généralise dans les centres d'archives et les mairies afin de diffuser à distance les registres utilisés par les généalogistes. Il est aussi omniprésent dans l'industrie et l'administration, qui rendent ainsi instantanément accessibles les archives de tout un groupe de travail, quel que soit sa répartition dans l'espace.

En aucun cas, il ne faut détruire les microfiches après numérisation, car leur nature leur permet de traverser le temps.

Les points communs des numériseurs de microformes

Les microformes (microfiches, microfilms 16 et 35mm, cartes à fenêtre) sont numérisées sur des appareils spécialisés. Certaines visionneuses actuelles intègrent un module de numérisation pour l'export du document sur un réseau ou pour l'impression directe.

Certaines machines sont dédiées toutes entières à la numérisation en série, le produit aboutit dans un disque dur intégré ou est exporté directement en réseau.

Ces numériseurs sont délivrés avec un logiciel d'acquisition et de retouche "fonctionnelle" de l'image:

reconnaissance des « leaders »

redressement des vues mal positionnées

élimination des bords noirs

correction automatique des niveaux vue par vue

L'acquisition peut se faire en niveaux de gris en bitonal avec ou sans diffusion de points (dithering); on choisira le mode le plus adéquat pour chacun des originaux. Le mode niveau de gris ne sera appliqué qu'aux documents montrant de grandes zones sombres ou des traits ténus qu'il n'est pas possible de traiter en mode bitonal.

On peut effectuer la numérisation à partir des microformes argentiques ou des copies en diazo (violet) si celles-ci ont été conservées dans de bonnes conditions.

La résolution présélectionnée dans la machine suit les recommandations ISO prescrites pour les microformes.

La résolution indiquée par le logiciel d'acquisition, par exemple 300 dpi pour un document A4, est celle du document restitué à sa taille d'impression, c'est pourquoi il faut bien prendre garde au taux de réduction appliqué lors du microfilmage pour ne pas sous-échantillonner ni sur-échantillonner le fichier.

Numériseur-visionneuse universel

Ces appareils se retrouvent dans beaucoup de centres d'archives et dans l'industrie. Ils peuvent être équipés soit d'une vitre d'exposition pour les microfiches et les cartes à fenêtre, ou d'un dérouleur de microfilm.

Pour les microfiches et cartes à fenêtre, l'opérateur place manuellement l'original et peut régler le rapport d'agrandissement, la position, etc. L'acquisition est ensuite unitaire ou en « batch », mais demeure manuelle et prend un certain temps. C'est un cas d'usage plutôt destiné aux cas particuliers de vues mal exposées, réalisées à partir d'originaux réputés non microfilmables, ou à l'exploitation de quelques vues d'une microfiche.

Si un dérouleur de microfilm est installé, la prise de vues est automatique par avancement du film image par image. L'appareil est alors utilisé pour la production de plans de numérisation.



Visionneuse et numériseur combinés : on voit sur l'écran la projection de la microforme

Numériseur de cartes à fenêtre automatique

Les cartes à fenêtre sont numérisées à l'aide d'un système à transport dynamique. Les cartes disposées dans l'introducteur à gauche sont transférées vers l'unité de lecture, puis ressortent dans le réceptacle à droite. Le logiciel de pilotage permet de prédéfinir la taille de restitution du document, le mode colorimétrique, la résolution à la taille d'impression. Il y a possibilité de traitement par lots. Certains logiciels lisent les cartes à fenêtre perforées et prennent ce code comme nom du fichier. Les formats TIFF, JPEG/jfif et PDF sont proposés par la plupart des fabricants.

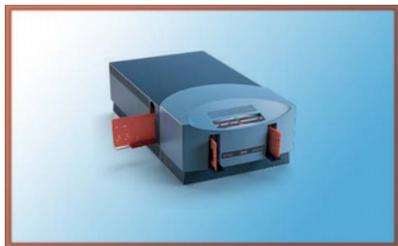


Image 19 Numériseur de cartes à fenêtre

Cet autre appareil travaille lui aussi à partir d'une collection de cartes à fenêtre qui sont numérisées successivement sans intervention de l'opérateur. Ces appareils nécessitent toutefois que chacune des microfiches soit de bonne qualité et avec des traits bien contrastés et des écritures bâton, autrement dit réputés microfilmables.



Numériseur carte à fenêtre

Numériseur de microfiches

Le numériseur de microfiches fonctionne à l'aide d'une caméra dynamique qui scrute les vues une par une et les met à disposition de l'opérateur dans un logiciel de tri et d'adressage. Les microfiches au format A6 les plus courantes comportent 98 vues. Les microfiches peuvent être chargées par séries, l'appareil les met en place automatiquement pour l'acquisition. 200 vues sont numérisées par minute, un opérateur peut alimenter 3 machines. C'est un appareil pour la très grande production. Le nommage reste l'activité la plus « longue » dans la mesure où l'on n'a pas aussi automatisé la reconnaissance des repères en marge de la microfiche.



Numériseur de microfiche

Numériseur de microfilms

Le numériseur de microfilms fonctionne par un transfert automatique des images vers un logiciel de tri et d'adressage. Le film avance devant une caméra de prises de vues qui acquiert les images alors que le film est immobile devant la fenêtre d'exposition. Les films 16mm et 35 mm sont traités par le même appareil, avec une permutation d'objectif et de bobines.



numériseur de microfilm

5.9 5.8. La pratique de la numérisation en centre d'archives

Intéressons nous à présent aux manipulations pratiques de la numérisation en centre d'archives. Ce chapitre ne concerne pas les organisations de grands projets de numérisation, mais bel et bien la numérisation au jour le jour, avec les moyens propres d'un centre d'archives ou d'une bibliothèque.

5.10 5.9. La mise en place des éléments du poste de numérisation

Il est important de constituer un poste de numérisation cohérent, pour éviter la fatigue de l'utilisateur, les difficultés de manipulation et les gestes inutiles.

Le numériseur à plat, s'il est du type «à poser sur le bureau» doit rester le véritable centre du système; les autres éléments seront considérés comme périphériques.

Le numériseur à livre ouvert est à disposer dans un lieu où l'éclairage est modéré et ne varie pas, et n'est pas sujet à des taches de lumière (rayons de soleil, lumières fugitives lors de l'ouverture de la porte...)

L'opérateur devrait disposer d'un écran de 21 pouces pour distinguer les détails. Dans le cas où des travaux de retouche de couleurs sont envisagés, les écrans de bureautique ne conviennent pas, et il est nécessaire d'étalonner les écrans graphiques.

Pour assumer les tâches de numérisation, il est nécessaire de disposer :

- du numériseur et de son logiciel d'acquisition,
- d'un logiciel sérieux de retouche d'image (Adobe Photoshop domine le marché et sa version "Photoshop éléments", peu coûteuse peut suffire à un poste d'acquisition),
- d'une unité centrale conçue pour le traitement d'images et capable de lire, d'exporter et de dialoguer dans tous les mondes: Mac et Unix pour les imprimeurs, PC pour les autres utilisateurs; les unités Mac conviennent mieux à cet usage, le système d'exploitation lit et enregistre tous les médias, et le système d'étalonnage des couleurs est simple à utiliser.

Les autres actions comme la mise en banques de données, la compilation de fichiers PDF, peuvent être faites sur un poste aux performances moyennes, de n'importe quel environnement.

5.11 5.10. Ségrégation des documents

Ségrégation des documents mono-page

Afin de travailler d'une manière efficace et sérieuse, il est conseillé de rassembler autant que possible les originaux en suivant une typologie colorimétrique.



Exemple

On se proposera de numériser en bloc toutes les gravures d'un fonds artistique, puis toutes les sanguines, puis toutes les aquarelles, etc...

La sérialisation permet de rester dans les réglages d'acquisition du numériseur et de gagner du temps de manipulation. Les numériseurs proposent l'acquisition par lots, sans quitter l'application-pilote du numériseur.

Ségrégation des pages d'un document relié

Avant de numériser un document relié rassemblant plusieurs types de pages, il est nécessaire de faire une reconnaissance du contenu pour adapter le mode colorimétrique à chaque page, ou à choisir un mode commun pour

On choisira le mode bitonal pour les pages de textes et gravures et le mode niveaux de gris ou couleur pour les pages contenant des illustrations en tons continus. En outre, dans les imprimés, il faudra activer les fonctions de détramage. Bien souvent, la solution consiste à numériser les documents en couleurs pour préserver les illustrations. Nous verrons plus loin que les fichiers en couleurs JPEG à haute résolution et forte compression donnent les meilleurs résultats pour ces pratiques sur les imprimés/reliés.

5.12 5.11. Post-traitement des images

Travail des images

La numérisation n'est pas forcément finie dès la sortie du numériseur. Il peut être nécessaire de pratiquer un post-traitement de l'image, pour atténuer les défauts de numérisation :

redresser l'image,

enlever les bords noirs, les mires de contrôle éventuellement

retoucher la colorimétrie pour améliorer le rendu tout en restant fidèle au document original que l'on a numérisé,

détourer les cartes postale,

etc.

Le post-traitement inclut aussi la génération des fichiers à basse résolution pour la diffusion en ligne.

Le post-traitement doit être confié à un personnel plus qualifié que celui pratiquant l'acquisition brute. Pour le post-traitement il est recommandé d'acquérir un logiciel exécutant des macrocommandes préenregistrées qui regroupent une série d'actions à déclencher automatiquement à partir d'un clic de souris (comme par exemple Photoshop).

On peut même automatiser le traitement par lots (par exemple le sous-échantillonnage des fichiers de diffusion en ligne), ce qui permet de déclencher ces travaux la nuit, sans occuper inutilement le poste et laisser l'opérateur dans des tâches ineptes et répétitives.

Identification des images

Il est préférable de donner des titres clairs aux fichiers, auquel on ajoute la cote d'archivage plutôt que des numéros d'ordre sans signification. La longueur des titres dans les systèmes actuels comprend au moins 35 caractères.

Ne pas laisser d'espaces pour la mise en ligne sur un serveur internet ou intranet.

Ne pas utiliser de caractères accentués pour permettre l'utilisation dans tous les systèmes d'exploitation de tous les pays sans changer le nom du fichier.

Aperçus (thumbnails)

Il faut toujours activer la fonction «générer l'aperçu» des fichiers image car une petite icône apparaîtra immédiatement dans les fenêtres des gestionnaires d'images, ou bases de données spécialisées sans perte de temps pour la générer lors de la consultation.

5.13 5.12. Informations textuelles dans le fichier image

Les métadonnées sont des textes descriptifs du sujet représenté et des données techniques sur la prise de vues générées automatiquement par les numériseurs et appareils photo. Il est possible d'inclure à l'image, dans les formats conseillés ci-avant, des textes de description de l'image, sa cote en archives, son auteur, sa date d'édition, le copyright, etc... à partir des logiciels de retouche ou de gestion d'image. Ainsi cette image, même détachée du contexte de la notice de son serveur peut encore «raconter son histoire» si on ouvre la fenêtre « informations » dans le logiciel de retouche ou une visionneuse équipée de cette fonctionnalité.

Ces informations sont enregistrées dans des zones qui sont exploitables aussi par des logiciels de base de données spécialisés que l'on trouve dans le commerce dits logiciels de «fetch». Ils sont utilisés par les journaux et agences de presse au niveau mondial, pour le choix et l'échange des photos d'actualité en ligne : IPTC, Héritage. Dans les logiciels professionnelles, compléter un champ commun dans un modèle, complète automatiquement les champs correspondants dans les autres modèles.

Il est aussi possible d'insérer un copyright dans l'image. A ce moment, elle ne pourra être ouverte que si les droits de diffusion sont acquittés ou déclarés. Ce copyright nécessite l'abonnement à un gestionnaire délégué des droits d'auteur et ne s'applique guère au monde des archives.

Importance des métadonnées en ligne

Les métadonnées sont très importantes pour retrouver, reconnaître des images parmi des milliers, dans son propre poste, ou sur le net. Les images contenant des métadonnées sont pistées par les moteurs de recherche comme Google, et elles amènent des chercheurs à votre site, et le font connaître... une image muette avec un tire idiot comme img 4 n'a aucune chance de ressortir dans les moteurs de recherche .

Tout le texte est reconnu par les moteurs de recherche, mais il faut prêter attention au champ « mots-clés » qui oblige l'archiviste à faire une synthèse du contenu du document, et répond particulièrement à certaines recherches par des moteurs spécialisés.

Des logiciels comme Adobe bridge®, des gratuits, permettent de changer globalement les données d'une série d'images ou de documents. Le format PDF a une entrée pour les métadonnées (XMP, qui répond à tous les modèles courants)

Pourquoi des métadonnées ?

Il est essentiel, pour ne pas dire capital d'introduire des métadonnées dans une image, car même détachée du contexte de la notice de son serveur elle doit garder un patrimoine génétique qui peut être récupéré en cas de disparition du support de diffusion (la base de données en ligne).

Les métadonnées sont à placer impérativement dans les fichiers maîtres, afin de toujours revenir au référentiel et non à diverses copies pouvant se trouver dans différents contextes. Il est aussi possible d'insérer un copyright dans l'image (IPTC extension). A ce moment, elle ne pourra être utilisée que si les droits de diffusion sont acquittés ou déclarés. Ce copyright nécessite l'abonnement à un gestionnaire délégué des droits d'auteur et ne s'applique guère au monde des archives, sauf pour les cas où il serait fait une diffusion en nombre vers des éditeurs.

L'importance des métadonnées en ligne

Les sous-fichiers doivent être créés avec des logiciels de conversion qui conservent les métadonnées de l'image. La plupart des petites « moulinettes » adossées à des logiciels de gestion de bibliothèques ne le font pas. Ainsi l'image devient muette, comme quelqu'un qui aurait perdu ses papiers d'identité. Si on la détache de sa notice HTML, elle devient amnésique et orpheline!

Les métadonnées sont très importantes pour retrouver, reconnaître des images parmi des milliers, que ce soit dans le propre serveur du centre d'archives, ou sur le net. Les images contenant des métadonnées sont pistées par les moteurs de recherche comme Google, et elles amènent des chercheurs au site du centre d'archives et le font connaître... une image muette avec un titre anonyme comme img_4 n'a aucune chance de ressortir dans les moteurs de recherche !

Tout le texte des métadonnées est reconnu par les moteurs de recherche, mais pour la reconnaissance en ligne il faut prêter attention au champ « mots-clés » qui est le plus scruté. Ces mots-clés doivent être astucieux, pertinents et même s'étendre à des secteurs connexes pour mieux « harponner » par association d'idées. Par exemple, si l'image représente une antique chaudière d'usine, les mots-clés directs décrivant l'image sont « chaudière, vapeur, nom du lieu, nom du constructeur » des mots secondaires accrocheurs seraient « charbon, appareil à pression, soupape de sécurité, chauffeur, eau traitée »

Des logiciels comme Adobe bridge®, et des gratuits, permettent de changer globalement les données d'une série d'images ou de documents. Les données

«description» du format XMP sont reconnues par tous les lecteurs.

Formats d'enregistrement pouvant contenir des métadonnées

Les formats conseillés pour l'archivage peuvent contenir des métadonnées. Voir à notre chapitre « formats d'enregistrement » pour les détails. En résumé, les formats TIF, PNG, JPG, JPF et PDF accueillent les métadonnées.

Que mettre dans les métadonnées ?

Il y a deux stades de renseignement des métadonnées, l'une étant la retranscription des caractéristiques de l'archive dans son contexte de conservation (cote, dimensions, nature), l'autre étant l'analyse de son contenu, la notice historique, faisant appel au savoir des chercheurs locaux ou collaboratifs en ligne.

Lors de la numérisation, les données suivantes sont connues et à intégrer :

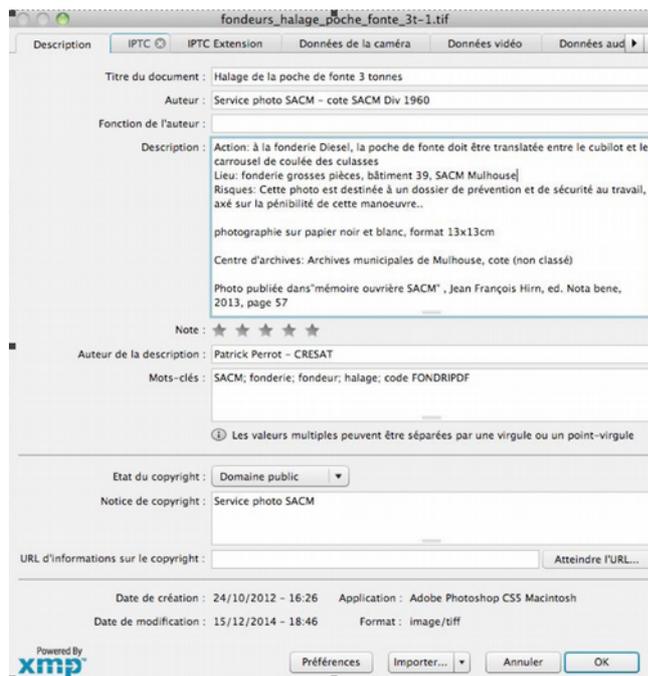
- la cote de l'archive physique et son lieu de conservation
- la nature de l'archive physique (document, photographie,...) et son format (il sera aussi renseigné par le numériseur dans les métadonnées EXIF, tant que les réglages machine auront respecté ce format)
- les textes utiles qui pourraient être inscrits au dos des photographies et documents.

Dans un deuxième stade, les informations plus complètes sur la description sont entrées par les archivistes et chercheurs. Un modèle de métadonnées de base peut

être appliqué à une série d'images, puis les descriptions sont personnalisées et complétées par les personnes compétentes. Ce sont ces informations recopiées que

l'on retrouvera dans les bases de données de diffusion.

- La description de ce que l'on voit sur l'image, lieu, identités des personnages
- Le nom de l'auteur de l'image s'il est connu (l'éditeur pour les cartes postales)
- Le nom de l'auteur de la notice .
- Pour des raisons pratiques, la description, la date, la cote d'archivage, le format et le type d'original peuvent être groupés dans le champ « description ». c'est le champ qui apparaît le plus souvent même dans certains explorateurs de fichiers.



5.14 5.13. Trucs et astuces

Il arrive que l'on ait à utiliser un matériel de numérisation en dehors de sa destination première, surtout pour les cas où il n'est pas économiquement envisageable d'acquérir des machines coûteuses pour des opérations ponctuelles.

5.14.1 5.13.1. Documents plus grands que la vitre d'exposition

Parfois, quelques documents à numériser sont plus grands que la vitre du numériseur à plat. Il est très simple de pallier cet inconvénient pour peu que l'on puisse couvrir le document en deux ou trois prises de vues.

Il suffit de placer le document dans un sous-verre démontable maintenu par des clips métalliques. Ce sous-verre va matérialiser une référence de position et maintenir le document parfaitement à plat. L'épaisseur de la vitre ne gêne pas la plupart des numériseurs (pourvus d'objectifs) dont la profondeur de champ est au moins de 3 à 4 mm. Les numériseurs professionnels sont à mise au point automatique ("auto focus"), la vitre du sous-verre peut troubler cette fonction.

Le dos du sous-verre doit comporter des repères délimitant la surface utile du numériseur, moins les marges de recouvrement, pour permettre ses positionnements successifs.

Il suffit de numériser l'image en déplaçant le sous-verre sur un guide fixé sur un côté du numériseur: ce guide assure le parallélisme des prises de vues successives, ainsi il sera toujours possible de les raccorder exactement dans le logiciel de retouches d'images. Au cours de ces numérisations, il faudra bloquer les fonctions automatiques du numériseur, pour obtenir toutes les images au même contraste et à la même densité.

Lorsque toutes les prises de vues sont exécutées, il faut créer dans le logiciel de retouches d'images un document vide à la taille du document fini, puis y coller une par une les images d'acquisition. Chacune des images importées se trouvera sur un calque distinct, il suffira de déplacer ces calques pour assurer la coïncidence des parties rapportées.

Lorsque les calques sont calés, et les zones de transition plus ou moins retouchées, il suffit d'aplatir l'image (fusionner les calques) puis de l'enregistrer au format désiré.

Il existe dans les logiciels d'image des accessoires de raccordement des images (photo merge) mais cela mène parfois à des déformations des lignes droites, si les originaux ne sont pas bien présentés face à face, nous déconseillons cet usage surtout pour les plans et dessins techniques.

5.14.2 5.13.2. Numérisation d'un ouvrage «à la baguette»

Lorsque l'on ne dispose pas d'un numériseur à «livre ouvert» et que quelques pages sont à numériser, il faudra se résoudre à faire cette acquisition sur un numériseur à plat.

Cette pratique est « meurtrière » pour les reliures, aussi devra-t-on la limiter à des cas extrêmes.

Il est difficile de numériser les livres car la reliure ouverte conduit à un gauchissement des pages qui apparaissent inclinées sur le document final. Par ailleurs, le placement sur la vitre est hasardeux et le décalage d'une page à 56

5. Numériseurs

peut être important.

Pour réduire ce phénomène, il suffit de fixer un objet cylindrique, peint en blanc au-dessus de la vitre du numériseur, à l'endroit où l'on veut placer la reliure (un goujon de menuisier, un tube de cuivre peint...).

Ce dispositif permet:

- de placer le livre dans une position répétitive d'une page à l'autre,
- de bien orienter les pages et de les maintenir parallèles au cadre de numérisation,
- d'épargner la reliure en évitant son écrasement.

La charnière du livre ouvert se centre sur le cylindre et force plus ou moins les pages à s'ouvrir de manière régulière et normale à la reliure, sans la casser. Il faut adapter le diamètre du guide à la profondeur du pli de l'ouvrage (on change le guide au cours du traitement). On obtient de très bons placements surtout si la vitre du numériseur est traitée aux silicones pour être bien glissante (bombe silicones auprès du technicien photocopieur). L'ombre entre le papier et le guide est ensuite éliminée par recadrage, ou tout simplement on règle le cadrage d'acquisition pour cacher l'accessoire.

Cette technique n'est toutefois pas applicable systématiquement aux registres, car les écritures s'étendent souvent jusqu'à la charnière de la reliure, et la baguette cacherait une partie de l'information.

On peut aussi faire cette acquisition sur une baguette de section triangulaire peinte en blanc, ce qui diminue l'ombre, mais guide moins bien la page.

Ce procédé reste un pis-aller à utiliser seulement pour les cas particuliers ou l'exploitation documentaire de quelques pages d'un ouvrage ou d'un magazine.

Si vous connaissez d'autres trucs et astuces, n'hésitez pas à les partager dans le *forum du PIAF*⁵

Les manuscrits, registres d'état civil, lettres et actes sont numérisés pour permettre une consultation en ligne. Les principales vocations de la numérisation sont :

- donner accès à un grand nombre de chercheurs à distance (les généalogistes, les historiens...),
- préserver les registres de trop nombreuses manipulations qui dégradent les reliures,
- préserver les papiers pelures ou papiers anciens fragilisés par le temps.
- Diffuser les microfiches en ligne

Ces vocations sont complémentaires.

Dans ce chapitre, nous avons englobé les véritables manuscrits et les documents dactylographiés ou imprimés, comprenant du texte principalement. Des illustrations au trait intégrées à ces documents peuvent être numérisées comme du texte : graphiques, tableaux, croquis...

Ce chapitre comprend une phase importante pour ce genre de document : la génération de fichiers de très petite taille pour la diffusion. On comprendra aisément que c'est bien le texte lui-même qui intéresse le lecteur, aussi la qualité graphique passe-t-elle au second plan. Cela permet d'utiliser des artifices divers.

Dans le contexte des numérisations de textes imprimés et dactylographiés, il est maintenant de pratique courante de faire une reconnaissance de caractères, et d'intégrer le texte en indexation sous-jacente tout en conservant l'aspect du document à l'écran. C'est un outil de recherche puissant et convivial

6.1 6.1. La nature des manuscrits

Entrent dans la catégorie des manuscrits vis à vis de la numérisation:

- les registres d'état civil, d'enregistrement, les chronos, les cahiers de bord, les comptes-rendus de séances, etc qui sont effectivement manuscrits.
- les actes notariés, les contrats, les testaments même dactylographiés...
- les lettres, les carbonés ou originaux de correspondances dactylographiées...
- et enfin les formulaires répétitifs complétés à la main: dans ce cas, on s'attachera au rendu des écritures manuscrites, la partie imprimée sera de fait rendue avec une qualité satisfaisante.

6.2 6.2. Choix d'un procédé physique pour les reliés



Les registres étant fragiles, il sera nécessaire de procéder à l'aide d'un numériseur à caméra, qui permet l'acquisition en livre ouvert. L'opérateur n'a qu'à tourner les pages pour réaliser les prises de vues successives. La productivité est optimale et l'original est préservé du retournement nécessaire avec un numériseur à plat.

Dans ce cas, les pages seront présentées en face à face, paire à gauche, impaire à droite. Dans le cas où l'on voudrait décomposer le registre en pages individuelles, il est nécessaire de faire un post-traitement informatique pour les séparer, ou utiliser une ressource embarquée de l'appareil, qui est fournie systématiquement dans les appareils actuels.

Le numériseur est généralement fourni avec un logiciel qui redresse l'image et efface les bords par des commandes simples voire automatisées (voir au chapitre numériseurs pour les détails techniques).

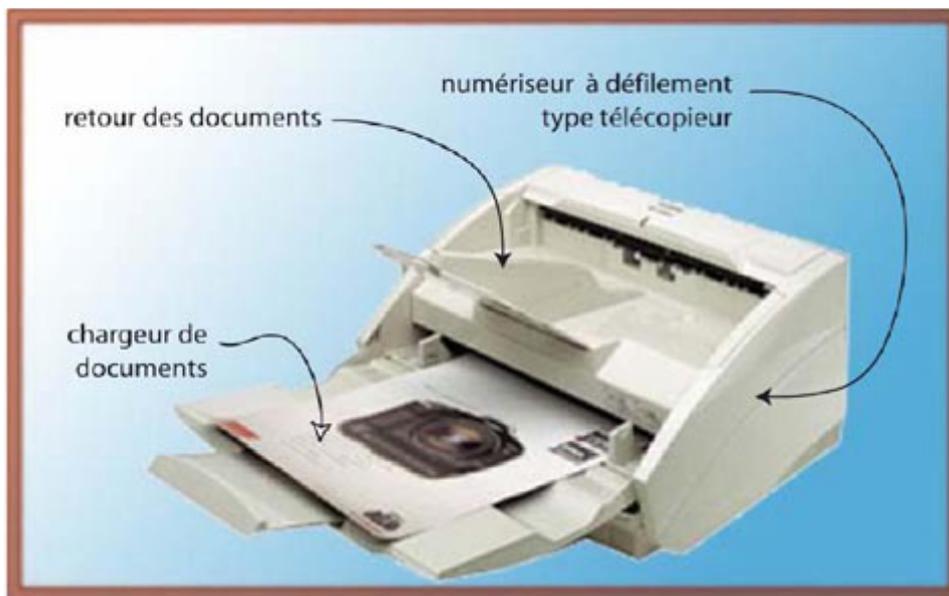
Image 20 Un numériseur de livres à caméra, conçu comme un banc de microfilmage.

6.3 6.3. Choix d'un procédé physique pour les feuilles volantes

Le numériseur à plat est recommandé pour les feuilles "volantes" : il permet de bien orienter l'original sur la vitre d'exposition.

Pour les pages écrites recto verso, il est nécessaire de placer derrière la page à numériser un fond en carton noir, si le couvercle du numériseur n'est pas déjà noir (la plupart des numériseurs professionnels ont un dos d'exposition noir, hormis les dos lumineux). Le noir "absorbe" l'écriture au verso tout en obscurcissant légèrement le fond de l'image à numériser. Il suffira d'augmenter légèrement le contraste de l'image numérisée pour effacer ce fond et bien ressortir l'écriture du recto.

Si l'on veut automatiser le processus pour un grand nombre d'originaux, on peut utiliser un chargeur automatique de documents, mais cette opération n'est possible que si les documents sont d'un grammage suffisant et exempts d'agrafes, de déchirures, plis etc.



6.4 6.4. Choix du mode colorimétrique

Le document manuscrit peut être acquis en trois modes différents:

- Le mode bitonal
- Le mode niveaux de gris
- Le mode couleur



L'exemple ci-contre contient du véritable manuscrit et de l'imprimé.

Image 22 Les modes d'acquisition du document manuscrit

6.4.1 6.4.1. Le mode bitonal

C'est le mode le plus adapté aux registres, tant que les écritures sont bien nettes et contrastées. Il convient aux encres noires et sépia foncé, mais peut se révéler dégradant si des parties du texte sont pâles (les écritures pâles vont disparaître si elles sont en-dessous du seuil de conversion).

Si l'acquisition est faite par défaut sur un numériseur à plat, des ombres dans la zone de reliure peuvent être traduites par de grandes taches noires. Ce problème ne survient pas avec les numériseurs à colonne (la lumière vient d'en haut et élimine les ombres).

Sur la figure ci-dessous, la tache d'humidité au-dessus du «i» de « ramollir » est partiellement traduite en points noirs. Le mode bitonal sera à éviter pour des registres très altérés. La difficulté est de bien régler le seuil de conversion en noir pour assurer un bon rendu au document : pas de lettres mangées, pas d'apparition gênante des écritures du verso ni défauts importants du papier.



Image 23 Le mode bitonal

En mode bitonal, on choisira une résolution suffisante pour permettre la lecture du manuscrit:

- la numérisation en 200 dpi suffit à une écriture assez grosse (équivalente à 10-12 points),
- si les écritures sont fines, il sera nécessaire de numériser en 300 ou 400 dpi pour conserver tous les détails.

Préférentiellement, ces fichiers **seront enregistrés au format PNG**; c'est dans ce mode et avec ce format que l'on obtiendra les fichiers les plus petits pour la mise en ligne et surtout une lecture universelle sur toutes les plateformes. Le format TIF compression LZW est aussi valable, mais le monde Windows a du mal de lire ces fichiers à partir d'une certaine taille, et ne lit pas naturellement le TIF compression Zip.

6.4.2 6.4.2. Le mode niveaux de gris

Le mode niveaux de gris permet de capturer les nuances du document, les densités de l'encre, mais traduit aussi les taches et ombres en niveaux de gris plus ou moins élevés, ce qui peut gêner la lecture.

Pour le fichier maître :

La numérisation à 300dpi restitue même le grain du papier, elle suffit pour cet usage.

Pour les fichiers de diffusion :

- La numérisation en 150 dpi suffit pour une écriture assez grosse
- Pour des écritures plus fines, une numérisation en 200 ou 250 dpi est d'usage courant



Image 24 Le mode niveaux de gris

Les sous-fichiers de diffusion seront enregistré en JPEG pour obtenir une compression élevée des données. Il est à noter que même si le JPEG crée des halos de contraste et du "sable" autour des écritures, elles restent lisibles.



Attention

En pratique, il n'est pas intéressant de numériser en niveaux de gris, car les algorithmes de compression du JPEG/jfif sont peu performants sur ce mode colorimétrique. Aussi surprenant que cela paraisse, la compression d'un document en couleurs au format JPEG donne des fichiers de taille identique, voire plus petits que leur équivalent en niveaux de gris.

Dans la pratique actuelle, le niveau de gris est cantonné à la numérisation des microfiches.

6.4.3 6.4.3. Le mode couleurs RVB

Le mode couleurs permet de capturer toutes les nuances du document, les densités de l'encre, mais traduit les taches et piqûres du papier dans leur couleur d'origine, ce qui facilite la lecture tout en reproduisant l'esthétique du document. Le résultat est flatteur et très lisible.

Le plus grand avantage de ce mode est de pouvoir obtenir des fichiers plus petits que ceux **en niveaux de gris**, avec une compression Jpeg dans les qualités moyenne et supérieure. En qualité basse, les tailles sont égales, parfois plus petites que le niveau de gris.

Pour le fichier maître :

- La numérisation à 300dpi restitue même le grain du papier, elle suffit pour cet usage.

Pour les fichiers de diffusion :

- La numérisation en 150 dpi suffit pour une écriture assez grosse
- Pour des écritures plus fines, une numérisation en 200 ou 250 dpi est d'usage courant



Image 25 Le mode couleurs RVB

Il est à noter que même lorsque le Jpeg crée du "sable" autour des écritures, elles restent nettes, et le fichier à taille égale est plus lisible que celui en noir et blanc.

6.4.4 6.4.4 Reliure en documents PDF

Si les documents sont à relier dans un fichier Pdf, le meilleur conseil est d'incorporer les images au format JPEG 2000 dans le fichier PDF, la taille pour une même qualité est divisée par 2 voir 3 dans certains cas. Le document est lisible sur toutes les plateformes avec le gratuit Adobe Reader®

6.5 6.5. La numérisation de microformes

contraste, pour pouvoir enregistrer les nuances de l'écriture, ainsi que les parties insuffisamment encrées. Il en résulte une image peu contrastée.

Lors de la numérisation, il est souvent nécessaire de pratiquer une retouche des niveaux et contrastes, pour rendre le document plus lisible. Certains numériseurs proposent des filtres qui corrigent automatiquement la densité de l'image pour aller du blanc pur au noir profond (au lieu de se limiter entre gris 20% et gris 60% comme beaucoup de microfiches).

Cependant la solution la plus rationnelle consiste à convertir ou à créer le fichier en mode bitonal. A cet effet, on utilisera avec avantage les commandes de réglage du seuil de conversion à partir de l'histogramme des densités que l'appareil présente dans l'un de ses modes de fonctionnement.

Les numériseurs à microfiches proposent des fonctions expertes pour rendre les fichiers en mode bitonal à "diffusion d'erreur", c'est à dire de nuages de points noirs sur fond blanc, à utiliser dès que les fonds sont colorés inégalement, le contraste insuffisant ou les écritures ténues. Le fichier produit est un peu plus volumineux qu'un fichier bitonal pur et dur, mais le rendu des demi-teintes et traits pâles est remarquable.

Si l'appareil n'est pas capable de faire cette bitonalisation à diffusion de points, il sera nécessaire de pratiquer un post-traitement dans un logiciel de retouche photo, en deux phases:

- accroître le contraste de l'image pour détacher les écritures,
- convertir en mode bitonal en diffusion d'erreur ou en bitonal pur.

Prêter attention à **ne pas générer un fichier en niveaux de gris à 2 couleurs, mais un vrai fichier bitonal**, c'est-à-dire ne pas oublier de changer de mode (de niveaux de gris à bitonal) après avoir effectué les retouches.

6.6 6.6. Importance de la taille des fichiers

Si l'on fait l'inventaire des formats courants que l'on peut utiliser pour enregistrer une page manuscrite, on aboutit à un histogramme tel que celui ci-dessous.

La taille des fichiers est très importante lorsque l'on engage des numérisations en grande série. les histogrammes ci-dessous démontrent de manière percutante que:

- **Notre meilleur choix sera: le TIF bitonal en premier lieu** (dans cette analyse la résolution en mode bitonal a été choisie au double de celle en niveau de gris ou en couleurs).
- Le Jpeg en couleurs sera recommandé pour les autres cas. Le JPEG en niveau de gris sera réservé à la numérisation des microfiches si on ne peut vraiment pas les transformer en mode bitonal.

Analyse comparée des tailles de fichiers :

Un même fichier d'un document manuscrit a été converti dans tous les formats recommandés dans ce cours. La taille de départ 10 000 Kbytes, (environ 10M0) est celle du fichier Tiff en couleurs, que l'on enregistrerait pour une préservation intégrale des caractéristiques du document et sans pertes de données. Cette taille donne aussi une idée de volume d'une **photothèque** de fichiers maîtres (que l'on enregistre en Tiff) par rapport à une **bibliothèque** de documents à consulter **en ligne** (tous les autres formats).

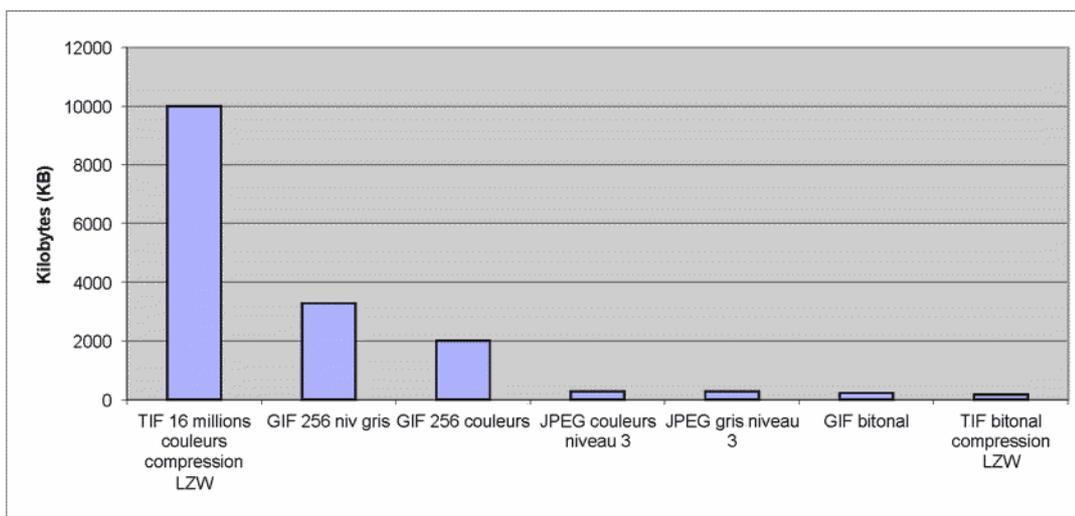


Image 26 tailles comparées de tous les formats

A la lecture de l'histogramme ci-dessus, on découvre une disproportion étonnante entre les modes d'enregistrement. Pour la mise en ligne, on va donc s'intéresser aux quatre derniers.

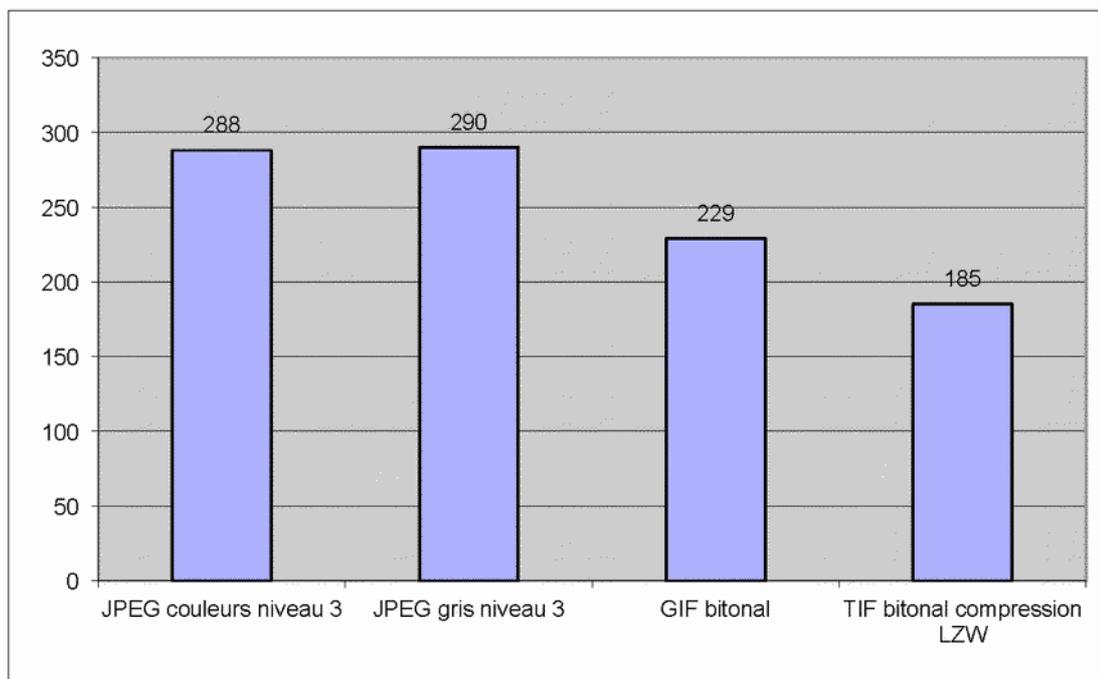


Image 27 tailles comparées des fichiers de diffusion

Ce deuxième graphe montre qu'à un niveau de compression élevé (3 ou basse qualité), le jpeg a la même taille en couleur et en niveaux de gris.

L'enregistrement des fichiers au mode bitonal en TIFF est plus performant qu'en GIF.

Pour donner une illustration comparée de ces modes concurrents, proposons nous de définir les aspects des fichiers à *Isotaille d'enregistrement*, c'est à dire la taille réelle en pixels de fichiers ayant la même *taille d'enregistrement après compression*.

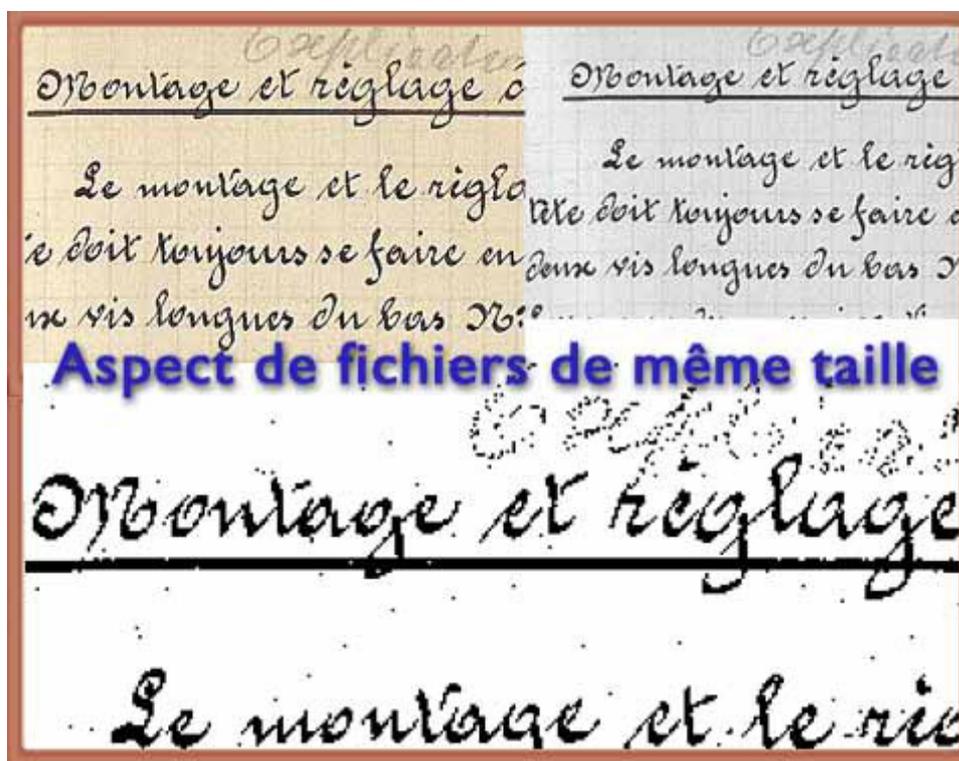


Image 28 Aspect de fichiers de même taille

La simulation ci-dessus tient compte des caractéristiques suivantes:

- Les fichiers couleur et niveaux de gris sont compressés en JPEG aux niveaux de qualité moyenne ou basse (entre 3 et 5 sur 12 niveaux de qualité)
- Dans ces conditions le fichier en niveaux de gris est juste un peu plus grand que celui en couleurs, donc pour

une même taille enregistrée on aura moins de pixels (ce que montre la figure)

- Le fichier bitonal offrira 4 fois plus de pixels pour cette taille, mais se montrera moins performant sur les textes pâles (indication au crayon en haut de l'image)

Pour toutes ces raisons, on peut **choisir des modes différents de colorimétrie pour traiter une même série d'archives** : c'est le contraste des écritures qui nous guidera vers le mode adéquat.

6.7 6.7. La reconnaissance optique des caractères

Le procédé s'appelle en Français Reconnaissance optique des caractères (ROC) et en Anglais Optical Characters Recognition (OCR). Il consiste à reconnaître les textes dans l'image d'une page numérisée.

Les textes manuscrits ne sont pas aptes à la Reconnaissance Optique des Caractères, pour une cause évidente de complexité. Généralement, ce sont des communautés participatives de paléographes qui retranscrivent les manuscrits.

Par contre les textes dactylographiés, en caractères d'imprimerie, ou en écriture bâton très soignée sont aptes à la reconnaissance optique des caractères.

Le chapitre qui suit s'applique autant à ces documents texte qu'aux imprimés comprenant des similigravures et autres images dont nous parlerons au chapitre 9

Comment fonctionne une ROC ?

L'algorithme fonctionne seul, dès qu'il a été mis en route par l'opérateur. Il traite des pages entières. Les phases successives sont :

- Le redressement de l'image de la page pour aligner les lignes sur la grille d'expertise.
- La décomposition de la page en éléments contenant du texte ou des images.
- La bitonalisation du texte pour la reconnaissance.
- L'établissement de la correspondance entre les pixels de chacun des caractères avec une matrice décrivant le caractère type.
- La comparaison des mots formés avec un dictionnaire, permettant d'intégrer certains « suspects » pour reformer un mot.

Ensuite, le texte est soit exporté vers une application indépendante (exemple ABBYY fine reader), soit directement disposé en couche sous-jacente à l'image dans le fichier PDF représentant l'archive (Adobe Acrobat®). C'est cette deuxième solution qui intéressera l'archiviste, qui veut montrer le document dans son aspect d'origine.

L'option vectorisation des caractères

Il existe dans Adobe Acrobat® une option qui permet de transformer le texte pixellisé en texte vectoriel (mode clearscan). Les contours des caractères sont dessinés en langage Postscript® emprunté à Illustrator®. Ils deviennent ainsi très nets rendant la lecture aisée et la réimpression très nette. La taille du fichier diminue drastiquement. Ce principe est applicable aux documents :

- Au mode bitonal (fond blanc)
- Au mode couleur ou niveaux de gris. Dans ce cas le fond situé sous les textes est flouté.

Les caractères vectorisés reprennent la couleur des caractères originaux. Les images, signatures, tampons sont préservés dans l'aspect d'origine et à leur place. Les tonalités du papier ancien sont conservées, mais le détail du grain disparaît.

L'aspect du papier n'étant pas entièrement préservé, l'archiviste peut vouloir ne pas activer cette option pour des documents anciens ayant une certaine esthétique, il conservera alors l'aspect original du document avec le texte indexé sous-jacent, mais pour une taille de fichier plus grande.

Ci-dessous, l'exemple qui a été choisi est dans les plus critiques pour une ROC : c'est un tirage de duplicateur à alcool, de couleur violette, mais avec des caractères bien formés.

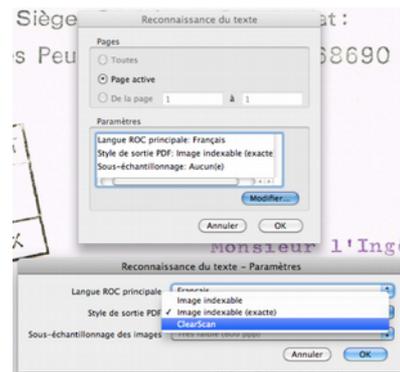
A quoi sert de faire la ROC d'une archive ?

La première utilité est l'indexation du document, c'est à dire que l'on peut faire des recherches d'un mot contenu dans ce document, directement à partir de la fenêtre de recherche d'un lecteur de document PDF. Mais aussi, ce document peut être localisé dans un répertoire ou un serveur à partir d'un mot qu'il contient. Très aisé si par exemple on recherche un nom propre !

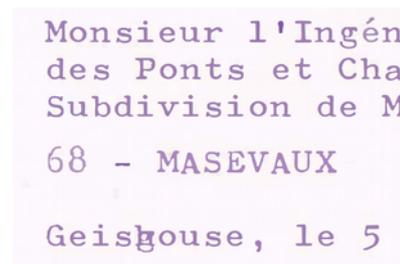
La seconde utilité est la récupération du texte contenu dans le document pour en faire un texte de citation, de travail etc. Ce texte intégré sert aussi à faire la lecture par la voix de synthèse destinée aux malvoyants.

Les logiciels Adobe permettent de sortir le texte par le copier – coller, page par page ou d'exporter tout le document en un seul fichier vers une application de traitement de texte, avec ou sans mise en page.

Exemple de fenêtre de choix de ROC, avec trois options : « Image indexable » pour la page redressée après ROC, « Image indexable exacte » pour une page non redressée, dans ce cas, la sélection de texte sous-jacent sera approximative ; et «clearscan » s'applique pour un texte vectorisé et une page redressée.



Aspect du texte vectorisé. Les caractères ont gardé leur couleur, et même les fautes de frappe sont reproduites comme sur l'original (h de Geishouse) ; le fond est légèrement violet comme l'original.



Comment faire une bonne ROC ?

La reconnaissance optique des caractères fonctionne à partir d'une bitonalisation de la page par le convertisseur. On comprendra aisément que plus la résolution est élevée lors de la numérisation, plus la reconnaissance sera fidèle. Lorsque la résolution atteint 600dpi, pour des caractères de 8-12 points la reconnaissance est au maximum de sa performance. En-dessous de 300dpi, la performance est réduite, surtout si le corps de caractère est en-dessous de 10 points. A ce moment apparaissent les erreurs de reconnaissances, surtout pour les caractères accentués.

Mais alors, pour avoir une ROC fidèle, il va falloir diffuser des fichiers lourds ?

Non car il existe deux solutions :

- après avoir pratiqué la ROC, on peut choisir dans Acrobat de réduire la taille du fichier par une réduction du nombre de pixels ou par une compression Jpeg plus importante.
- Au moment de pratiquer la ROC, choisir l'option de vectorisation des caractères. Ceci est très adapté aux documents modernes en mode bitonal, mais aussi aux documents anciens et se traduit par un flou du papier pour les documents en tons continus.

Les options de correction

On peut faire la chasse aux mots mal reconnus dans l'application même. Le logiciel présente alors en surbrillance tous les mots ne correspondant pas au dictionnaire, et on peut en refaire la saisie dans une fenêtre dédiée. Si l'on veut être très sélectif, et gagner du temps, cette correction peut être limitée aux noms propres et/ou caractéristiques au document, et elle sera rapide si on sait les localiser.

Testez les fonctions de recherche textuelle et de copie du texte dans la galerie « 6.7 reconnaissance de texte (cf.) »,

Lors de la copie du texte depuis l'image en PDF, les endroits où la reconnaissance a eu des difficultés sont marqués par des « ? », on peut aussi trouver des mots décomposés en syllabes suite à des espaces importants de la typographie originale du document. A la fin de chaque ligne de l'original, il y a une fin de paragraphe, il faut donc rattacher les paragraphes pour refaire un texte formatable.

Lors de l'export du texte et s'il est justifié à droite, les retours chariot en fin de ligne sont automatiquement supprimés. Le saut de paragraphe se fait dans ce texte si la ligne du document original n'était pas pleine.

Les centres d'archives conservent de plus en plus de photographies. Ces photographies sont demandées par les chercheurs pour illustrer les bulletins historiques et les ouvrages. La photo est un témoin important du XXe siècle et les publications électroniques donnent des moyens rapides de les divulguer avec simplicité.

Les photographies sur papier, les plaques, les diapositives sont numérisées dans plusieurs buts:

- donner accès à un grand nombre de chercheurs à distance en mettant en ligne des copies à basse résolution,
- préparer les documents à être intégrés dans des publications imprimées,
- préserver les archives photographiques de trop nombreuses manipulations qui les dégradent,

constituer une sauvegarde des documents en photo couleur dont la dégradation est inéluctable et très rapide.

Ces vocations sont elles aussi complémentaires.

Appel aux archivistes

Actuellement, les fonds les plus menacés de détérioration sont les photographies, films et diapositives en couleurs, parce que la destruction des colorants par les produits chimiques restant dans l'émulsion est rapide et irréversible*. Une prise de conscience de cet état de fait est nécessaire, les archivistes devraient contrôler l'état de leurs fonds et déclencher des actions d'urgence sur les images les plus dégradées. Les images dégradées virent au rose, au violet, au brun. La densité de l'image se réduit. Des commandes spéciales de « restauration d'image » existent sur les numériseurs récents, rendant cette opération simple et rapide : l'image ressort avec des couleurs ravivées, ou au moins des couleurs acceptables. Ce sera la dernière chance pour ces images, parce que leur agonie continuera inexorablement et il n'y a rien à faire, même la mise sous vide est inefficace.

On peut faire deux exceptions : les diapositives Kodachrome® et les diapositives Agfachrome 50® professional, développées par des labos professionnels. Ces émulsions ne se dégradent pas rapidement, mais ne leur donne pas encore la durabilité des clichés argentiques en « noir et blanc »

7.1 7.1. Première approche en numérisation de photographies

Il est recommandé de procéder à la numérisation avec une résolution et dans un format d'enregistrement qui préservent les qualités originales de la photographie.

Cette numérisation en très haute qualité restera en principe dans le serveur du centre d'archives, ou sur des supports de sauvegarde physiques comme le CD-ROM, le DVD...

Ce sont des copies en basse résolution (catégorie 2) ou des vignettes qui sont placées en ligne, sur le site internet. Les documents en haute définition (haute résolution et large échantillonnage de couleurs) sont à distribuer à titre onéreux, et le diffuseur doit être identifié pour respecter la législation sur les droits d'auteurs (et reverser les royalties à l'ayant-droit, s'il existe encore).

Ces copies en basse résolution sont obtenues par une réduction de taille de l'image, c'est à dire une réduction du nombre des pixels constituant le fichier original par le truchement d'un logiciel de retouche photo. La résolution de l'image sera donc réduite pour ces copies du fichier maître.

La réduction de la résolution est irréversible, la pratiquer uniquement sur des copies !



Attention

Cette opération doit être faite dans un logiciel qui conserve les métadonnées incluses au fichier maître. Faire des tests avant de pratiquer à la réduction en série.

7.2 7.2. Plaques de verre

Résolution et taux d'agrandissement

Les négatifs sur plaque de verre peuvent être numérisés directement avec un numériseur équipé d'un dos lumineux pour les transparents.

La résolution à choisir est celle permettant la reproduction à 300 dpi pour un tirage d'environ 30x40 cm. Cette

7. Numérisation des photographies

été déterminée par la Direction des Archives de France et fait partie du cahier des charges de la numérisation. Il faudra multiplier 300 par le rapport d'agrandissement pour obtenir la résolution à la numérisation, tout comme on fait pour les diapositives et négatifs de mini-format.

Exemple :

plaque 9x13cm, tirage 40cm divisé par 13 = 3 fois, la résolution minimale sera donc de $3 \times 300 = 900$ dpi. En pratique, il faudra choisir la résolution la plus proche proposée par les réglages prédéterminés du numériseur, ce sont les plus performants en finesse d'image.

Échantillonnage de l'image

Les plaques de verre ne se numérisent pas comme des clichés sur papier, parce que ce sont des négatifs aux nuances très subtiles, dont les variations de densité peuvent être très fines comme très brutales. L'acquisition devra se faire dans un mode qui permettra le « tirage » numérique lors du post-traitement, afin d'obtenir un beau cliché, bien équilibré allant du blanc pur au noir profond. Pour ce faire il faut acquérir la plaque en profondeur 16 bits niveaux de gris, pour permettre au logiciel de retouche de travailler sur un spectre suffisant. Si le numériseur est ancien, il peut ne pas posséder cette option, à ce moment l'astuce est d'acquérir en couleurs (soit 3 couches 8 bits qui totalisent 24 bits) puis lors du post-traitement utiliser ces nuances par une conversion propre au logiciel qui préservera les nuances, voire permettra de les ajuster. (par exemple dans Photoshop, Images > réglages > noir et blanc)

Manipulations

Il faut dépoussiérer délicatement la plaque avant numérisation, mais se garder de toute autre retouche qui pourrait lui être fatale dans des mains inexpertes



Attention

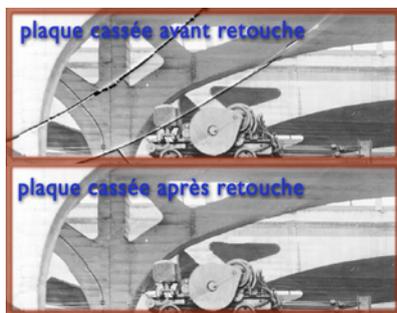
Le meilleur moyen de casser une plaque de verre, c'est de la poser sur la vitre d'un numériseur!

En effet:

- *lorsque la plaque est posée directement sur la vitre du numériseur, l'air entre la plaque et la vitre se raréfie et il peut en résulter un effet ventouse dû à la pression atmosphérique;*
- *lorsque l'on veut relever la plaque, on risque de la casser, car elle est plaquée sur le numériseur par l'action de la pression atmosphérique.*
- *Pour éviter ce dommage, introduire doucement entre la plaque et la vitre du numériseur une bandelette de cartoline qui permettra à l'air de s'introduire, puis soulever précautionneusement la plaque.*

Retouches des défauts techniques

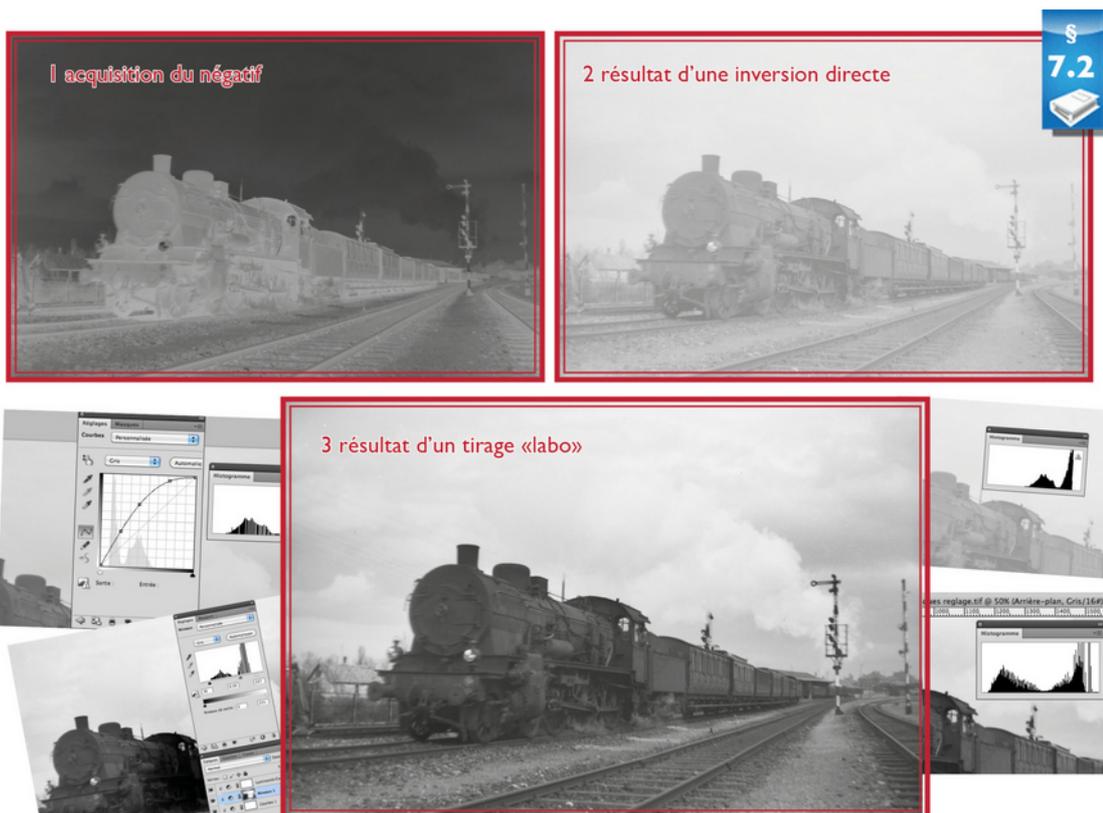
Si l'on décide de restaurer le cliché, les rayures, taches, cassures du verre et autres défauts sont éliminés dans le logiciel de retouche. La figure ci-dessous montre un exemple de retouche numérique. Les métadonnées doivent inclure une mention de cette restauration numérique.



Retouche rapide d'une plaque cassée et altérée à la cassure
Cela a l'air simple, cependant, ce n'est pas à la portée des débutants.

(fonderie SACM à Mulhouse, actuellement FSESJ de l'Université de Haute Alsace)

Image 29 Retouche d'une plaque cassée et altérée à la cassure



Cet exemple montre les trois étapes de la numérisation d'un négatif. Le négatif original est acquis en 16 bits niveaux de gris (1), l'inversion directe(2) montre une image plate. Le post-traitement (3) consiste à refaire le travail du tireur au laboratoire: restituer l'image dans une gamme de tons complets du noir profond au blanc pur. Ensuite l'image est conservée en 8 bits soit 256 niveaux de gris. (Photo Jean Serain)



Tirage des négatifs



Cet exemple montre les trois étapes de la numérisation d'un négatif. Le négatif original est acquis en 16 bits niveaux de gris (1), l'inversion directe(2) montre une image plate. Le post-traitement (3) consiste à refaire le travail du tireur au laboratoire: restituer l'image dans une gamme de tons complets du noir profond au blanc pur. Ensuite l'image est conservée en 8 bits soit 256 niveaux de gris. (Photo Jean Serain)



Tirage des négatifs

7.3 7.3. Diapositives et négatifs de petit format

Les négatifs et positifs (diapositives) de moyen et petit format (6x9 cm, 6x6 cm, 24x36mm...) sont à acquérir avec un numériseur apte à ces tâches. Il est équipé de châssis permettant de disposer les films à distance de la glace d'exposition, en évitant les spires de Newton et l'effet électrostatique qui attire les poussières.

Les négatifs et positifs photographiques sur film nécessitent une numérisation à une résolution supérieure à celle appliquée à la taille d'impression. Pour ces documents, tout comme pour les petites plaques, il conviendra de calculer le rapport d'agrandissement nécessaire et le multiplier par la résolution recherchée à la taille d'impression pour en déduire la résolution nécessaire lors de la numérisation. On appliquera par défaut, comme pour les plaques, le rapport d'agrandissement permettant la reproduction en 300 dpi en 30x40cm, conformément aux instruction de la DAF.



Exemple

Je veux reproduire un paysage industriel dans une revue de patrimoine:

- l'image sur le négatif mesure 24 x 36 mm,
- la photo est susceptible d'occuper une pleine page A4.
- Il devrait exister une réserve pour le recadrage
- Il faut déterminer le rapport d'agrandissement : pour un tirage 30x40cm, il sera de 400mm / 36 mm = 11 fois.

La résolution à l'acquisition théorique est de 3300 dpi, le 3200 dpi est la résolution prééglée du numériseur, ce sera notre choix.

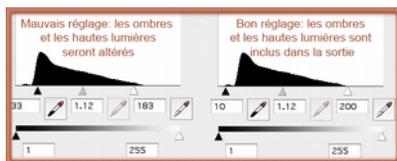
Le 4800 dpi sera choisi si l'on veut couvrir une double page et/ou avoir de la marge pour recadrer un détail de l'image.

Négatifs en noir et blanc

Tout comme les plaques, il faudra numériser avec une profondeur d'analyse de 16 bits niveaux de gris pour conserver les nuances lors du « tirage numérique » de l'épreuve. Bien sûr, les options d'exposition automatique des numériseurs

7. Numérisation des photographies

contrastées, en bref une image flatteuse, mais qui parfois « grille » les hautes lumières et bouche les ombres. Il faudra veiller à ce que le noir 0% n'existe nulle part dans l'image, non plus que le noir 100% dans de grandes zones d'ombre. Le contrôle de l'histogramme permet en un coup d'oeil de corriger ces défauts.



Les numériseurs ont tendance à altérer les extrêmes pour rendre l'image plus flatteuse ; observer la position des curseurs sous les histogrammes, à droite toutes les densités de pixels sont conservées et l'image sera restituée dans toutes ses nuances.

Image 30 Histogramme

Lors de cette acquisition, on s'attachera à capturer toutes les nuances de la plaque ou du négatif. Cependant, l'acquisition de cette manière « monotone » pourra montrer une image d'aspect fané, et il sera nécessaire de pratiquer un post-traitement pour améliorer le rendu et donner à l'image l'aspect d'un « tirage professionnel » tel que les photographes l'exécutaient sur le papier au bromure.

L'enregistrement du fichier maître se fera dans un format sans perte de données.

Diapositives en couleur

Les diapositives en couleurs représentant des sujets documentaires bien contrastés et équilibrés sont à acquérir en RVB avec une profondeur de 24 bits, cela est généralement suffisant. Les Ektachromes® montrant des nuances délicates sont à acquérir avec un échantillonnage de couleurs supérieur. Ce sont par essence les diapositives à nuances subtiles comme les reproductions de tableaux et documents imagés, les paysages de mer, les images scientifiques et spatiales. On fera le choix d'un numériseur ayant une profondeur d'analyse de 42 bits ou 48 bits, soit 14 ou 16 bits par couche au lieu de 8 bits traditionnellement.

Les formats recommandés TIF et PSD conservent ces échantillonnages. Il ne faut pas abuser de cet usage, le 24 bits restitué déjà 16 millions de couleurs.

L'enregistrement du fichier maître se fera dans un format recommandé sans perte de données.

Négatifs en couleur

La numérisation du négatif en couleur se fait à travers le filtre dédié du numériseur qui corrige automatiquement le voile orangé. Ce filtre inverse l'image et affiche un tirage positif de l'image à l'écran. Il y a des filtrages pré-établis pour chacune des marques de pellicules, cependant il faut tenir compte du fait que les négatifs archivés ont été altérés par le temps, il faudra parfois choisir plutôt la correction qui donne les couleurs les plus naturelles, ou créer un profil pour une série d'images.

Dans la photo couleur, il est souvent nécessaire d'effectuer un post-traitement.

Le post-traitement consistera à régler les niveaux et la balance des couleurs pour obtenir une photographie dont les couleurs sont vraisemblables. Là encore, il est nécessaire d'avoir une certaine connaissance technique pour pratiquer cette opération, c'est un travail de photographe. Lors du traitement d'une série de clichés, il est recommandé :

- soit d'enregistrer la courbe de correction appliquée à la première image, et de l'appliquer à toutes les images prises dans les mêmes conditions,
- soit de garder à l'écran la première image retouchée, et appliquer les autres corrections par comparaison visuelle

L'enregistrement du fichier maître se fera dans un format sans perte de données.

7.4 7.4. Photographies sur papier

Les photographies sur papier sont acquises avec un numériseur à plat.

On peut procéder à des numérisations individuelles (une par une) ou placer un lot de photos sur la vitre du numériseur et faire une seule acquisition. Il faudra ensuite copier-coller les zones de l'image globale pour créer des fichiers individuels.

Tout comme pour les documents, on regroupera les photographies par lots de typologie et d'aspect similaire.

On ne mélangera pas des clichés ayant de grandes différences de contraste ou de densité en une même prise de vue: la cellule de mesure du numériseur devrait faire un moyen terme entre des clichés très différents et aucun ne serait traité correctement.

7.4.1 7.4.1. Photographies monochromes

Les photos monochromes sont issues d'un négatif noir et blanc.

Il existe deux catégories de tirages d'après ce négatif :

- les tirages en noir et blanc qui se traduisent en niveaux de gris,
- les tirages noir et blanc qui ont ensuite été virés en couleur: une seule couleur dont l'intensité varie et parfois un mélange de noir dans les ombres et de couleur dans les tons moyens (dont nous avons vu 70 %)

7. Numérisation des photographies

l'équivalent numérique: la bichromie).

Les photos en noir et blanc contemporaines (après 1945) sont généralement à acquérir en niveaux de gris, le virage étant passé de mode.

Certains tirages noir et blanc plus anciens présentent des tons chauds (papiers au bromure) qui font partie du message à transmettre. Les noirs ne sont pas noir pur mais brun très foncé, les gris sont légèrement beiges. Certains papiers présentent des zones irisées bleuâtres là où les sels d'argent sont les plus denses (photos du début du XXe siècle).

La traduction de ces clichés en niveaux de gris va enlever une "partie de la vérité" et pourra être ressentie comme une altération du document. On choisira dans ce cas une numérisation en couleurs; les seules retouches porteront sur la densité du cliché au cas où le numériseur n'aurait pas rendu correctement l'étagement des noirs. En outre, ce procédé présente l'avantage de saisir l'image originale avec 16 millions de nuances, ce qui ouvre un champ intéressant pour le travail ultérieur en niveaux de gris. (voir ci-avant concernant les plaques de verre)

Enfin les photos virées en sépia, virage à l'or, bleu cyanhydrique, seront traitées en couleur pour préserver leur aspect d'origine.



Les photos anciennes sont parfois tirées sur des papiers ayant subi un "virage". De gauche à droite: virage en noir gravure, virage sépia, virage cyan, ou "au prussiate" (Comtesse Sophie Potocka, collection Braun, ADHR).

Image 31 Comtesse Sophie Potocka

L'enregistrement du fichier maître se fera dans un format recommandé sans perte de données.

7.4.2 7.4.2. Photographies en couleur sur papier

Les photos en couleur contemporaines présentent pour certaines des dégradations dues au vieillissement des colorants. Une dominante rouge-violette apparaît sur l'ensemble de l'image, les verts et les bleus s'affaiblissent. Cette dégradation est inexorable et se produit même à l'abri de l'air. Il est donc urgent de numériser ces images pour sauver ce qui peut encore l'être.

Il existe pour ce faire deux chemins :

- la saisie brute avec tous les défauts existants en vue d'un post - traitement
- la « réparation » des couleurs avec un filtre expert du numériseur.

Dans ce second cas, le numériseur utilise des systèmes experts et parfois une exploration de l'image par les infra-rouges, ce qui donne un résultat beaucoup plus rapide et efficace qu'une correction par tâtonnements. L'image peut parfois avoir un aspect criard ou peu réaliste, mais il sera plus facile de repartir de cet aspect pour refaire quelque chose de vraisemblable.

Il faut garder à l'esprit que la photographie en couleurs sur papier est toujours un « mensonge arrangé » qui a suivi les modes et les goûts du public, on trouve même des dominantes différentes selon les contrées : violacé aux Etats-Unis, vert en Angleterre, couleurs chaudes en Allemagne...

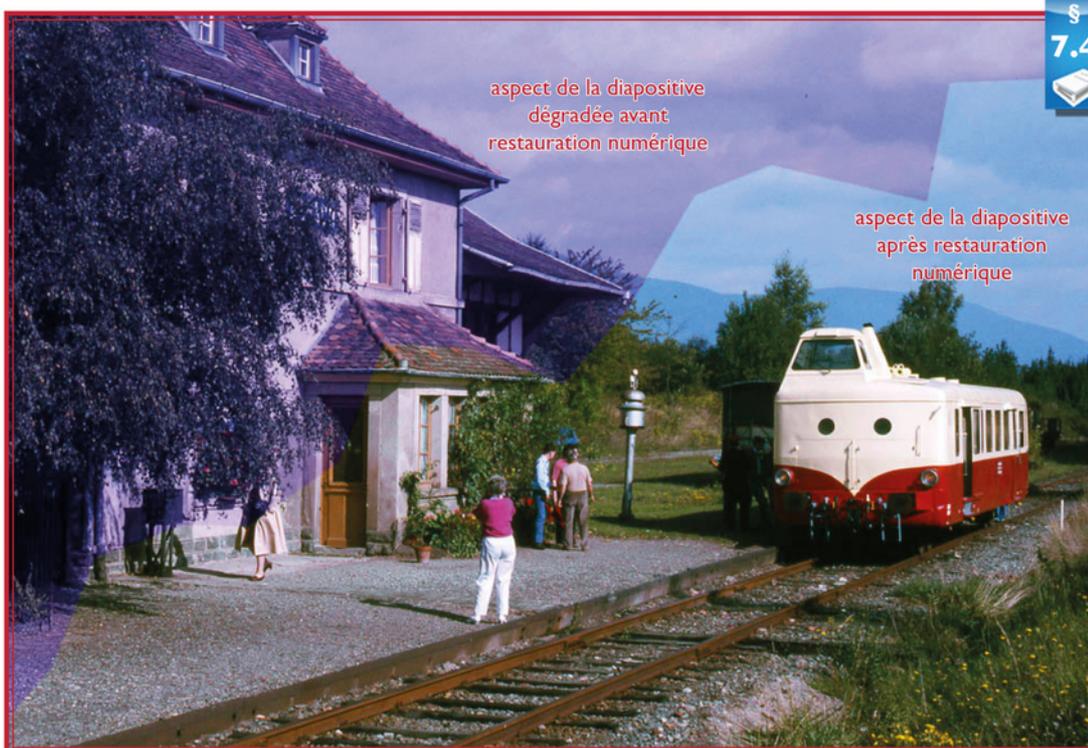


Complément : La restauration des couleurs

Il est à noter que ce travail est à faire dès que possible sur les fonds de photos en couleurs, car le processus de vieillissement est inexorable, et s'accélère après 20 ans de conservation à température ambiante.



La restauration de cette photo en couleurs, même si elle n'a pas restitué totalement les couleurs d'origine, est cependant nécessaire pour retrouver lisibilité et pertinence d'usage. Dans un premier stade de dégradation, l'affaiblissement des colorants formant les bleus et les verts font apparaître une dominante magenta qui s'installe en voile sur l'image, mais l'image est encore récupérable. Lorsque les colorants se seront totalement dégradés, il ne restera que le fantôme magenta de l'image, il sera trop tard pour récupérer quelque ton que ce soit.



Cet exemple illustre en vue composite le résultat de la restauration numérique d'une photographie en couleurs avec une «boîte à outils» de commandes standard: niveaux, balance des couleurs, teinte/saturation, correction sélective. Ce procédé est valable pour les diapos et les tirages photo. Les diapositives et tirages en papier couleur se dégradent inexorablement et plus vite dans des ambiances chaudes. (photo P. Perrot)

PIAF
PORTAL INTERNATIONAL
ARCHIVISTIQUE
FRANCO-PHONIE

Restauration des couleurs

Ce chapitre décrit les règles nécessaires à l'exécution des travaux de numérisation des gravures (eaux-fortes ou estampes et similaires) ainsi que les pièges à éviter.

La numérisation des gravures est de pratique facile dès que l'on a appréhendé le procédé de traduction des traits en mode bitonal.

8.1 8.1. Numérisation au bon mode



Attention

Il est important de bien lire ce chapitre, car le rendu des gravures (estampes) est communément réalisé en tons continus, une approximation récurrente dans les plans de numérisation, alors que ce sont par essence des modèles de dessin bitonal.

La méthode de numérisation décrite ci-dessous est élaborée pour un résultat sans altération de la gravure, notamment pour la reproduction en ouvrages imprimés. Pour la mise en ligne de documents destinés à l'impression domestique (**catégorie 3**), la résolution choisie pourra être arbitraire (voir tableaux des solutions de numérisation) et les gravures seront dégradées partiellement dans leur finesse et dans leur tonalité, du fait de l'épaississement ou de l'effacement partiel des traits.

Les gravures sont des dessins faits de lignes au trait noir sur fond de papier (assimilable à du blanc) qui présentent des zones de haute densité (traits très gros et proches) et des zones de faible densité (traits très fins et espacés).

La numérisation se fera prioritairement en mode **bitonal** (bitmap) et non en niveaux de gris comme on le voit couramment (erreur fatale et commune). Le mode niveaux de gris est satisfaisant pour une lecture seule à l'écran, mais lors de l'impression, les traits de la gravure sont mangés par la trame de l'imprimante et, ce qui est plus grave, la gravure n'est pas numérisée correctement, car les zones franches entre le noir et le blanc ont été transformées en flous gris, ce qui est une **altération de la réalité**. Par ailleurs, le fichier en niveaux de gris a une taille plus importante qu'un fichier en mode bitonal à plus haute résolution, tous deux décrivant un document de la même dimension physique !

Gravure acquise en mode bitonal et gravure en niveaux de gris. Le niveau de gris altère les traits de la gravure dans le fichier enregistré.



Image 32 Gravure acquise en mode bitonal et gravure en niveaux de gris



Complément : Réimpression des gravures

L'acquisition en mode bitonal semble moins fine, car l'affichage à l'écran est synthétisé par groupement de pixels noir et blanc, qui apparaissent découpés en escaliers.

Cependant les pilotes d'imprimantes et flasheuses d'imprimerie lissent les «escaliers» pour donner un aspect de trait continu et net. Reproduites de cette manière, les gravures retrouvent toute leur précision et leurs détails subtils.

Si en revanche on choisit un mode de reproduction en demi-tons (niveaux de gris), l'imprimante ou la flasheuse vont transformer le fichier en une similigravure photographique. A ce moment, ce sont les points de la trame qui vont représenter tant bien que mal les fins traits de la gravure. La figure ci-dessous montre le résultat très agrandi, et on peut juger de la dégradation des traits d'origine.

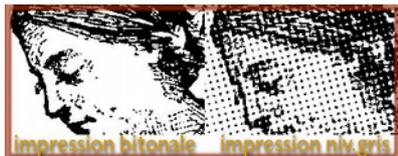


Image 33 Réimpression d'une gravure

Résultats de l'impression à partir d'un fichier bitonal et à partir d'un fichier en niveaux de gris. La gravure traitée en simili est complètement dénaturée

Réglage de la résolution

Pour numériser correctement une gravure et restituer les traits dans leur intégralité, il convient de régler la résolution en dpi par rapport à l'épaisseur du trait le plus fin. On peut déterminer pour une série de gravures de la même provenance une finesse limite du trait, par sondages à l'aide d'un compte-fils gradué en centièmes de mm ou en millièmes de pouces.

La définition sera correcte, si le trait le plus fin est reproduit sur 3 pixels de largeur. Une approche pratique consiste à faire des essais de numérisation sur une zone à traits fins, puis d'agrandir à l'écran pour constater si la règle des 3 pixels est atteinte (le trait ne sera pas interrompu et sa largeur variera entre 2 et 3 pixels aux endroits les plus ténus) . Cette approche sera faite chaque fois que l'on voudra optimiser la taille du fichier pour le stockage.

Lorsque la gravure est destinée à l'impression offset, la méthode est de systématiquement numériser à 2400 dpi, ce qui évite les tâtonnements et les conversions ultérieures.



A gauche, une résolution insuffisante(72dpi): les traits de la gravure sont hachés, et les détails sont occultés.

A droite, une numérisation à 400dpi, résolution suffisante pour rendre les traits de l'original.



Complément : Réglage des densités

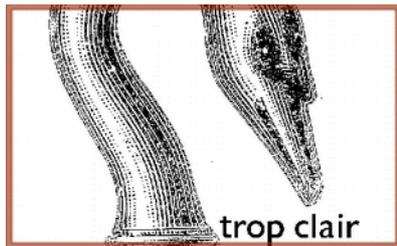
Pour une série de gravures d'un ouvrage, il est nécessaire de faire des essais préalables, car la difficulté est d'obtenir une gravure dont le rendu est complet sans apparition des mouchetis provenant du grain du papier. Le numériseur va générer une image à seuil de densité, c'est à dire qu'il va interpréter les nuances de l'original soit en noir, soit en blanc (0 ou 1).

- Si le seuil de densité moyenne est trop décalé vers le bas, le grain du papier va être interprété en noir et un mouchetis couvrira toute la gravure. Par contre les traits les plus fins seront rendus sans interruption, et même épaissis. Cet épaississement se ressentira aussi dans les zones sombres, et les traits serrés seront transformés en un bloc noir sans détails.



Si le seuil de densité moyenne (conversion) est trop haut, les traits fins vont disparaître et la gravure perdra ses nuances

Image 34 Gravure numérisée avec un seuil de conversion trop bas: le grain du papier apparaît et les détails sont bouchés.



Les traits qui ne sont pas d'un noir absolu sont dégradés et amincis.

Image 35 Gravure numérisée avec un seuil de conversion trop haut

- La numérisation correcte sera obtenue en réglant manuellement le seuil de conversion et en procédant à des essais successifs.

Le seuil de conversion correspond à la densité moyenne des traits de la gravure : les zones ténues sont traduites en noirs, alors que le grain du papier est correctement éliminé.

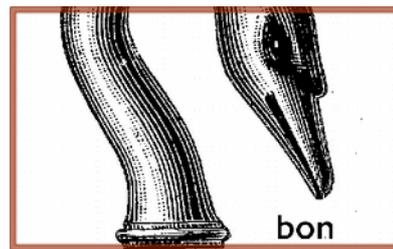


Image 36 Gravure numérisée correctement

La numérisation correcte d'une gravure laisse toutefois apparaître de petits points isolés correspondant à des défauts du papier ou à des petites taches d'encre. Ils sont à enlever à la gomme numérique, si l'on veut incorporer la gravure à un ouvrage.

Possibilités du numérisateur

Il est important lors de l'achat d'un numérisateur de s'assurer que le seuil peut être réglé manuellement à travers le logiciel de capture (ce qui est le cas de tous les numérisateurs professionnels).

A propos de l'acquisition en nuances et le post-traitement de bitonalisation

La transformation en mode bitonal étant irréversible, on n'est pas certain d'obtenir un résultat optimal par un traitement machinal et aveugle en série.

Le choix peut être fait de numériser les originaux en tons continus et de faire ensuite un post-traitement des images en centre d'archives. Ce processus a aussi l'avantage de disposer :

- de fichiers en couleurs donnant une représentation du document original,
- de fichiers en mode bitonal, constitués uniquement de la matière même de la gravure,

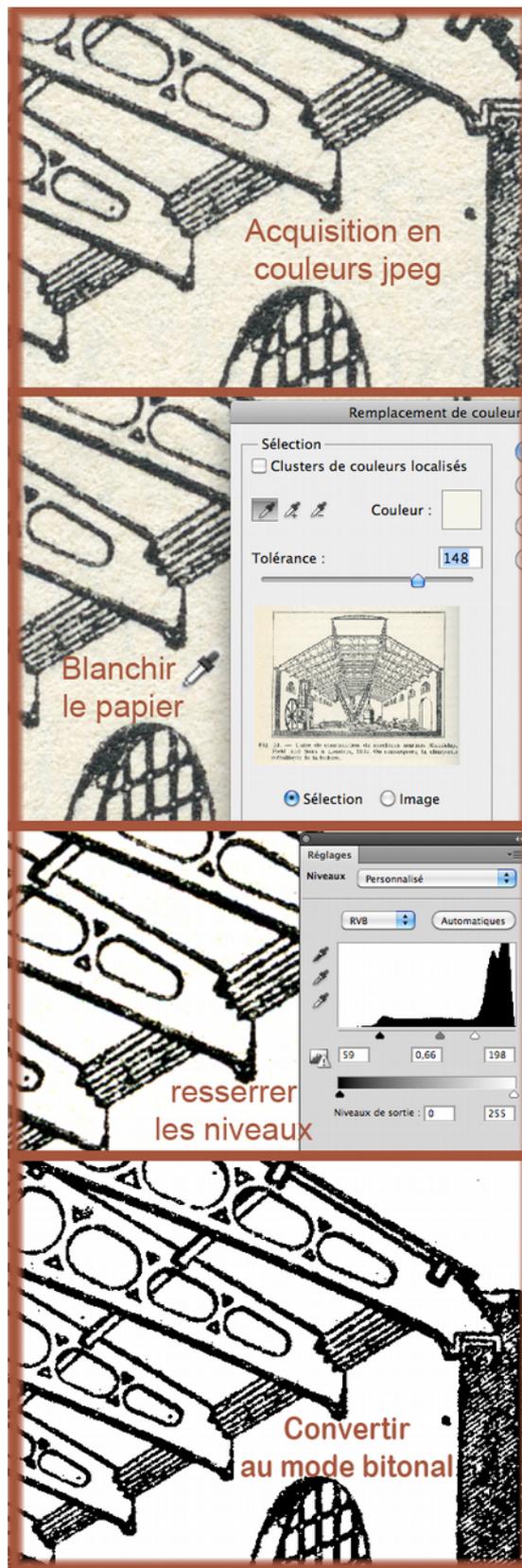
La bitonalisation différée bien menée permet d'éliminer les taches, les rousseurs, les jaunissements du papier.

Le post-traitement sera fait en série dans un logiciel de traitement d'image, dans lequel une macro-instruction mettra en place des outils de transformation pré-réglés. Il suffira à l'opérateur de manipuler quelques curseurs pour obtenir en moins d'une minute une image bitonale optimale.

Les principales étapes sont :

- numérisation en couleurs à la résolution définitive du fichier bitonal (par exemple 1200 dpi,
- enregistrement en JPEG avec une compression moyenne qui n'altère pas le trait. Le fichier est ainsi plus petit pour le stade intermédiaire. La perte de nuances de couleurs n'a pas d'importance puisqu'on va les éliminer,
- Réglage des niveaux avec un calque réglable,
- Elimination du fond par remplacement de couleur,
- Bitonalisation avec un calque réglable,
- enregistrement au format bitonal 1 bit par pixel, sans changer la résolution.

Résultat : un fichier de 11 MO se trouve réduit à moins de 500 kO (indication de proportion). Tout ce qui a été dit ici s'applique aussi à des plans, planches de dessins techniques du type encyclopédie de Diderot et Dalember, etc. Voir des exemples dans la galerie.



La figure montre les étapes principales de la préparation à la bitonalisation. C'est en quelque sorte une "extraction de l'encre" de la gravure imprimée.

A l'origine, l'image est acquise en couleurs 24 bits, mais enregistrée en JPEG de qualité moyenne à supérieure, les artefacts autour des traits ne sont pas gênants dans cet usage.

Le papier est à "effacer" par divers moyens. Sa tonalité jaune est un avantage, on peut sélectionner les pixels de cette couleur et les blanchir.

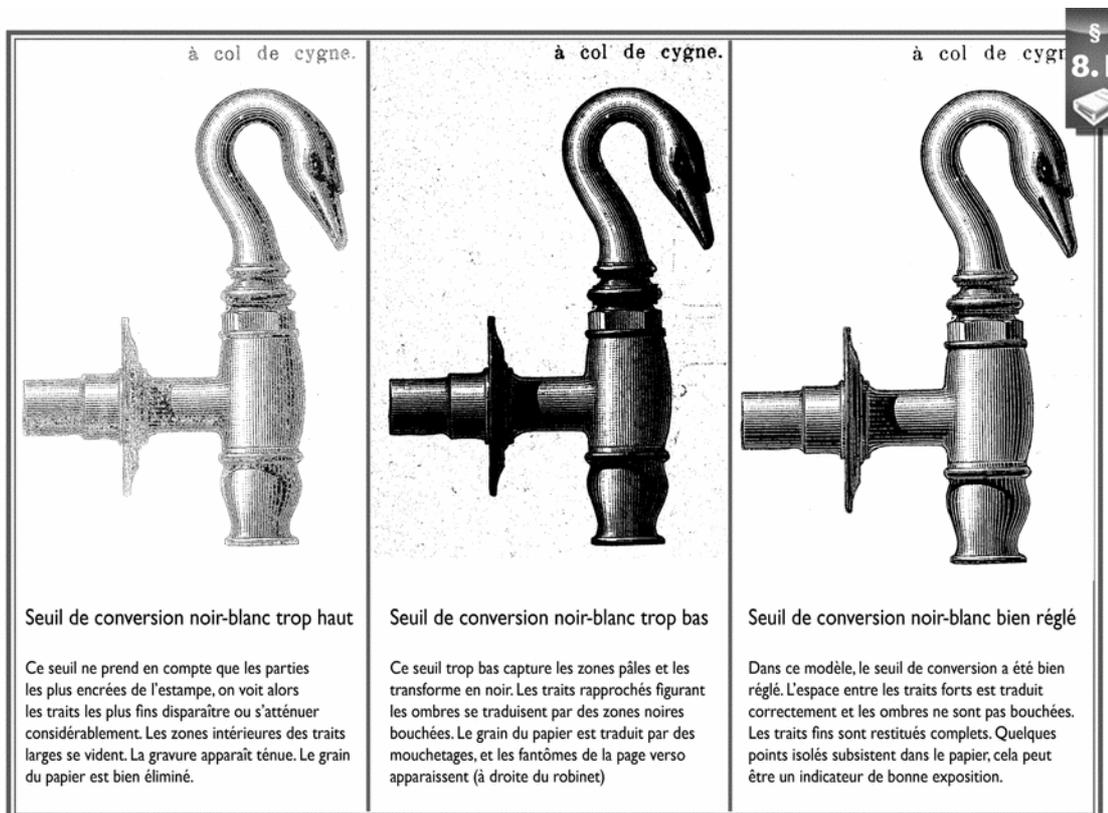
8. Numérisation des gravures

Dans une deuxième étape, on fait de même avec les rousseurs, les éclats de bois brunâtres inclus au papier. Ensuite, le resserrement des valeurs de niveaux (par un calque de réglage) permet de noircir la gravure et transformer le papier en blanc pur.

Il ne reste qu'à régler le seuil de bitonalisation (par un calque de réglage) pour inclure ou éliminer les zones d'encrage ténu.

L'usage des calques permet de revenir en arrière sur les réglages à appliquer, et de les modifier jusqu'à obtenir un résultat satisfaisant.

A la fin des opérations, aplatir l'image et la convertir en mode bitonal. C'est à ce moment que la taille est...divisée par 24, sans compter les avantages de l'algorithme de compression!

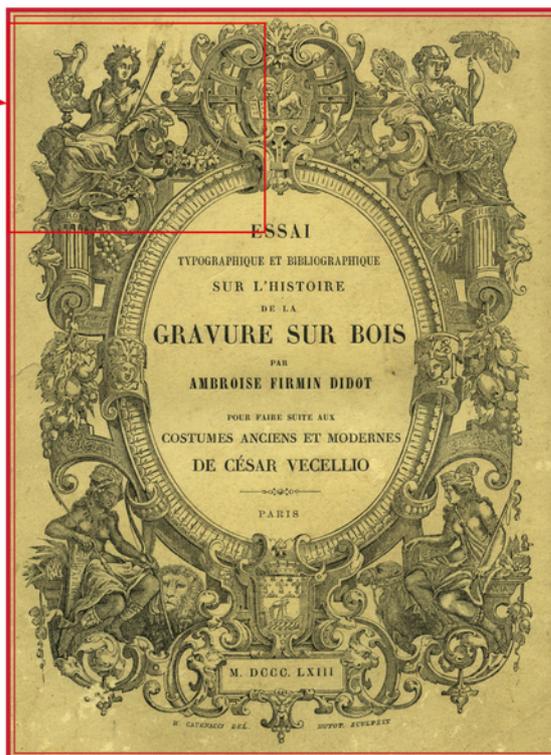


Cette image de la galerie montre en détail les numérisations du robinet à col de cygne du complément du chapitre 8.1.

PIAF
PORTAL INTERNATIONAL
ARCHIVISTIQUE
FRANCOPHONE

Détermination du seuil bitonal

voir ce détail sur
le fichier bitonal



Cette numérisation montre un ornement de livre de Firmin Didot. Le fichier maître est en 2400 dpi, pour une taille de 108 MO compressé. En couleurs, le document conserve l'aspect du document d'origine mais à cette résolution il est trop grand pour la diffusion. Si on en réduit la résolution, la finesse de la gravure en sera altérée. La solution consiste à extraire la gravure au mode bitonal pour en conserver tous les détails. C'est en quelque sorte une « extraction de l'encre » (fichier suivant dans la galerie)



Gravure estampes en bitonal



Voici un fragment de la gravure en mode bitonal en 2400 dpi **agrandi 4 fois** sur cet exemple. Le fichier fini pèse 6,8 MO, soit une taille 15 fois plus petite que celui en couleurs, avec un contenu de grande qualité. Pour la diffusion, un fichier à 600 dpi montre suffisamment de détails. L'impression du fichier bitonal est excellente, testez avec cet exemple. Cette image de la galerie est entièrement au mode bitonal.



Gravures estampes en bitonal

8.2 8.2. Cas spéciaux

Il est parfois nécessaire de numériser une gravure en niveaux de gris :

- lorsqu'elle doit être réduite en taille pour une impression domestique,
- si le rapport de taille d'impression (en ouvrage) est inférieur à 50%.
- lorsqu'elle est à diffuser sur la toile ou pour consultation à l'écran uniquement.

On peut :

- soit acquérir directement la gravure en niveaux de gris,
- soit convertir un fichier bitonal en niveaux de gris, en réduisant la résolution pour l'adapter à l'usage final

Ce sont les seuls cas pratiques de transformation du fichier bitonal en niveaux de gris. Il est à noter que cette transformation, avec un rééchantillonnage du nombre de pixels formant l'image, peut se faire dans un logiciel de retouche photo, muni de convertisseurs performants.

La figure ci-dessous montre le rendu des précédentes gravures après transformation en niveaux de gris et réduction de taille en pixels.

On s'aperçoit en outre que si la première acquisition en mode bitonal n'est pas correcte, la conversion en niveaux de gris ne rattrapera pas les défauts de densité.

Gravures du cygne converties en niveaux de gris pour diffusion à l'écran : les défauts inhérents à l'acquisition ne pourront pas être corrigés par ce biais.



Image 37 Gravures du cygne

Les imprimés traités dans ce chapitre sont les publications contenant des textes, des photographies en similigravure, des dessins, issus de divers procédés d'impression dits photomécanique.

Ne sont pas concernés par ce chapitre, les imprimés administratifs ou formulaires, qui ne comprennent que du texte et des filets. Ceux-ci sont à traiter comme les manuscrits.

Les ouvrages qui sont faits d'illustrations au trait et de texte, sans photographies en similigravure, sont à numériser au mode bitonal.

Comme le traitement des imprimés est assez particulier, il est nécessaire de comprendre dans un premier temps leur procédé d'élaboration, pour bien les appréhender lors de la numérisation.

9.1 9.1. Contenu des imprimés, procédés d'obtention

Les imprimés contiennent du texte et des images en «similigravure», reproduites à partir des photographies par des procédés dits «photomécaniques». Leur nature a été transformée pour s'adapter à la création d'une forme imprimante, c'est-à-dire que les valeurs tonales variant en continu ont été transformées en points élémentaires de nature bitonale, à travers une trame : c'est la similigravure qui est à la base de la plupart des reproductions imprimées modernes.

Nous verrons que cette similigravure va imposer des choix lors de la numérisation. Il n'est pas toujours possible d'acquérir ces originaux en mode bitonal, aussi faudra-t-il «détramer» certaines images photographiques imprimées en similigravure.

Avant de classer les imprimés, il est indispensable de comprendre le principe de la similigravure pour bien évaluer les difficultés qu'elle apportera dans la numérisation

9.2 9.2. Principe de la similigravure

La similigravure est nécessaire à la confection d'une **forme imprimante** appelée communément le **cliché**. C'est cette forme imprimante qui va reporter l'encre sur le papier. Comme cette forme est encrée de manière uniforme par la presse d'impression, il est nécessaire de fragmenter l'image en points de différentes grosseur afin de rendre les nuances de densité de l'image. Ces points sont répartis de manière régulière selon une trame, dont le nombre de points par pouce constitue la linéature. Dans les zones claires, les points sont petits et de fait espacés, donnant l'illusion d'un gris clair; dans les zones foncées les points sont gros et se recouvrent, donnant l'illusion du noir dense.

On l'a appelée similigravure, car elle a remplacé les gravures que les artistes traçaient à la main par des photographies traitées par ce procédé photomécanique, peu avant 1900. L'image tramée était obtenue en projetant l'image en niveaux de gris sur des émulsions à haut contraste à travers des plaques de cristal empreintes de micro prismes. Ensuite, le cliché en simili était obtenu par une attaque à l'acide des zones non exposées d'une plaque de zinc photosensibilisée.



Image 38 Image en demi-tons avant similigravure

Pour reproduire le ventilateur ci-dessus dans un document imprimé, on partira d'un fichier en niveaux de gris à 300 dpi. Il sera converti en points (au mode bitonal) pour l'impression. Le carré délimitant le "détail trames" su 80 :

9. Numérisation des imprimés

figure correspond aux reproductions tramées montrées ci-dessous

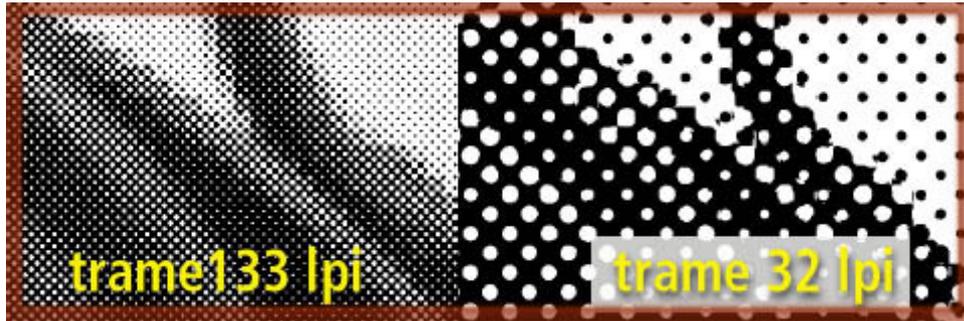


Image 39 Résultat de la similigravure sur la zone détail trame de la figure précédente

La figure ci-dessus montre l'agrandissement de deux trames d'imprimerie appliquées à la reproduction du détail ciblé sur la figure ci-avant. La trame 133 lpi s'applique à un magazine, la trame 32 lpi à une affiche de grande taille.

"lpi" veut dire 'Line per Inch' c'est à dire lignes par pouce. Ce terme provient du procédé de fabrication des trames en cristal, qui étaient rainurées en lignes pour former les pyramides condensant la lumière.

Ce principe est appliqué pour les procédés d'impression par typographie avec une **forme imprimante en relief** nommée « cliché » Les actuels procédés d'impression « offset » ou calcographie utilisent une forme imprimante sans relief constituée d'une plaque d'aluminium recouverte d'émulsion de gélatine. La plaque est mouillée avant encrage, et les zones humides repoussent l'encre, et les autres sèches, prennent l'encre.

Interaction entre la trame et les pixels

Pour constituer une image apparaissant à l'oeil comme à tons continus, la trame va constituer un réseau régulier de points noirs dans le cas des illustrations en « noir et blanc ».

Pour les images en couleurs il faut générer quatre plaques que l'on encrera avec des encres dites primaires, le Cyan, le Magenta, le Jaune et le Noir (CMJN).

Le résultat de l'impression est une image, faite de points alignés selon la linéature (ou fréquence) de la trame. Lors de la numérisation, le numériseur va y superposer sa propre trame, celle des pixels formant l'image numérique. Et c'est là le grand problème, il va en résulter un moirage disgracieux, comme lorsque l'on superpose deux épaisseurs de l'étoffe d'un voileage.

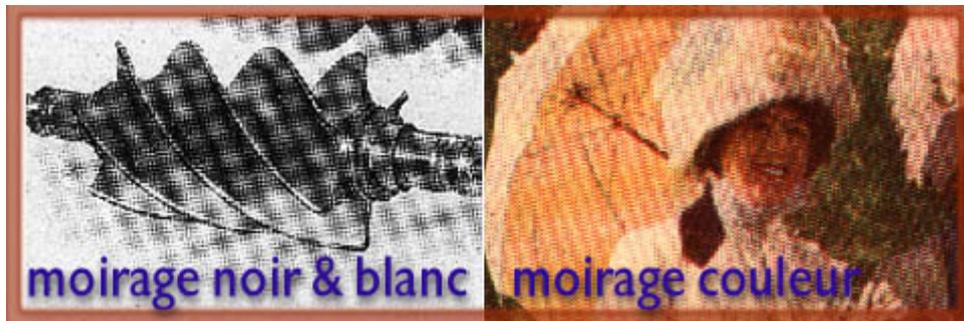


Image 40 Moirage

Lors de la consultation à l'écran, en changeant d'échelle, il se produit des aspects de moirage changeant selon l'interférence entre la trame originale et la grille des pixels de l'écran.

Lors de la réimpression, il se produit le même phénomène entre la trame originale du document et la trame de l'imprimante qui reproduit le document.



Complément : Procédé bitonal

Dans le procédé bitonal appliqué à des originaux comprenant et/ou des gravures et des similigravures à trame grosse (moins de 100 lpi ou trame inférieure à 100 lignes par pouce), on peut prendre le parti de reproduire le tout en 300 dpi ou plus. C'est infiniment plus simple pour l'opérateur à partir du moment où les essais de densité sont effectués. Il est à noter que dans ce cas, la consultation à l'écran n'est pas «agréable», mais l'impression est de la meilleure qualité pour chacune des imprimantes utilisées. La photographie cependant peut montrer des contrastes variables et un aspect granuleux sur les imprimantes à basse définition (moins de 600dpi).



Image 41 Numérisation bitonale en 300 dpi

Numérisation bitonale en 300 dpi d'un catalogue de facture soignée : la lecture est aisée, les images sont un peu dégradées, voire moirées, mais acceptables.



Image 42 Numérisation bitonale en 600 dpi

Numérisation bitonale en 600 dpi : les images sont plus fines et le document est reproductible dans de bonnes conditions, même en imprimerie de type "bulletin d'histoire locale"

Attention !

Sur les exemples ci-dessus, les images ne sont pas réellement au mode bitonal, il s'agit d'une simulation permettant d'apprécier des différences, comme si l'on lisait le document réimprimé, sans s'approcher pour voir les détails.



Complément : Numériser des pages entières en tons continus

La numérisation des pages entières en niveaux de gris avec détramage conduit à rendre un peu flous les caractères du texte, les filets et gravures au trait, au profit du rendu des photos.

Ce procédé a l'avantage d'obtenir assez facilement une page apte à l'impression domestique, mais parfois dans un volume supérieur à celui d'un fichier bitonal. Il faut veiller à détramer efficacement les images, car ce sont les points de la trame qui augmentent la taille du fichier (leur description est plus compliquée qu'une étendue de gris uniforme)



Image 43 Page numérisée en niveaux de gris

Aspect légèrement flou d'une page numérisée en niveaux de gris avec détramage global

Rappel !

Il n'est pas nécessaire de détramer les ouvrages en héliogravure. L'acquisition est fait en mode niveau de gris sans autre artifice qu'une bonne balance des luminosités.



Complément : Choix avancés concernant les numérisations d'imprimés

Avant de se lancer dans les détails exécutoires d'une numérisation d'imprimés, il est nécessaire de se rappeler :

- que la numérisation de l'imprimé sera toujours un compromis entre la qualité et le temps ou l'espace disque que l'on veut y consacrer,
- que la numérisation des imprimés est très techniques (sa réussite se fait au moment même de la numérisation: on veillera donc à la confier à du personnel compétent).

Pour la reproduction de pages composites

Il est possible de reproduire assez simplement une page dans un bulletin d'information, à l'aide d'un fichier photo unique, comme indiqué ci-dessous.

- 1. Acquérir la page entière en mode niveau de gris, sans détramage et en 300dpi: les textes apparaissent assez nets, les photos apparaissent non détramées.



Image 44 Page brute de numérisation

Page brute de numérisation : la photo est moirée, le fond est gris, les caractères sont "mous" - Cette page au format tiff avant retouche fait 2 mégaoctets et, après retouche seulement, 700Koctets. La grande taille du fichier de la page brute est due à la description complexe de la trame et du fond en papier.

- 2. Ouvrir le fichier dans un logiciel de retouche, sélectionner uniquement l'image (ou les images) et appliquer un filtre de détramage en post-traitement sur ces zones.

Intervertir la sélection et augmenter le contraste ou changer les niveaux sur le texte: le fond en papier disparaît ou s'éclaircit et les caractères gagnent en vigueur.

La page est maintenant prête à être intégrée dans une composition ou en ligne, de manière simple. Mais il faut avoir à l'esprit que l'impression de ce document se traduira par un tramage des textes par l'imprimante ou la flasheuse de l'imprimeur.



Image 45 Page retouchée

Page retouchée : la photo est détramée, sa densité est adaptée, le fond est clair et les caractères bien nets

Pour la reproduction fine en ouvrages imprimés

Pour la réédition d'ouvrages anciens en imprimerie, il faudra assurer une reproduction des textes et gravures sans trame, et des photographies (similigravures) avec une trame. IDEM On procède par une bitonalisation intégrale des documents sources, puis par la méthode des "deux couches" si les similigravures ne se traduisent pas bien à la réimpression.

- 1. Acquérir les pages entièrement au mode bitonal en haute définition (1200 à 2400 dpi), la photographie apparaît avec la trame d'origine dans ce premier fichier qui sera utilisé pour la restitution des textes, gravures et filets. Si l'original a été imprimé sur un papier glacé ou couché de bonne qualité, on pourra conserver ce document dans son intégralité en bitonal et le reproduire en offset tel quel, si les points de la trame sont bien formés et nets. Dans le cas contraire, on devra passer à la seconde phase: transformation de la photo tramée en niveaux de gris.



Première couche de la page acquise en mode bitonal: les caractères qui vont être conservés pour la reproduction sont bien nets et seront reproduits sans tramage

Image 46 Première couche

Acquérir à part les similigravures en mode niveau de gris avec un détramage et pratiquer une correction globale de l'image en matière de contraste, niveaux, etc.

- 2. Les fichiers en niveaux de gris sont ensuite replacés par dessus la page bitonale, dans un logiciel de composition (préresse). La réserve de l'image en niveau de gris est très légèrement agrandie pour couvrir le fantôme en mode bitonal, ou celui-ci est effacé dans le fichier et remplacé par un cadre fin plus petit que l'image, juste pour l'aide au positionnement. Ces images en niveaux de gris peuvent être retravaillées et même leur rendu amélioré par rapport à l'original pour augmenter la lisibilité.



La page finie : la photo est détramée, sa densité est adaptée, le fond est clair et les caractères seront imprimés au mode bitonal avec la définition maximale de la photocomposeuse

Image 47 Page finie

Attention !

L'acquisition d'un document au mode bitonal est relativement aisée et peu coûteuse, par contre le procédé des deux couches, nécessitant des retouches et de la mise en page reviendra plus cher, et de plus en plus cher selon le nombre des photos en similigravure. L'imprimeur devra en outre être très compétent (inutile de demander ceci à une imprimerie de labeur).

9.3. Solutions de numérisation

Devant ce problème, la numérisation devient complexe et il faudra jouer des compromis. Par essence, l'imprimé est déjà une reproduction, la numérisation devient une reproduction de la reproduction. Dans ces circonstances, l'objectif devient unique : la diffusion. Les critères de qualité passeront donc au second plan, et l'objectif sera de faire :

un document léger pour sa diffusion,

une reconnaissance (indexation) du texte pour faciliter les recherches.

Les images seront dans la plupart des cas détramées, c'est à dire qu'un genre de flou sera appliqué sur les similigravures pour escamoter les points de la trame. Ce même flou s'appliquera globalement sur les textes, sauf sur certaines machines qui ont des logiciels de reconnaissance permettant de différencier les textes des images se traduisant par une image « tuilée » avec des flous sur les images et un texte en bitonal (ou vectorisé dans le cas d'Adobe Acrobat ®)

Ci-dessous, une image d'un programme de théâtre montre les zones non détramées (encadrées de turquoise) sur un arrière-plan entièrement détramé. La résolution est de 300dpi, ce qui n'est pas indispensable avec les imprimés courants où l'on peut se limiter à 200dpi avec un détramage global. Extrayez l'image de la galerie et affichez la à diverses échelles, la trame entrera en résonance à certains rapports d'agrandissement, mais pas à tous.

En observant cette image, on voit que la trame de l'illustration de la danseuse est peu sensible au moirage, et que le détramage global ne dégrade pas excessivement le document.

Cette solution passe-partout est très pratique, on peut même l'appliquer aux coupures de journaux anciens. Le détramage est à adapter à la taille de la trame sur le numériseur.

Il existe aussi un argument massue à l'utilisation d'un détramage efficace : un fichier détramé voit sa taille après compression réduite drastiquement, parce que la description de zones en demi-tons est moins complexe que celle des points de la trame. Un bon détramage réduit la taille des fichiers d'au moins 50%.



Pierre DARCK
Chef d'orchestre adjoint

L'orchestre de Viennoises est dirigé par une certaine Franzi. Arrivent Fonségur et Moussy. Franzi accepte à souper avec Fonségur, mais c'est tout. Tous deux sont attablés dans un pavillon de verdure quand surviennent Joachim et Lothar, à la recherche de Maurice de Fonségur. Les deux hauts personnages se trouvent soudain entourés de musiciennes. Ils apprennent par elles que Franzi soupe avec un officier. Si c'était lui! Mais voici de nouveaux personnages: Hélène a voulu constater elle aussi l'infidélité de celui qu'elle aime. Dans une conversation avec Franzi, elle apprend sans peine l'amour de la Viennoise pour son mari. Elle s'étonne que cette femme ait si facilement fait la conquête d'un



Les grands artistes, les grands hommes
d'affaires, les personnalités du monde entier
... voyagent par **AIR FRANCE**

Services réguliers au départ de:

MULHOUSE

pour

PARIS et VIENNE



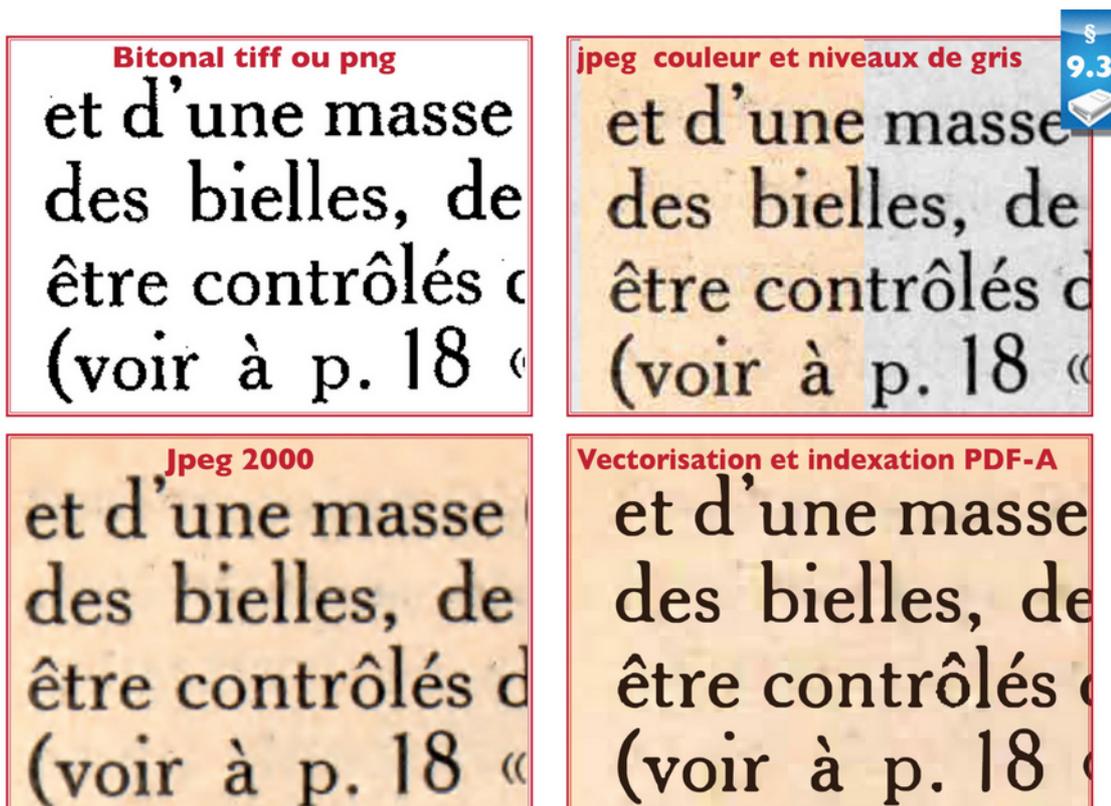
Air France dessert 155
centres répartis dans 70
pays d'Europe, d'Afri-
que, des Amériques et
d'Asie.

AIR FRANCE



AÉRODROME DE BLITZHEIM ET TOUTES AGENCES DE VOYAGES

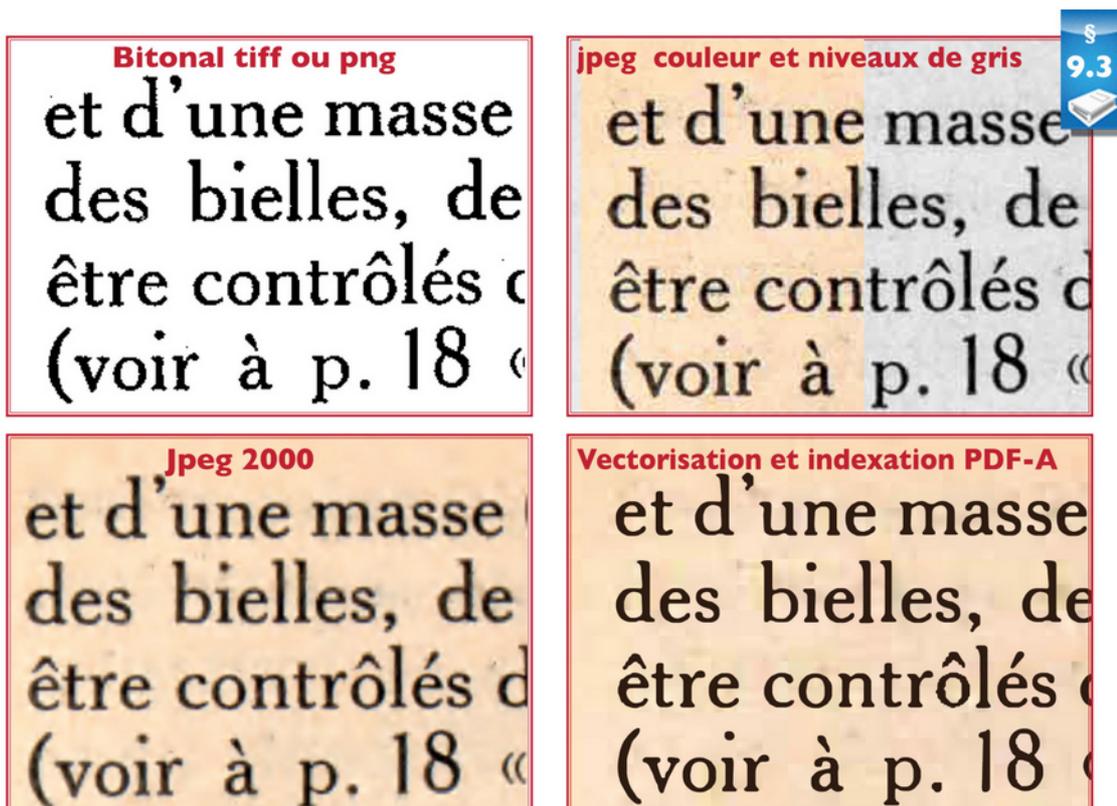
DP 20 Yves Alexandre



Aspects très agrandis des fichiers de diffusion d'imprimés numérisés dans différents modes: le mode bitonal, les différents modes Jpeg avec une taille comparable au fichier bitonal, et enfin l'aspect d'un texte indexé et vectorisé dans Adobe Acrobat. C'est la meilleure solution en manière documentaire puisque l'indexation du texte est sous-jacente, et on peut le copier d'un coup de curseur. (nota dans cette reproduction en pixels, le texte vectorisé n'est pas aussi net que dans le fichier PDF)



Imprimés texte reconnu



Aspects très agrandis des fichiers de diffusion d'imprimés numérisés dans différents modes: le mode bitonal, les différents modes Jpeg avec une taille comparable au fichier bitonal, et enfin l'aspect d'un texte indexé et vectorisé dans Adobe Acrobat. C'est la meilleure solution en manière documentaire puisque l'indexation du texte est sous-jacente, et on peut le copier d'un coup de curseur. (nota dans cette reproduction en pixels, le texte vectorisé n'est pas aussi net que dans le fichier PDF)



Deux exemples de solutions pour numériser les journaux anciens: en couleurs à 200 dpi et au format jpeg, on peut conserver l'aspect du document et une certaine légèreté de fichier, mais la lecture peut s'avérer difficile si le papier a beaucoup jauni. Sur la second image, le même journal au mode bitalonal en 600 dpi sera d'une lecture très agréable surtout au format PDF qui contient un algorithme de lissage à l'écran. L'impression donnera un parfait fac-similé.



Journaux

9.3 9.4. Principe de l'héliogravure

La reproduction des documents à tons continus peut aussi se faire par le truchement de **l'héliogravure**, du grec Hélios pour le soleil = procédé utilisant la lumière.

Pour obtenir la forme imprimante, l'image est décomposée en petits réservoirs d'encre **gravés en creux** sur un cylindre de cuivre. C'est le procédé d'**héliogravure**, qui utilise la lumière pour le report d'image.

Le résultat imprimé est une **héliotypie**, le processus complet est l'héliographie.

Le cylindre est gravé en creux avec des cloisons déterminant des minuscules réservoirs d'encre très fluide qui se diffuse dans le papier donnant l'illusion d'une continuité parfaite des tons.

Il est presque impossible de distinguer la trame d'une héliogravure, tant la linéature est fine. Si la fluidité de l'encre est bien choisie par rapport au papier, l'illusion de continuité est parfaite, car l'encre fuse et recouvre les zones blanches laissées par les cloisonnements.

L'héliographie a fait les beaux jours des livres de photographie tout au long du XXe siècle, et dans l'édition d'images d'art (par exemple les guides Arthaud à Grenoble, Braun à Mulhouse)

L'héliogravure est toujours pratiquée en imprimerie, c'est le procédé utilisé pour les journaux à très grand tirage comme les magazines illustrés, les programmes télé... Mais la monochromie a cédé la place à la quadrichromie, c'est de **l'héliochromie**.

La phototypie

La phototypie est un procédé à mi-chemin entre l'héliographie et l'offset. Il en résulte des épreuves photographiques non tramées, mais dont on peut distinguer en les observant de très près, une forme de réticulation au dessin irrégulier, ce qui nous met à l'abri d'une résonance avec la grille des pixels du numériseur. Ce procédé est utilisé à l'impression des cartes postales, des portefeuilles d'images d'architecture, et certains ouvrages de luxe.

Comment numériser ces originaux ?

Les héliotypies, phototypies et héliochromies sont à numériser sans détramage, ainsi l'on conserve la finesse des détails. Les impressions en héliotypie sont parfois en couleurs. Des journaux et guides touristiques sont imprimés en sépia chaud, ou en « vert américain » ou « bleu de Prusse » pour rompre la monotonie d'une impression en noir.



Image 48 Image en héliogravure

Image en héliogravure agrandie, acquise sans détramage. Le rendu est quasi-photographique. Les publications en héliogravure et cartes postales utilisaient des encres de couleur tel que montré dans la partie droite de cette image. (doc Arthaud- l'Alsace par Hélène Mégret et Robert Minder-impression Braun à Mulhouse)

9.4 9.5. Lithographies et Chromolithographies

La lithographie inventée par Aloys Senefelder au début du XIXe siècle est l'impression à partir d'une pierre de calcaire extrêmement fine, sur laquelle le dessin a été directement exécuté avec un crayon gras. Il suffit ensuite de mouiller la pierre pour différencier les zones qui attirent ou repoussent l'encre. Ce principe est repris dans la plaque offset où la pierre est remplacée par de la gélatine.

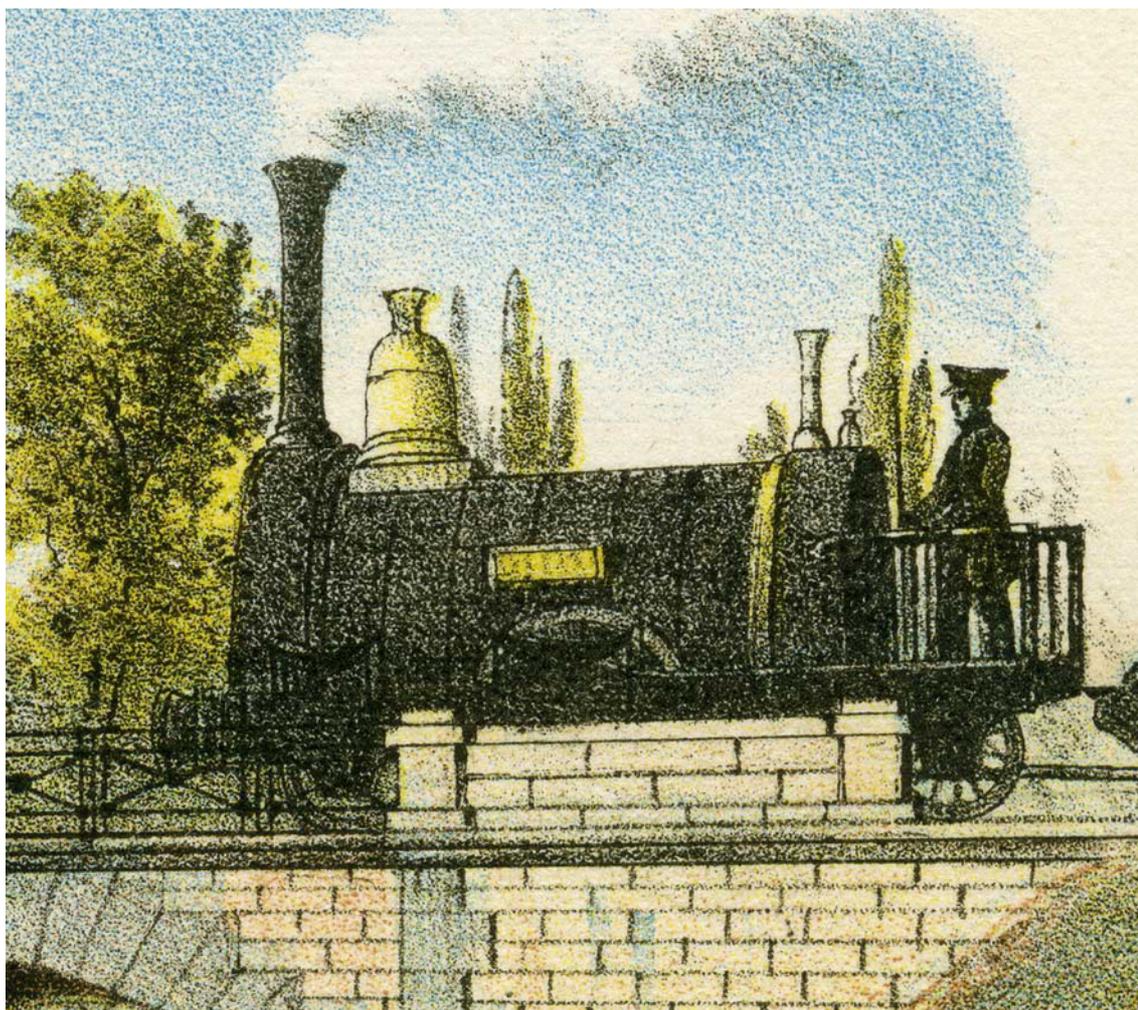
La chromolithographie a été mise au point par le mulhousien Godefroy Engelmann en 1836. Le mérite d'Engelmann sur ses nombreux concurrents est d'avoir mis au point une méthode à la fois théorique : l'emploi des trois couleurs primaires, le cyan (bleu), le jaune et le magenta (rouge), auxquelles on ajoute le noir, pour obtenir toutes les teintes et les nuances possibles, et une indexation précise des couleurs sur les machines d'impression.

Les chromolithographies sont connues de nos jours sous le nom de « chromos ». L'image est composée de points ronds juxtaposés, il y a parfois plus de teintes employées que dans le système CMJN.

Comment numériser ces originaux ?

Les lithographies peuvent présenter des zones d'encre parfaitement noire, en pointillés créés par le grain de collophane appliqué sur la pierre pour donner du grain à l'illustration. La numérisation en couleurs, en 300 dpi constitue une solution standard, qui inclut l'image et le papier. Mais pour la reproduction en ouvrage ou l'élimination des rousseurs, on peut opter pour une numérisation à 1200 dpi et une bitonalisation en suivant le processus utilisé pour les gravures.

Les chromolithographies sont à numériser en couleurs RVB, mais on s'aperçoit que certaines chromolithos utilisent des encres tellement vives qu'il n'est pas possible de les reproduire sans qu'un affadissement apparaisse au premier coup d'oeil. Cela est dû au fait que ces encres peuvent échapper au spectre RVB (spectre incident), vu qu'il est à l'opposé du spectre CMJN (couleurs complémentaires). Les encres actuelles sont beaucoup plus ternes.



Détail d'une chromolithographie de Engelmann, le chemin de fer de Mulhouse à Thann, 1839, collection SIM, cabinet des estampes de la bibliothèque municipale de Mulhouse.

Observer l'arrangement en foisonnement de points de couleur, il n'y a pas de trame régulière. Un principe analogue existe pour les imprimantes à jet d'encre.

Les plans en centre d'archives sont décomposés en deux grandes familles:

- les tracés,
- les tirages.

On appelle **tracés** tous les plans "originaux" qui ont été élaborés par le dessinateur, sur des supports divers: le papier fort, le papier entoilé, le calque.

Les plans modernes issus de la DAO (Dessin Assisté par Ordinateur) sont imprimés sur du papier ordinaire par des imprimantes à jet d'encre ou à poudre. On les placera dans la catégorie des tracés originaux.

On appelle **tirages**, les épreuves diazoïques ou de Van Dycke exécutées par un procédé de reproduction héliographique. Le papier sensible est insolé (lampe à ultra-violet) à travers le tracé original sur son calque ou papier mince. Le papier sensible est ensuite révélé pour obtenir un tirage destiné à l'utilisateur de l'information.

Le contre-calque retouché à l'encre de chine est à considérer comme un tirage.

Les tirages diazoïques se dégradent dans le temps, sous l'action de la lumière. Le pâlissement est irréversible.

Les tirages cyano-fer (ferricyanure) fond bleu, trait blanc sont aussi sensibles à la lumière, mais se régénèrent dans le noir et en présence d'humidité.

10.1 10.1. Plans tracés, cartes

Domaine de définition

On regroupera dans cette famille tous les plans originaux tracés sur papier ou calque. Les cartes originales, mais aussi les tirages sont associées parce que leur typologie est analogue aux plans en couleurs.

Plans tracés au crayon

Ces plans sur papier ou sur calque ont des nuances délicates et des traits très fins. La numérisation se fera en tons continus. Le mode couleurs sera préférable, car il permet de mieux discriminer le trait du papier, surtout s'il est jauni, et une compression JPEG amènera à une économie intéressante d'espace disque.

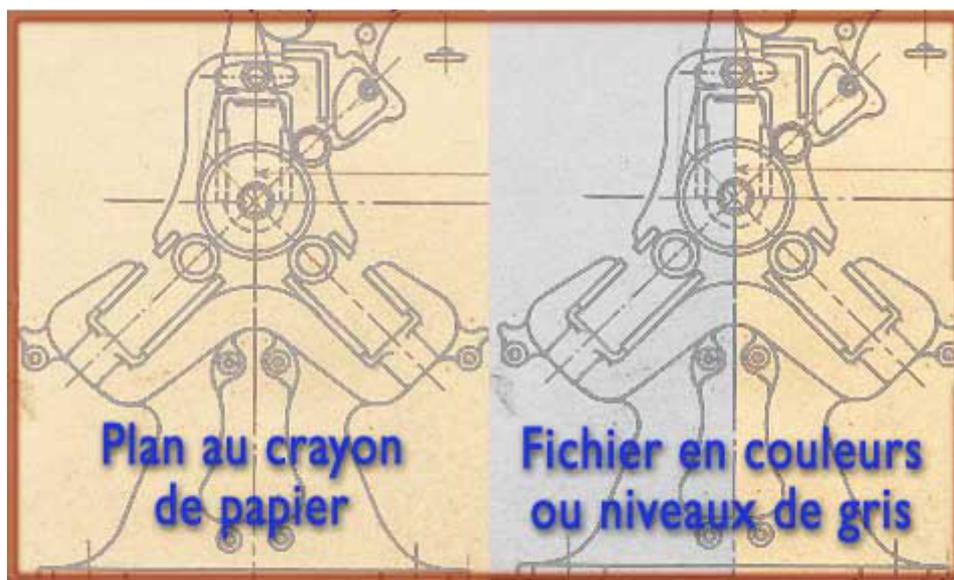


Image 49 Plans tracés au crayon

Les plans à l'encre de chine, sur calque le plus souvent, ou sur papier, sont particulièrement adaptés à la traduction par un fichier bitonal. L'acquisition est aisée, mais la fragilité du calque desséché peut poser des problèmes de manipulation (cassure des plis éventuels).

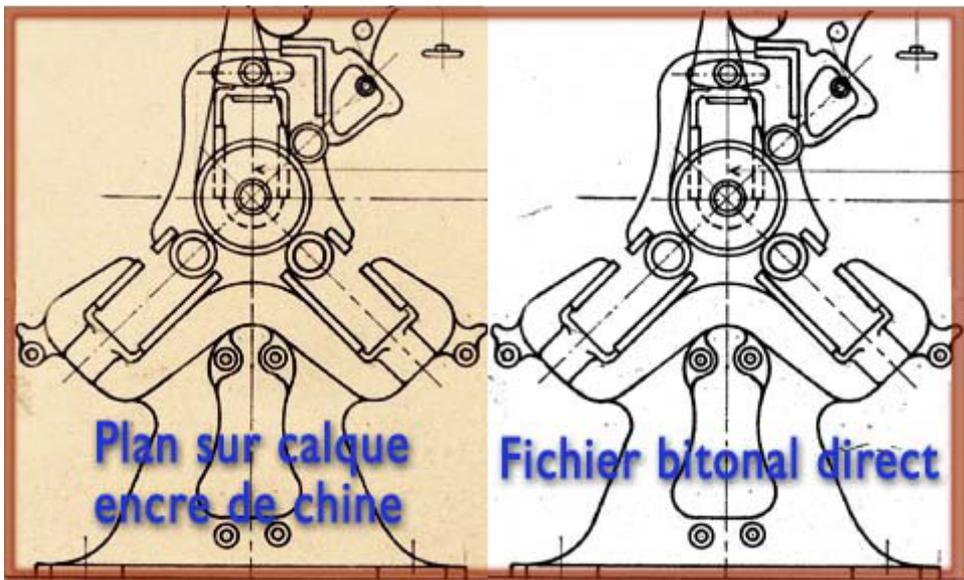


Image 50 Plans tracés à l'encre de chine, sans lavis ni coloriage

Les plans coloriés sont réalisés sur papier fort ou entoilé. Ils représentent le plus souvent des ensembles architecturaux ou de machines.

Les plans coloriés sont à acquérir en couleurs.

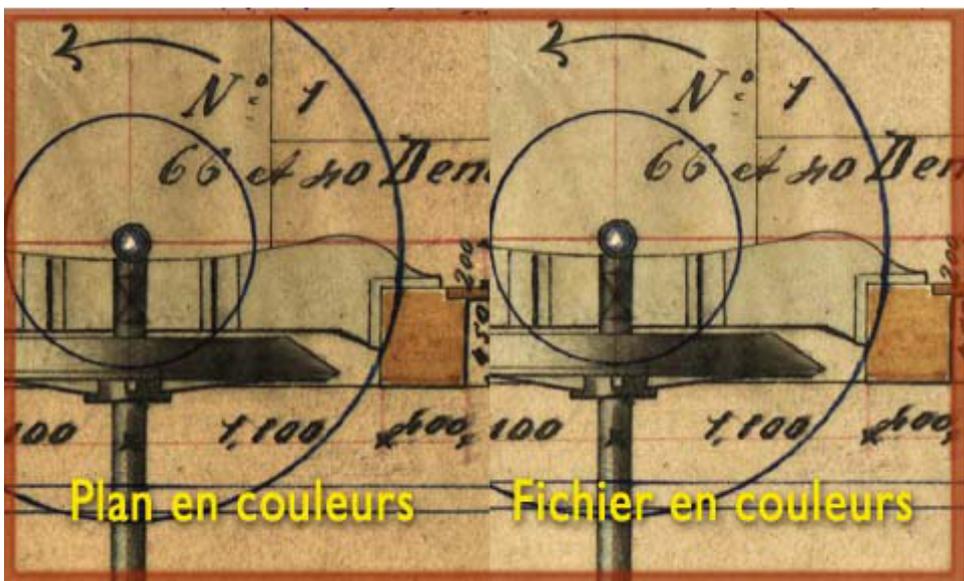


Image 51 Plan et fichier en couleurs

10.2 10.2.Tirages de plans

Tirages au ferricyanure ou aux sels d'argent

Les plans au ferricyanure sont bleus, les traits apparaissent en blanc.

Les plans aux sels d'argent ont un fond brun et des traits blancs.

On procède à une acquisition bitonale, puis le fichier est inversé pour obtenir un plan que l'on peut imprimer facilement.

Il peut exister une exception pour les plans au ferricyanure qui ont été surchargés au crayon rouge ou jaune: il faudra les numériser en couleurs. Les fichiers de diffusion avec une importante compression jpeg demeurent très lisibles malgré les artefacts.

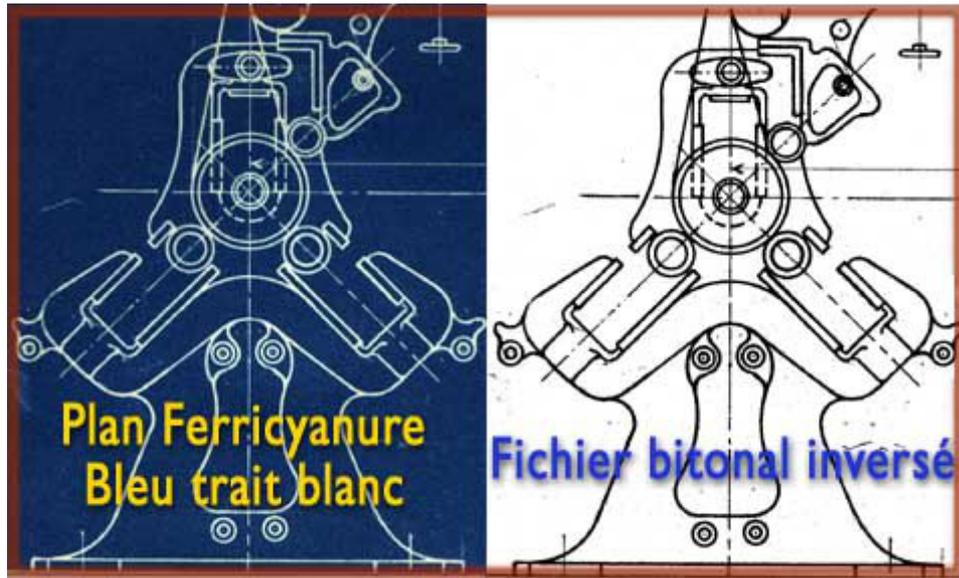


Image 52 Tirages au ferricyanure

Tirage diazo violets

Les plans diazo violets sont à traiter en bitonal.

On passera au fichier à tons continus en couleurs si le fond est foncé (tirage sous-exposé ou traits ténus) afin de restituer les détails incertains.

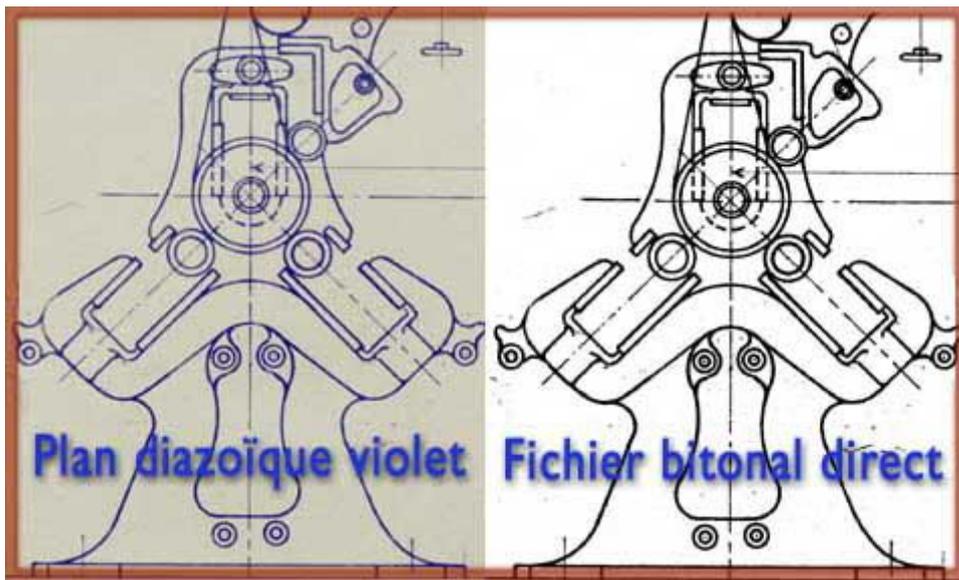


Image 53 Tirage diazo

Tirages diazo sepia et contrecalques

Les "contre-calques" diazoïques et certains tirages sur papier présentent un aspect trait brun sur fond papier, avec du tain, car on a tendance à les sous exposer pour ne pas perdre les traits fins.

Avec un bon réglage de seuil bitonal, on arrive à les traduire "au trait". Dans le cas où certains plans de la série montreraient un contraste insuffisant (contre-calque de Nième génération ou tracé au crayon), on les traiterait en tons continus comme les plans au crayon de papier.

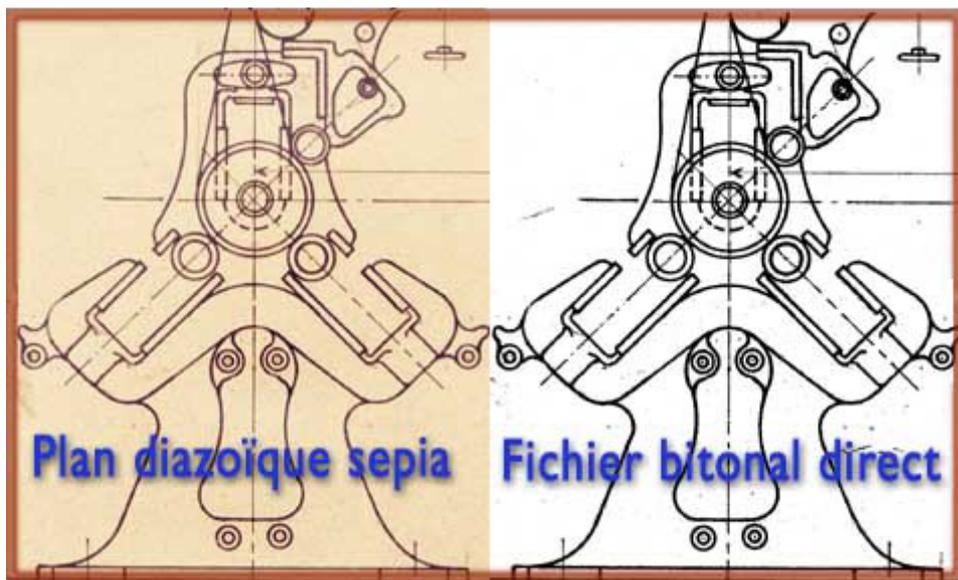


Image 54 Tirages diazo sepia

S 10.2

<p style="text-align: center;">Niveaux de gris (contenu de la microfiche)</p>	<p style="text-align: center;">Bitonal avec seuil de conversion fixe noir/blanc</p>	<p style="text-align: center;">Bitonal avec diffusion de points (simulant les demi-tons) Mode le meilleur pour ce cas</p>																																																						
<table border="1" style="font-size: small; border-collapse: collapse;"> <tr><td>223</td><td>100</td><td>501</td><td>101</td><td>403593</td><td>500</td></tr> <tr><td>2</td><td>5</td><td>5</td><td>4</td><td>N° DLT</td><td></td></tr> <tr><td>203</td><td>10</td><td>21</td><td>21</td><td>30283</td><td></td></tr> </table>	223	100	501	101	403593	500	2	5	5	4	N° DLT		203	10	21	21	30283		<table border="1" style="font-size: small; border-collapse: collapse;"> <tr><td>223</td><td>100</td><td>501</td><td>101</td><td>403593</td><td>500</td></tr> <tr><td>2</td><td>5</td><td>5</td><td>4</td><td>N° DLT</td><td></td></tr> <tr><td>203</td><td>10</td><td>21</td><td>21</td><td>30283</td><td></td></tr> </table>	223	100	501	101	403593	500	2	5	5	4	N° DLT		203	10	21	21	30283		<table border="1" style="font-size: small; border-collapse: collapse;"> <tr><td>223</td><td>100</td><td>501</td><td>101</td><td>403593</td><td>500</td></tr> <tr><td>2</td><td>5</td><td>5</td><td>4</td><td>N° DLT</td><td></td></tr> <tr><td>203</td><td>10</td><td>21</td><td>21</td><td>30283</td><td></td></tr> </table>	223	100	501	101	403593	500	2	5	5	4	N° DLT		203	10	21	21	30283	
223	100	501	101	403593	500																																																			
2	5	5	4	N° DLT																																																				
203	10	21	21	30283																																																				
223	100	501	101	403593	500																																																			
2	5	5	4	N° DLT																																																				
203	10	21	21	30283																																																				
223	100	501	101	403593	500																																																			
2	5	5	4	N° DLT																																																				
203	10	21	21	30283																																																				

Ces images proviennent de la numérisation d'une microfiche filmée à partir d'un PLAN INAPTE AU MICROFILMAGE: écritures mal formées, traits ténus, parties colorisées (les flèches). Le mode niveaux de gris est le moins lisible et la restitution varie selon les imprimantes. Le mode bitonal à seuil fixe a effacé les traits de cotes. Le mode bitonal à diffusion de points restitue et sa reproduction est identique sur toutes les imprimantes. C'est le mode utilisé dans l'industrie.



Plans à faible contraste

10.3 Moyens de numérisation

La numérisation des plans, cartes, affiches peut se faire avec les numériseurs de grand format définis au chapitre 5.7. Pour les originaux en bon état, les numériseurs à défilement donnent les meilleurs résultats.

Lorsque les originaux sont dégradés, fragiles, raides et difficiles à déplier, la numérisation sur un numériseur à caméra dynamique représente une bonne solution.

10. Numérisation des plans et cartes

Les centres d'archives utilisent aussi la reproduction avec une chambre photo numérique. L'original est maintenu sur un plan vertical par exemple avec des aimants ou avec une table à aspiration, puis il est photographié en une ou plusieurs prises de vues. C'est un excellent procédé pour les affiches et estampes de taille moyenne. Pour les cartes modernes, ce procédé montre vite ses limites, les petites écritures ne sont pas lisibles faute de résolution de l'appareil de prise de vues.

La section qui suit commente les solutions de numérisation définies dans le cahier des charges pour la numérisation défini par la Direction des Archives de France. Un complément explicatif est inclus à ce chapitre pour mieux comprendre les expressions du contenu des colonnes de solutions de ce document.

Elle s'adresse aux archivistes qui s'engagent dans un plan de numérisation, ou qui veulent numériser à l'unité quelques archives en appliquant des normes reconnues et constantes.

Solutions de numérisation prédéfinies :

Les résolutions, les usages, les formats de restitution sont définis dans le site de la Direction des Archives de France, sous forme de tableaux récapitulatifs. En suivant ce lien, ou en se référant à la bibliographie du présent cours, on peut retrouver ces tableaux.

Lien vers le cahier des charges techniques de la numérisation :

<http://www.archivesdefrance.culture.gouv.fr/static/1309>⁶

11.1 11.1. Introduction aux choix faits dans les solutions de la DAF

Le document fournit des recommandations sur les caractéristiques techniques des fichiers images produits par la numérisation en fonction de la nature des documents traités et des usages envisagés. Ces recommandations, qui s'appliquent aux cas les plus fréquemment rencontrés, restent indicatives et peuvent être adaptées selon les objectifs précis poursuivis.

Un classement par grandes familles rassemblant des typologies cohérentes entre elles a été adopté : il permet ainsi de déterminer des lots lors du lancement d'un appel d'offres, afin d'optimiser la qualité en fonction des prestataires plus spécialisés dans un domaine que dans un autre.

Pour chaque type de document à numériser, trois usages distincts (et donc trois spécifications techniques différentes) sont en principe pris en compte dans ces tableaux de solutions :

- La conservation : il s'agit d'obtenir une reproduction du document la plus fidèle possible à l'original à des fins de conservation pérenne et d'impression de haute qualité.
- La diffusion : il s'agit de réaliser, à partir du fichier de conservation, une image du document en fonction de la qualité que l'on souhaite mettre à disposition du public. Cette reproduction doit permettre une transmission fluide sur les réseaux et par conséquent l'affichage rapide à l'écran des informations essentielles du document original, auquel les utilisateurs ne pourront plus accéder directement. Cette reproduction permettra également des travaux d'impression courants .
- La visualisation : à l'inverse du fichier de diffusion, aucune impression de qualité ne sera possible à partir de ce fichier (notamment en raison de droits d'accès restreints pour certaines oeuvres comme les affiches, les photographies de grands noms...).

Selon les cas, un plan de numérisation peut viser dès l'origine ces trois cas de figure ou seulement une partie d'entre eux. De plus, la technologies de diffusion en ligne à partir de fichiers JPEG 2000 permettent l'utilisation directe du fichier de conservation pour la consultation, sans nécessiter la préparation spécifique d'un fichier de diffusion. De même, le logiciel de base de données en ligne peut générer ou contenir des copies d'aperçu à des résolutions inférieures, en général entre 1 200 et 2000 pixels sur le plus grand côté.

Les cas exceptionnels ou trop particuliers n'ont pas été abordés ici. Il est conseillé de procéder à des tests avec des sociétés compétentes ou spécialisées. **Ne pas oublier de maintenir les métadonnées à l'intérieur des fichiers de diffusion.**

Les formats de fichiers numériques recommandés dans ces spécifications (TIFF, JFIF, JPEG 2000 et PDF) ont été retenus en raison de leur normalisation publique ou de fait, de leur usage généralisé et des avantages qu'ils procurent à l'exploitation (couverture large des documents concernés, optimisation du poids, type de compression...). Le choix du format pour les fichiers de conservation a été en principe le suivant :

- le format TIFF pour la numérisation noir et blanc (2 bits) ; un format PNG peut le remplacer avantageusement à cause de son universalité de lecture.
- le format JFIF/JPEG faiblement compressé pour la numérisation niveaux de gris ou couleur, pour les manuscrits, dactylographies, registres, le rapport qualité/poids étant dans ces cas meilleur ;
- le format TIFF (24 à 48 bits) pour la numérisation couleur, pour les documents cartographiques,

iconographiques ainsi que les photographies en raison du degré de précision supérieur ; le format PNG là aussi peut être utilisé, surtout lorsqu'on veut détourner les cartes postales.

- le format JPEG 2000 semble offrir les avantages du JFIF en terme de poids et du TIFF en terme de qualité lorsqu'il est « sans pertes de données ». Il est cependant d'un usage moins répandu. Il est par conséquent parfois proposé, mais seulement comme alternative avec un autre format. Le format JPEG 2000 est particulièrement adapté aux contenus complexes (conjuguant images et textes, comme par exemple la presse) en raison de sa technologie (compression par ondelettes qui sait distinguer les différentes couches des documents). Ce format peut être encapsulé dans des documents PDF, il devient ainsi « transparent » pour le lecteur de l'utilisateur. Lorsqu'on souhaite produire un fichier rassemblant plusieurs images, par exemple plusieurs pages d'un registre, on utilisera le format PDF/A .

Les formats de fichiers image proposés sont capables de contenir des métadonnées encapsulées, à incorporer au choix de l'utilisateur. Certains champs des métadonnées (taille d'image, résolution, date...) sont automatiquement remplis par les machines de numérisation et les appareils photographiques et demeurent exploitables tant que l'image n'a pas été l'objet d'un copier-coller.

11.2 11.2 Complément d'information sur le contenu des colonnes de solutions de la Direction des Archives de France

Le document de référence peut être trouvé à l'adresse ci-dessous ou dans la galerie associée à ce cours.

Annexe 1 : recommandations sur les caractéristiques des fichiers images

<http://www.archivesdefrance.culture.gouv.fr/static/1309>⁷

1. Colonne 'Rapport de formats'

Le rapport de formats (ou rapport d'agrandissement) est le rapport entre la taille originale du document numérisé et la taille d'impression enregistrée dans le fichier numérique produit. Les documents originaux opaques ne sont en principe pas agrandis ni réduits lors de la numérisation ; on dit qu'ils sont numérisés à leur taille réelle (1/1 ou 100 %). Il en va autrement lorsque la numérisation porte sur des supports destinés à être agrandis lors de la visualisation, comme les films, les diapositives, les plaques de verre, les microfilms... Dans ce cas, le rapport d'agrandissement est supérieur à 1/1. La restitution à l'échelle 1/1 permet aux archivistes de retrouver les dimensions originales du document, sans l'avoir sous les yeux. Il est important de stipuler dans les marchés de numérisation, que la taille apparaisse dans les propriétés EXIF des métadonnées contenues dans l'image.

Il est à noter qu'en cas de prises de vues avec une chambre photographique, il n'y a pas d'échelle, la dimension du document doit être stipulée manuellement dans les métadonnées

2. Colonne 'Résolution de restitution'

La résolution correspond à la précision et à la finesse de détail d'une image numérique. Il importe de parler de résolution « optique », par opposition à la résolution « interpolée », où des pixels sont générés par un logiciel pour augmenter la taille des images, faussant le degré de précision. Par exemple, numériser un document à une résolution de 4800 dpi interpolée par logiciel ne permet pas d'obtenir plus de détails qu'avec une résolution optique réelle de 2400 dpi, mais risquerait au contraire d'engendrer des artefacts dommageables aux bords contrastés.

- Les solutions sont axées sur les résolutions suivantes :
- 300 dpi pour les fichiers de conservation,
- 150 dpi pour les fichiers de diffusion,
- 72 dpi pour les fichiers de visualisation .

Les résolutions pour la conservation et la diffusion s'entendent pour la taille de restitution du document à ses dimensions d'origine, ou à une dimension prédéterminée pour les agrandissements de supports film. Dans ce dernier cas, si l'on décide de pratiquer à l'enregistrement l'agrandissement en ramenant l'image à 300dpi en restitution, il faudrait prévoir une indication de la taille du négatif ou de la diapositive d'origine dans les métadonnées (entrée en texte légende ou en métadonnées EXIF par la machine)

Remarque : La résolution en points par pouce est une notion qui s'applique principalement aux numériseurs, et non aux appareils photographiques numériques, qui produisent des images de taille constante (définition), par exemple 2 000 x 3 000 points. Il est néanmoins possible, lorsqu'on utilise de tels appareils, d'évaluer une résolution approximative. Par exemple, si le document original mesure 10 x 15 pouces (1 pouce = 2,54 cm), la résolution obtenue avec un appareil photographique de définition 2 000 x 3 000 points est de 200 points par pouce.

3. Colonne 'modèle chromatique'

Les modèles chromatiques préconisés - noir et blanc, niveaux de gris, couleurs RVB (Rouge, Vert Bleu), couleurs CIE Lab - déterminent la manière de coder numériquement la tonalité de l'image. Leur utilisation dépend de l'apparence du document à reproduire et du résultat attendu. Un microfilm est naturellement en niveaux de gris, il sera toujours numérisé dans ce mode.

L'espace de stockage occupé par les images couleur compressées en JPEG est inférieure à celui des images en niveaux de gris. La numérisation en couleur doit donc généralement être préférée à la numérisation en niveaux de gris, surtout que les machines actuelles le font « naturellement », il n'y a aucune raison pour qu'il y ait un surcoût.

4. Colonne 'Echantillonnage' (profondeur d'analyse des couleurs)

La profondeur d'analyse, ou échantillonnage, est la quantité d'information utilisée pour représenter chaque point (pixel) de l'image. Elle est exprimée en nombre de bits par pixel, c'est à dire le nombre d'informations 0 ou 1 décrivant la couleur du pixel.

Le but de cette préconisation est de numériser avec la plus grande qualité possible, parce que la manipulation d'un original de valeur est plus coûteuse que sa numérisation proprement dit. Par ailleurs il vaut mieux éviter des transports, des manipulations qui deviennent finalement néfastes aux originaux fragiles : par exemple plaques de verre, originaux en parchemin, enluminés...

Selon les documents à numériser et les objectifs à atteindre, on pourra demander, lors de la phase d'acquisition, d'appliquer une fréquence d'échantillonnage nettement supérieure à la fréquence souhaitée en restitution (" sur-échantillonnage "), par exemple 48 bits par pixel au lieu de 24 bits/pixel, afin de faire des corrections ou des traitements informatiques visant à améliorer le résultat. Cette méthode nécessite une puissance de calcul très importante, elle n'est donc pas applicable à toutes les situations. Dans les tableaux ci-après, seul l'échantillonnage de restitution est précisé. Dans le cas des numérisations de sauvegarde des œuvres d'art, le fichier maître devrait être conservé avec ce « suréchantillonnage » et ses calques de correction parce qu'on ne peut pas encore présager des progrès de la restitution de l'impression ou de l'affichage.

5. Colonne 'profil colorimétrique'

De manière schématique, le profil colorimétrique est la gamme de couleurs qu'un numériseur, une imprimante, un écran etc, peut acquérir ou restituer. Adobe RGB et sRGB sont des espaces de couleurs, de même que ceux décrits par des profils ICC d'appareils photos, d'écrans ou d'imprimantes. Pour ne pas faire l'amalgame avec des profils ICC, généralement personnalisés (c'est en quelque sorte la fiche d'identité colorimétrique de l'appareil caractérisé) les profils préconisés ici sont des standards reconnus à appliquer suivant l'objectif visé : Adobe RGB (1998) pour l'édition, sRGB IEC 6 1966 2.1 pour la diffusion en réseau, Gray Gamma 2.2 pour les images en niveaux de gris.

Le Lab, de son côté, est atypique car c'est un modèle chromatique et un espace colorimétrique à la fois. La couleur Lab définit une couleur constante sans tenir compte du périphérique (moniteur, imprimante, ordinateur ou scanner, par exemple) de création ou d'affichage de l'image. Son spectre est plus étendu que celui du RGB, c'est pourquoi on le préférera pour la reproduction des diapositives et œuvres d'art, mais encore faut-il que le scanner fabrique lui-même un fichier LAB, ce qui est assez rare.

6. Colonne 'format de fichier image'

Le format du fichier image détermine la manière dont les données qui le constituent sont structurées et codées dans la mémoire d'un ordinateur. Il existe de nombreux formats d'images, dont les caractéristiques sont définies ci-avant dans le cours : TIFF (Tagged Image File Format), JFIF (JPEG File Interchange Format), PNG (Portable Network Graphics)... En règle générale, des formats sans pertes de données sont choisis pour les fichiers-maîtres qui constituent l'archive numérique proprement dite (archive permanente).

Dans certains cas, par exemple pour les documents microfilmés, le format de conservation est le microfilm lui-même et il est inutile de prévoir en sus un format de conservation numérique.

Dans la règle de base, la distinction entre les deux types de formats avec et sans perte de données se justifie. Seules des vignettes ou des versions dégradées des images numériques sont diffusées, afin de protéger la propriété intellectuelle ou des droits d'exploitation, et améliorer la portabilité.

L'indication du format de fichier dans le cahier des charges doit préciser le nom du format et sa version (exemple : TIFF version 6). Il importe, une fois le choix du format arrêté, de maintenir ce choix tout au long de l'opération de numérisation des documents concernés, pour une même typologie donnée. En effet, un changement de format (par exemple de TIFF à JPEG/JFIF) au milieu d'une opération peut occasionner des problèmes de traçabilité, de nommage, voire des difficultés en cas de retraitements ultérieurs (par exemple pour des opérations de reconnaissance de forme ou des ajustements de tonalités qui ne sont plus possibles après pertes de données).

7. Colonne 'type et taux de compression'

Les techniques de compression sont des méthodes destinées en priorité à réduire l'espace nécessaire pour stocker des fichiers numériques.

Certains types de compression sont dits sans perte de données, lorsque l'information d'origine est entièrement conservée. C'est le cas de la compression LZW (Lempel-Ziv-Welch), qui peut être utilisée pour des images en couleur, ou encore de la compression CCITT T.4 groupe 3, qui s'applique aux images en noir et blanc dites bitonales (type fax). En règle générale ces compressions sont appliquées aux fichiers maîtres (archives permanentes)

D'autres types de compression sont dits avec perte de données, lorsqu'une partie de l'information d'origine est perdue. C'est le cas de la compression JPEG par exemple. Dans ce cas, il est possible de faire varier l'intensité de la réduction appliquée, donc de la perte d'information consentie.



Complément : spécifier le taux de compression JPEG/jfif

L'indication du taux de compression JPEG n'est pas simple car la norme JPEG elle-même ne définit pas de niveaux de compression. Il existe donc plusieurs manières d'indiquer ce paramètre : ko/dm², rapport, pourcentage, niveau sur l'échelle de Photoshop. La première méthode donne l'assurance d'obtenir des images de taille connue à l'avance en fonction de la taille d'un document (par exemple, si on fixe 200 ko/dm², le poids d'une image obtenue à partir d'un document A4 sera de 21 x 29,7 cm = 6,237 dm² x 200 = 1 247,4 ko). Mais les réglages à pratiquer par le prestataire sont plus complexes car, en fonction des variations d'un document à l'autre, le poids de fichier attendu ne s'obtient pas forcément de la même manière. L'expression du taux de compression par un rapport peut apparaître comme la

11. Solutions de numérisation en tableaux

également de pourcentage de compression (exemple : 60 % signifie que le fichier a été comprimé 2,5 fois). Mais l'usage le plus courant est certainement l'utilisation des niveaux sur l'échelle de Photoshop ; aussi est-ce cette solution qui a été adoptée dans ces tableaux. La compression JPEG 2000, qui ne doit pas être confondue avec JPEG, dont elle diffère entièrement, est plus récente et peut être avec ou sans perte. Le mode commun à tous ces procédés de compression est la réduction de la gamme des nuances, mais jamais du nombre de pixels .

8. Colonne 'poids moyen'

Dans ces recommandations, un poids moyen d'image est fourni. Il s'agit d'une taille indicative, déterminée sur un échantillonnage de cas réels, après compression. Cette donnée permet de définir statistiquement la taille du média de stockage et d'évaluer le temps d'accès pour les serveurs.

9. Colonne 'cadrage et orientation'

Le cadrage indique la partie du document qui fait l'objet de la prise de vue : seulement la partie du document qui contient des informations, ou bien tout le document mais sans bordure (plein cadre), ou encore le document avec une bordure autour. Lors de la numérisation, le fichier maître doit restituer le document dans son intégralité, y compris avec le cadre ou sa présentation sur un support collé par exemple. La résolution de restitution s'entend toujours par défaut avec un plein cadrage du document à numériser. Il faudra par conséquent en tenir compte si le choix pour le cadrage à l'acquisition est différent, notamment lors de l'introduction de mires, échelle centimétrique ou autres chartes nécessaires à des contrôles. Pour le cas des microfilms par exemple, le cadrage peut-être plein cadre du document original reproduit ou de la vue microfilmée, le rapport de ce fait peut-être différent et cela aura une incidence sur la résolution à appliquer.

L'orientation correspond au sens de l'image. Ce peut être le sens adopté à la numérisation pour des raisons pratiques s'il s'agit de fichiers maîtres, sauf dans le cas où l'orientation ne peut être définie en regardant l'image : les photos industrielles de détail par exemple. Le fichier de diffusion, sera toujours livré dans le sens de lecture.

Certains prestataires négligent l'orientation des originaux et les numérisent « de travers ». Il apparaît utile de spécifier dans le cahier des charges un défaut d'orientation maximal et de refuser lors du contrôle qualité les images présentées en « chamboule tout » ; la tolérance maximale nous semble être 3° sur le plus grand côté, ou le côté de la reliure.

10. Colonne 'cadrage et orientation'

Les traitements appliqués aux images une fois la prise de vue effectuée sont multiples : découpage de pages, détramage, recadrage, rectifications colorimétriques, améliorations de la netteté, redressement des images, effacement des accessoires de fixation, etc. Il implique un temps machine ou humain en sus de la numérisation, ce qui génère un coût. De ce fait, il doit être appliqué de préférence sur des quantités de documents limitées, ou être pratiqué avec des outils de reformage massifs, comme les macro-commandes à exécution programmée .

Il existe trois familles d'opérations de post-traitement :

Les opérations triviales comme le redressement, le recadrage, l'introduction de métadonnées, sont appliquées sans réserve au fichier maître.

Les transformations qui peuvent changer l'aspect de l'image comme le filtrage colorimétrique, la retouche des outrages du temps sont appliquées de manière graduelle. Ces opérations peuvent se faire sur des calques de réglage, qui soit seront conservés dans le fichier maître, soit aplatis après avis favorable du conservateur. L'avantage du calque de réglage ou du calque de retouche est de ne pas altérer l'image originale, l'inconvénient est l'espace disque important occupé par un fichier multicalques.

11. Colonne 'Images assemblées'

La numérisation de très grands documents peut nécessiter plusieurs prises de vue, qui seront ensuite assemblées. Les différentes vues produites pour aboutir à l'image reconstituée ne doivent pas nécessairement être conservées, notamment si la résolution sur le fichier une fois assemblé est la même que sur les parties non assemblées. Pour le cas contraire il sera peut-être judicieux et prudent de garder les fichiers des parties non assemblées. Dans tous les cas, une vérification attentive devra être exercée sur la justesse et la conformité de l'assemblage. Il devra y avoir un minimum de déformation dans la taille du fichier en regard de l'original, et on ne devra pas distinguer par des différences de densité, de contraste ou de netteté l'assemblage (à la coupure et aux alentours).

Ce chapitre n'est en rien une bibliographie exhaustive, elle est constituée de retours d'expériences servant à donner quelques conseils pratiques.

12.1 Numérisation, sites en ligne

Cahier des charges de la numérisation, Direction des Archives de France

Écrire un cahier des charges de numérisation du patrimoine : guide

<http://www.archivesdefrance.culture.gouv.fr/static/4132>⁸

Annexe 1 : recommandations sur les caractéristiques des fichiers images

<http://www.archivesdefrance.culture.gouv.fr/static/1309>⁹

Une association de digitaliseurs pour la généalogie: utilisation du "bookeye" (anglais)

<http://www.rod-need.co.uk/index.htm>¹⁰

Robotisation de la numérisation des documents avec un numériseur à livres équipé d'un tourne-pages, par l'université de Stanford

<http://www-sul.stanford.edu/depts/diroff/DLStatement.htm>¹¹

Principe de base de l'image numérique, propriétés physiques de la lumière, encodages...

<http://perso.univ-rennes1.fr/pierre.nerzic/IN/Cours/S4P4%20-%20Outils%20IN.html>¹²

Documentation concernant les numériseurs

Doc. du numériseur qui tourne les pages automatiquement, fabriqué en Suisse

<http://www.4digitalbooks.com/>¹³

Doc. de numériseurs à livre ouvert : Zeutschel

<http://www.zeutschel.de/>¹⁴

Doc. de numériseurs grand format

<http://www.i2s-digibook.com/fr/>¹⁵

Distributeurs de matériel d'archivage

Spigraph, distributeur de numériseurs spécialisés pour l'archivage (France et Europe)

<http://www.spigraph.fr/>¹⁶

Bibliographie pour les pratiques de la numérisation (en Anglais)

Flatbed Scanner Review (Flaar), un banc d'essai des numériseurs et des publications relatives à l'imagerie de grand format

<http://www.flatbed-scanner-review.org/index.html>¹⁷

Comment choisir un scanner (bibliographie américaine)

<http://www.scansoft.com/scannerguide/resources/books.asp>¹⁸

La numérisation de plans de très grand format

Un exemple de numérisation de plans de navires de grand format en TIFF bitonal (site anglophone). Observez les

8 - <http://www.archivesdefrance.culture.gouv.fr/static/4132>

9 - <http://www.archivesdefrance.culture.gouv.fr/static/1309>

10 - <http://www.rod-need.co.uk/index.htm>

11 - <http://www-sul.stanford.edu/depts/diroff/DLStatement.htm>

12 - <https://perso.univ-rennes1.fr/pierre.nerzic/IN/Cours/applets/GLediteur.htm>

13 - <http://www.4digitalbooks.com/>

14 - <http://www.zeutschel.de/>

15 - <http://www.i2s-digibook.com/fr/>

16 - <http://www.spigraph.fr/>

17 - <http://www.flatbed-scanner-review.org/index.html>

12. Bibliographie en ligne

tailles des fichiers listés sur la page dont l'URL est ci-dessous et ouvrez les ensuite (il faut ensuite télécharger un des fichiers, parce que sa taille en pixels est trop importante pour un afficheur de navigateur internet)

Chacun des plans référencés mesure...de 1,10m à 1,60m de large !

http://dreadnoughtproject.org/French%20Warship%20Plans/Pingouin_EtAl_1884/¹⁹

12.2 Bibliothèques en ligne

Bibliothèques et archives en ligne (exemples)

Nous prenons ici seulement deux cas d'espèce qui illustrent des manières pertinentes de traiter les documents, bien sûr il en existe des milliers d'autres tout aussi valables

Bibliothèque numérique Gallica de la Bibliothèque nationale de France

pour accéder à la bibliothèque de Gallica, activer l'adresse ci-dessous:

<http://gallica.bnf.fr/>²⁰

et orienter la recherche sur l'auteur "Louis Figuié", par exemple La bibliothèque vous propose de télécharger tout ou partie des ouvrages de l'auteur; ce sont des ouvrages contenant des gravures au trait et du texte (catégorie 3).

Pour voir directement un document numérisé en mode bitonal, activer l'adresse ci-dessous. Choisir « les oiseaux ». Pour voir une numérisation en couleurs, choisir un autre ouvrage de la liste. La numérisation est de haute qualité, mais les gravures souffrent d'un affadissement et ne seront pas reproduites dans leur nature bitonale par essence.

<http://gallica.bnf.fr/Search?ArianeWireIndex=index&q=Louis+Figuié&p=1&lang=fr>²¹

Bibliothèque numérique de l'université de Heidelberg

Heidelberg est la capitale Européenne de l'imprimerie, la bibliothèque de l'Université en est le reflet avec un site reproduisant avec une grande fidélité, des incunables remarquables. De plus, des adresses permanentes sont implémentées pour ne jamais perdre la trace de l'archive. Suivre ce lien :

<http://digi.ub.uni-heidelberg.de/diglit/braun1582bd1/0027>²²

Remarquer que sur le document téléchargé, la source et l'adresse permanente subsistent. Ceci est une solution très pragmatique et simple pour retrouver la source...même à partir d'une page imprimée. Observer pour les autres pages texte de cet ouvrage, que le livre est numérisé demi-ouvert à 120° pour que la page reconnue soit la plus plane possible. Ce procédé épargne la reliure. Cependant, dans l'exemple présent, il a fallu quand même ouvrir parfois à 180° pour acquérir les doubles pages.

Bibliographie

Manuel de la numérisation, collectif sous la direction de Thierry Claerr et Isabelle Westeel, éditions du cercle de la librairie, 317pp, 2011.

Conduire un projet de numérisation, collectif, Tec & Doc, 326pp, 2002

19 - http://dreadnoughtproject.org/French%20Warship%20Plans/Pingouin_EtAl_1884/

20 - <http://gallica.bnf.fr/>

21 - <http://gallica.bnf.fr/Search?ArianeWireIndex=index&q=Louis+Figuié&p=1&lang=fr>

22 - <http://digi.ub.uni-heidelberg.de/diglit/braun1582bd1/0027>

13. Evaluation des connaissances

13

Avez-vous bien compris tout ce qui vient de vous être enseigné?

Si vous voulez le vérifier, faites les exercices proposés ci-dessous.

Si vous ne savez pas répondre, ne regardez pas trop vite le corrigé, travaillez à nouveau la (les) section(s) précédente(s) où vous découvrirez les solutions.

Bien sûr, si vous n'y arrivez vraiment pas, vous pouvez consulter les réponses. Ne les lisez pas avec précipitation mais avec une grande attention et surtout essayez de comprendre.

A vous de jouer...

Exercice 1

[Solution n°1 p 118]

Parmi les énoncés suivants concernant les archives virtuelles, lequel est faux?

- a. Les archives informatiques sont uniquement des documents créés originalement au moyen d'équipement informatique
- b. Archives numériques et archives électroniques sont des synonymes
- c. Les photographies numériques sont des archives électroniques
- d. Les dessins assistés par ordinateur sont des documents numériques
- e. Aucune de ces réponses

Exercice 2

[Solution n°2 p 118]

Associer les concepts et les énoncés suivants.

1 - a. Le mode d'encodage bitonal

2 - b. Le format

3 - c. Le nombre pixels

1. Il détermine la résolution (ou définition) de l'image

2. Il s'utilise pour la reproduction des documents et des textes illustrés de gravures

3. Il détermine le type de numériseur à utiliser

Exercice 3

[Solution n°3 p 119]

Parmi les énoncés suivants, lesquels sont-ils vrais ?

- 1. Il est recommandé de numériser le document physique dans la plus petite résolution possible puisqu'une grande résolution exige la création de fichiers énormes, ce qui est plus dispendieux
- 2. Il existe une pratique uniforme pour la numérisation de tous les types de documents
- 3. On peut transformer un document de haute résolution en basse résolution, mais pas l'inverse
- 4. Les formats d'enregistrement sans pertes de données doivent être utilisés pour constituer une banque de fichiers source
- 5. Le choix des formats d'enregistrement doit seulement se faire en fonction du type de documents numérisés et des besoins de l'organisme

13. Evaluation des connaissances

Exercice 4

[Solution n°4 p 119]

Parmi les énoncés suivants, lesquels sont-ils vrais ? (suite)

- 6. La mise de photographies en ligne requiert que l'on offre aux internautes des documents de la plus haute qualité possible, donc comprenant une excellente définition
- 7. Le format PDF, bien que populaire, est à déconseiller puisqu'il s'agit d'un format propriétaire
- 8. Les formats les plus usuels sont les formats libres
- 9. De tous les supports de stockage, ce sont les supports informatiques qui ont la plus longue longévité

Exercice 5

[Solution n°5 p 119]

Associer les besoins de la situation au choix de la qualité de la résolution.

- 1 - a. Vous visez une portabilité (petite taille) permettant un flux rapide dans les réseaux et la diffusion par Internet et vous optez pour une résolution bitonale.
- 2 - c. Vous optez pour une basse définition et la compression des documents. Vous offrez des documents lisibles, mais dont l'impression est de piètre qualité.
- 3 - d. Vous optez pour la lisibilité tout en acceptant une certaine perte de qualité au niveau de la définition.
- 4 - b. Vous optez pour la plus haute résolution possible.

- | | | | |
|--|--|--|--|
| 1. Vous mettez en ligne la copie numérique de lettres du XIIIe siècle que les internautes pourront consulter et faire imprimer pour leurs besoins personnels | 2. Vous mettez en ligne des aperçus des publications de votre organisme. Ils comportent à la fois du texte et des images | 3. Vous numérisez un testament du XVIe siècle en très mauvais état | 4. Vous mettez en ligne des documents administratifs textuels en noir et blanc qui vont servir temporairement dans votre organisme |
|--|--|--|--|

Exercice 6

[Solution n°6 p 120]

Les formats d'enregistrement avec perte de données permettent de créer des fichiers de petite taille qui facilitent le téléchargement. Associer les différents formats aux utilisations que l'on peut en faire.

- 1 - a. JPEG
- 2 - c. PNG
- 3 - d. PDF
- 4 - b. GIF

- | | | | |
|---|--|-------------------------|-----------------------|
| 1. Documents multipages fait de plusieurs formats | 2. Diffusion sur le Web mais sans perte de données | 3. Diffusion sur le Web | 4. Diffusion d'images |
|---|--|-------------------------|-----------------------|

Exercice 7

[Solution n°7 p 120]

Parmi les énoncés suivants, lesquels sont-ils vrais ?

13. Evaluation des connaissances

- | | |
|--------------------------|--|
| <input type="checkbox"/> | 1. Une fois la numérisation d'un microforme terminé, on peut le jeter et utiliser la version numérique car elle est plus facilement accessible |
| <input type="checkbox"/> | 2. Afin d'améliorer la visibilité lors de l'opération de numérisation, il est préférable d'installer le poste de travail près d'une fenêtre afin de bénéficier de la lumière du jour |
| <input type="checkbox"/> | 3. Pour le traitement des images et leur enregistrement sur des supports physiques, il est préférable d'opter pour des unités de travail Mac |
| <input type="checkbox"/> | 4. Dans certains cas, il est utile de traiter l'image après sa numérisation |
| <input type="checkbox"/> | 5. La mise en ligne de photographies, documents fort recherchés, nécessite des copies à haute définition |

Exercice 8

[Solution n°8 p 121]

Parmi les énoncés suivants, lesquels sont-ils vrais ? (suite)

- | | |
|--------------------------|--|
| <input type="checkbox"/> | 6. Lors de la numérisation d'un négatif sur plaque de verre, il faut s'assurer que celui-ci soit bien plaqué contre la vitre de l'appareil afin d'éviter un bris |
| <input type="checkbox"/> | 7. Lors de la numérisation de photographies couleur, il y a parfois lieu d'effectuer des retouches |
| <input type="checkbox"/> | 8. La numérisation des gravures se fait prioritairement en mode bitonal |
| <input type="checkbox"/> | 9. Le format TIFF est recommandé pour la sauvegarde des archives |
| <input type="checkbox"/> | 10. Le format TIFF est recommandé pour la mise en ligne d'images, le téléchargement d'images et leur transmission par courriel |

Exercice

Parmi les énoncés suivants, lequel ne figure pas parmi les bonnes pratiques en matière de numérisation?

- Il est recommandé de numériser le document physique dans la plus grande des résolutions adaptée à l'usage que l'on en fera.
- Il est recommandé d'utiliser le mode RVB (rouge vert bleu) pour la reproduction de tous documents en couleurs
- Il est recommandé de tenir compte de la taille du document pour déterminer le mode colorimétrique à utiliser

Exercice

Parmi les catégories de documents à numériser suivantes, quelle est la catégorie dont l'impression est une reproduction au trait ou bitonale?

- Textes et images pouvant être consultés sans exigence graphique
- Ouvrages ou documents en téléchargement
- Documents iconographiques numérisés pour la sauvegarde ou la reproduction imprimée
- Documents iconographiques pour le Web

Exercice

Je suis un format d'enregistrement avec pertes de données qui est recommandé pour les images numériques en couleur destinées au téléchargement, à une diffusion sur le Web, à un envoi par courriel, ou à un usage bureautique. Qui suis-je?

- Le format TIFF
- Le format PDF
- Le format JPEG
- Le format GIF
- Le format PNG



§
4.2
(png)

Acquisition de la carte postale sur fond noir

Résultat du détourage et enregistrement au format png

Exemple de traitement des cartes postales anciennes: l'acquisition se fait sur un fond noir mat, puis ce fond est éliminé avec un outil «baguette magique» qui se trouve dans la plupart des logiciels de retouche photo. La photo est enregistrée avec un fond transparent au format png, ce qui permet de conserver ces propriétés de transparence, et sans pertes de données pour le fichier maître. Cette CPA colorisée sur une similigravure comprend des trames qui peuvent légèrement moirer à l'écran à certaines échelles.



Cartes postales au format png



Numérisation d'un plan sur un numériseur de grand format «suprascan» à caméra dynamique. Le bloc lumineux est placé à proximité de l'original pour un éclairage symétrique sans effet d'ombre. La caméra équipée d'un capteur linéaire scrute le document à travers la fenêtre d'exposition. L'ensemble du bâti caméra et source de lumière se déplace de gauche à droite pour balayer tout le document (Aux Archives de l'Etat en Belgique - Bruxelles photo P.Perrot)

PIAF
PORTAL INTERNATIONAL
RECHERCHE
FRANCOPHONE

Numériseurs grand format



Exemple de poste de travail sur un numériseur de grand format. Le registre est placé sur les plateaux compensés. On peut rabattre la vitre d'exposition pour obtenir une meilleure planéité de l'original s'il n'est pas fragile. Sur le moniteur à droite, le document numérisé. Un petit écran au pied de la colonne renseigne l'opérateur sur le nombre de vues et les modes de fonctionnement sélectionnés. (Aux Archives de l'Etat en Belgique - Bruxelles photo P. Perrot)

PIAF
PORTAL INTERNATIONAL
ARCHIVISTIQUE
FRANCOPHONE

Numériseurs grand format



Cet exemple montre les trois étapes de la numérisation d'un négatif. Le négatif original est acquis en 16 bits niveaux de gris (1), l'inversion directe(2) montre une image plate. Le post-traitement (3) consiste à refaire le travail du tireur au laboratoire: restituer l'image dans une gamme de tons complets du noir profond au blanc pur. Ensuite l'image est conservée en 8 bits soit 256 niveaux de gris. (Photo Jean Serain)



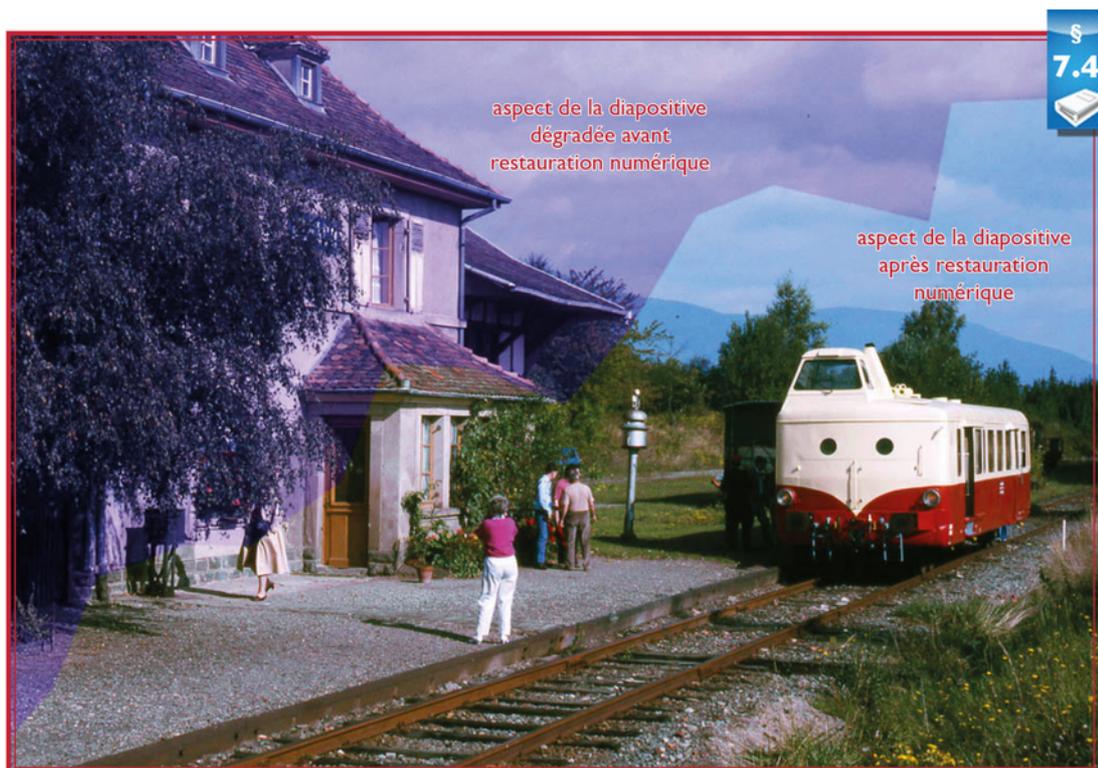
Tirage des négatifs



Cet exemple montre les trois étapes de la numérisation d'un négatif. Le négatif original est acquis en 16 bits niveaux de gris (1), l'inversion directe(2) montre une image plate. Le post-traitement (3) consiste à refaire le travail du tireur au laboratoire: restituer l'image dans une gamme de tons complets du noir profond au blanc pur. Ensuite l'image est conservée en 8 bits soit 256 niveaux de gris. (Photo Jean Serain)



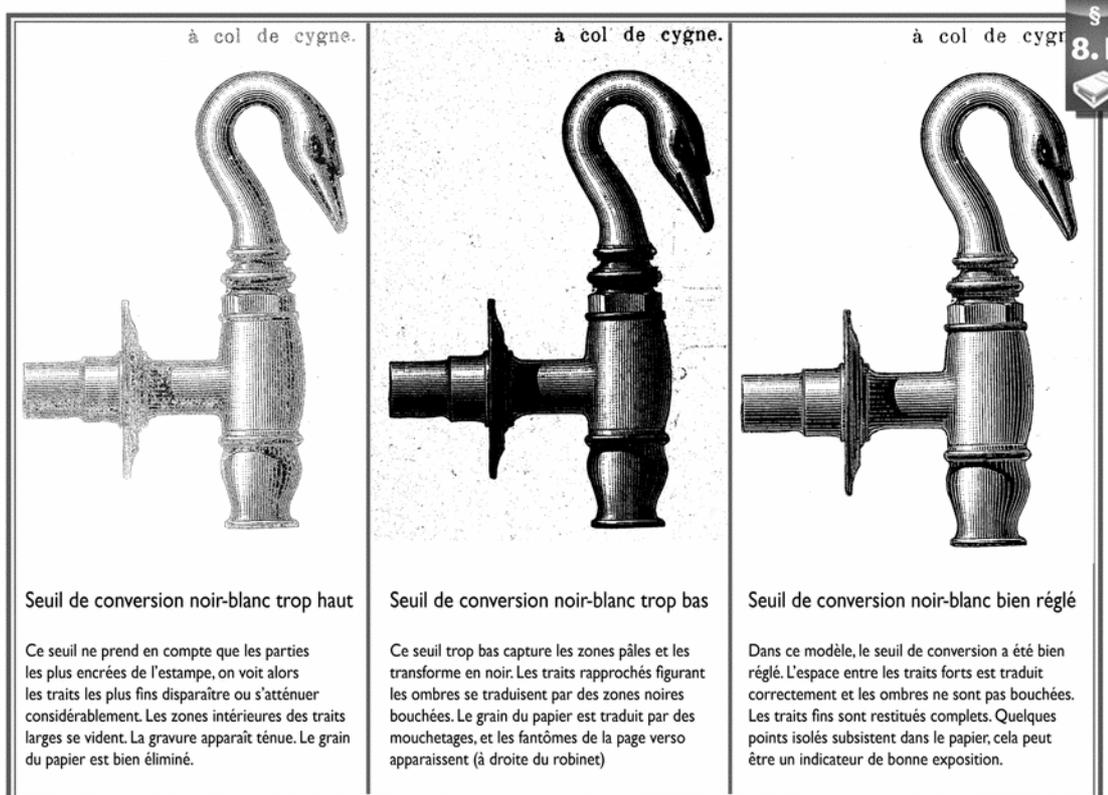
Tirage des négatifs



Cet exemple illustre en vue composite le résultat de la restauration numérique d'une photographie en couleurs avec une «boîte à outils» de commandes standard: niveaux, balance des couleurs, teinte/saturation, correction sélective. Ce procédé est valable pour les diapos et les tirages photo. Les diapositives et tirages en papier couleur se dégradent inexorablement et plus vite dans des ambiances chaudes. (photo P. Perrot)

PIAF
PORTAL INTERNATIONAL
ARCHIVISTIQUE
FRANCOPHONE

Restauration des couleurs



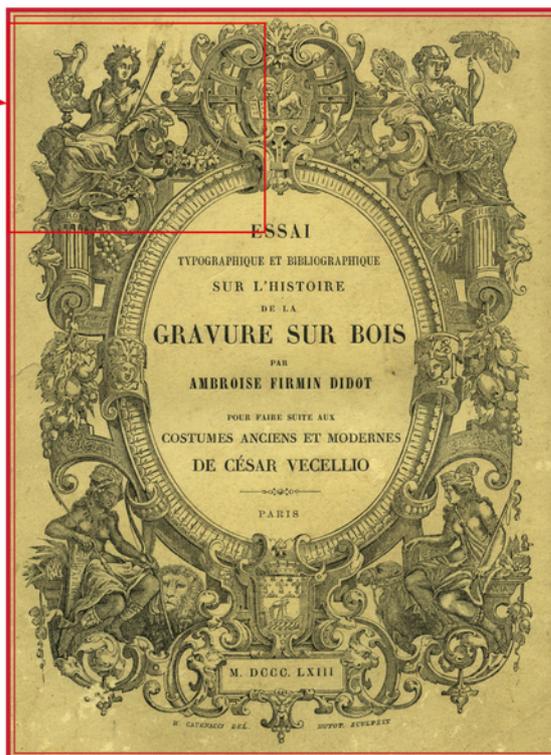
§
8.1

Cette image de la galerie montre en détail les numérisations du robinet à col de cygne du complément du chapitre 8.1.

PIAF
PORTAL INTERNATIONAL
RECHERCHES
FRANCOPHONE

Détermination du seuil bitonal

voir ce détail sur
le fichier bitonal



Cette numérisation montre un ornement de livre de Firmin Didot. Le fichier maître est en 2400 dpi, pour une taille de 108 MO compressé. En couleurs, le document conserve l'aspect du document d'origine mais à cette résolution il est trop grand pour la diffusion. Si on en réduit la résolution, la finesse de la gravure en sera altérée. La solution consiste à extraire la gravure au mode bitonal pour en conserver tous les détails. C'est en quelque sorte une « extraction de l'encre » (fichier suivant dans la galerie)



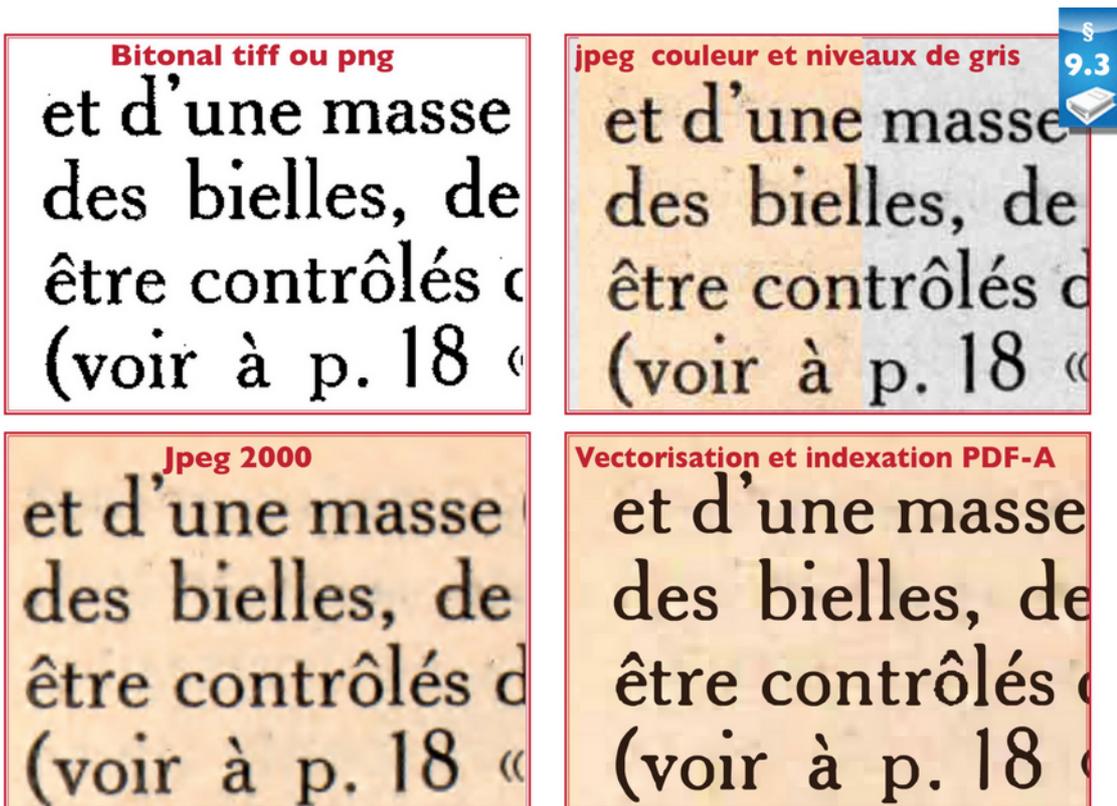
Gravure estampes en bitonal



Voici un fragment de la gravure en mode bitonal en 2400 dpi **agrandi 4 fois** sur cet exemple. Le fichier fini pèse 6,8 MO, soit une taille 15 fois plus petite que celui en couleurs, avec un contenu de grande qualité. Pour la diffusion, un fichier à 600 dpi montre suffisamment de détails. L'impression du fichier bitonal est excellente, testez avec cet exemple. Cette image de la galerie est entièrement au mode bitonal.

PIAF
PORTAL INTERNATIONAL
RECHERCHES
FRANCOPHONE

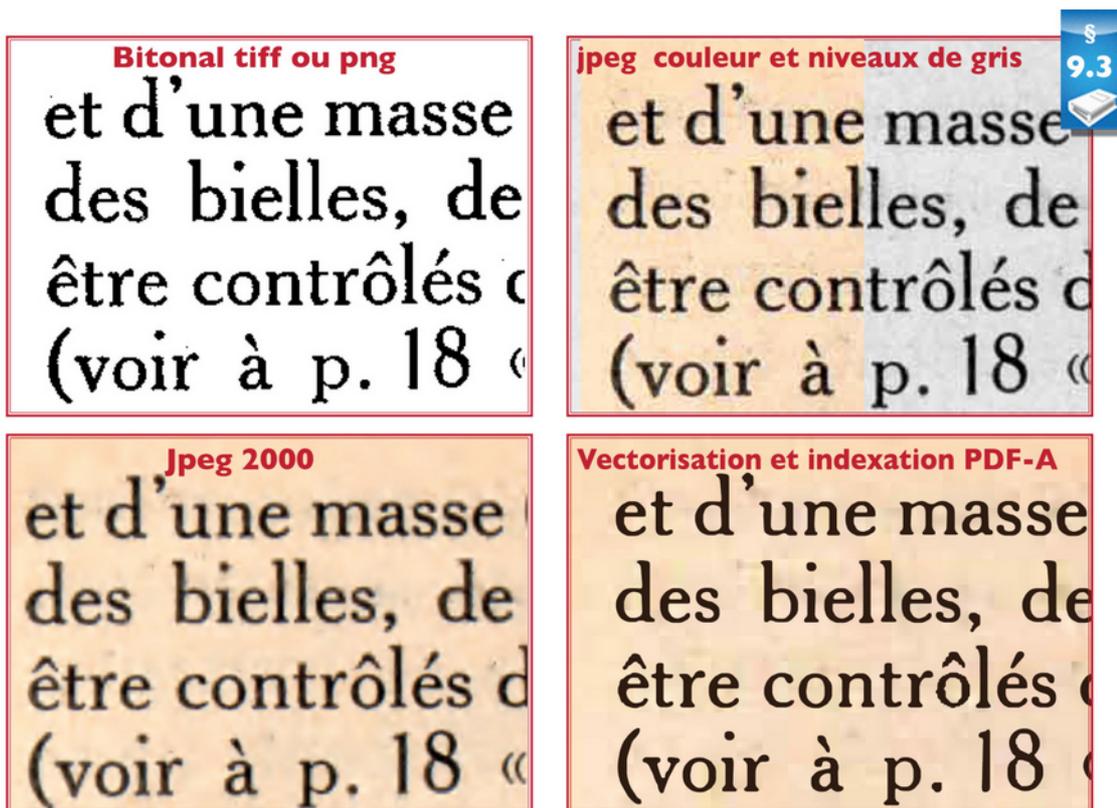
Gravures estampes en bitonal



Aspects très agrandis des fichiers de diffusion d'imprimés numérisés dans différents modes: le mode bitonal, les différents modes Jpeg avec une taille comparable au fichier bitonal, et enfin l'aspect d'un texte indexé et vectorisé dans Adobe Acrobat. C'est la meilleure solution en manière documentaire puisque l'indexation du texte est sous-jacente, et on peut le copier d'un coup de curseur. (nota dans cette reproduction en pixels, le texte vectorisé n'est pas aussi net que dans le fichier PDF)



Imprimés texte reconnu



Aspects très agrandis des fichiers de diffusion d'imprimés numérisés dans différents modes: le mode bitonal, les différents modes Jpeg avec une taille comparable au fichier bitonal, et enfin l'aspect d'un texte indexé et vectorisé dans Adobe Acrobat. C'est la meilleure solution en manière documentaire puisque l'indexation du texte est sous-jacente, et on peut le copier d'un coup de curseur. (nota dans cette reproduction en pixels, le texte vectorisé n'est pas aussi net que dans le fichier PDF)

Paris, 19. zieber Marce d'Aussigny, 31. Juli ds. ihm 400 000 ten.

Er gab da Absicht begeben und verbringen ; gelegenheit hab vervollkomm

In Berlin Gerichtsvollz er unter Hint Kassendefiz hatte. Mme. hin in den nac Paris.

Begleitet v richtsvollzieh Petit Parguet hat sich der stellt. Gleich richtsschreib bene Summe zahlt.

Substitut Haftbefehl d Bourges zur ihn wegen den Anklage ler ist alsdan führt worden.

Ein gefä treibt in de

Blois, 19. hier ein Ein

Paris, 19. zieber Marce d'Aussigny, 31. Juli ds. ihm 400 000 ten.

Er gab dar Absicht begeben und verbringen ; gelegenheit hab vervollkomm

In Berlin a Gerichtsvollz er unter Hint Kassendefiz hatte. Mme. hin in den nac Paris.

Begleitet vo richtsvollzieh Petit Parguet hat sich der J stellt. Gleich richtsschreib bene Summe zahlt.

Substitut B Haftbefehl der Bourges zur K ihn wegen Ve den Anklagezu ler ist alsdann führt worden.

Ein gefäh treibt in der

Blois, 19. Au hier ein Ein



General BERENGUER



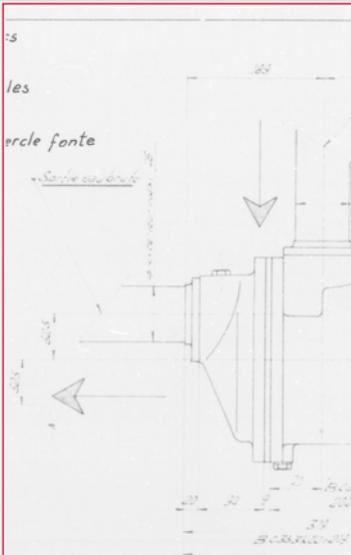
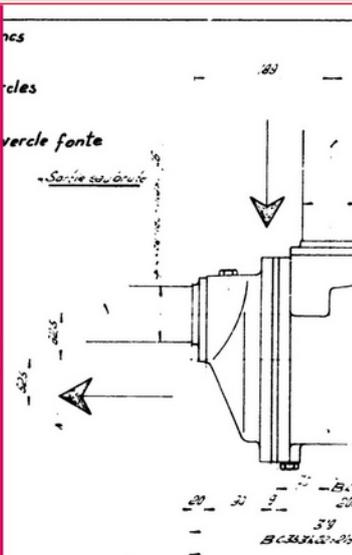
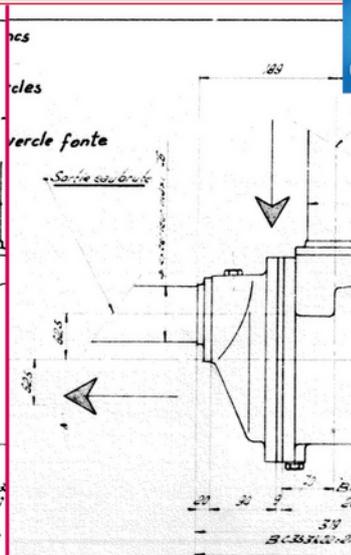
General BERENGUER

Deux exemples de solutions pour numériser les journaux anciens: en couleurs à 200 dpi et au format jpeg, on peut conserver l'aspect du document et une certaine légèreté de fichier, mais la lecture peut s'avérer difficile si le papier a beaucoup jauni. Sur la second image, le même journal au mode bitonal en 600 dpi sera d'une lecture très agréable surtout au format PDF qui contient un algorithme de lissage à l'écran. L'impression donnera un parfait fac-similé.



Journaux



																																																																	
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> </table>	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> </table>	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> <tr> <td>223</td><td>100</td><td>501</td><td>101</td><td>603593</td><td>N°</td><td>DLT</td></tr> </table>	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT	223	100	501	101	603593	N°	DLT
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
223	100	501	101	603593	N°	DLT																																																											
Niveaux de gris (contenu de la microfiche)	Bitonal avec seuil de conversion fixe noir/blanc	Bitonal avec diffusion de points (simulant les demi-tons) Mode le meilleur pour ce cas																																																															

Ces images proviennent de la numérisation d'une microfiche filmée à partir d'un PLAN INAPTE AU MICROFILMAGE: écritures mal formées, traits ténus, parties colorisées (les flèches). Le mode niveaux de gris est le moins lisible et la restitution varie selon les imprimantes. Le mode bitonal à seuil fixe a effacé les traits de cotes. Le mode bitonal à diffusion de points le restitue et sa reproduction est identique sur toutes les imprimantes. C'est le mode utilisé dans l'industrie.



Plans a faible contraste



Développement



Tables de montage

Solution des exercices

> Solution n°1 (exercice p. 101)

- a. Les archives informatiques sont uniquement des documents créés originalement au moyen d'équipement informatique
Bravo, cet énoncé est faux. Les documents numérisés (de la correspondance manuscrite ou une vieille photographie, par exemple) sont considérés comme des archives informatiques si le centre d'archivers ne possède pas les originaux. (chapitre 1.1)
- b. Archives numériques et archives électroniques sont des synonymes
Cet énoncé est vrai. Les archives numériques ou électroniques sont disponibles seulement par le biais de l'utilisation d'un ordinateur et sont en cela synonymes, mais le terme "archives électroniques" paraît désormais démodé (voir glossaire).
- c. Les photographies numériques sont des archives électroniques
Cet énoncé est vrai. Le modèle original ayant servi à les générer n'existe plus et les tirages sur papier en sont les copies et non les sources.
- d. Les dessins assistés par ordinateur sont des documents numériques
Cet énoncé est vrai. Les dessins assistés par ordinateur sont créés à l'origine sur un ordinateur et n'ont pas d'originaux sous forme papier. La copie papier est en fait une reproduction ou une impression du document électronique.
- e. Aucune de ces réponses
Voir les autres possibilités.

> Solution n°2 (exercice p. 101)

1. Il détermine la résolution (ou définition) de l'image	c. Le nombre pixels
2. Il s'utilise pour la reproduction des documents et des textes illustrés de gravures	a. Le mode d'encodage bitonal
3. Il détermine le type de numériseur à utiliser	b. Le format

> Solution n°3 (exercice p. 101)

- 1. Il est recommandé de numériser le document physique dans la plus petite résolution possible puisqu'une grande résolution exige la création de fichiers énormes, ce qui est plus dispendieux
FAUX. La plus petite résolution (économie d'espace sur le disque dur) est souvent insuffisante pour les besoins archivistiques. (chapitre 2.2)

- 2. Il existe une pratique uniforme pour la numérisation de tous les types de documents
FAUX. Une pratique uniforme ne répondrait pas aux besoins archivistiques de respecter la valeur intrinsèque du document. (chapitres 2.4.2 et 3)

- 3. On peut transformer un document de haute résolution en basse résolution, mais pas l'inverse
VRAI. C'est pourquoi il est préférable de sélectionner la qualité de la numérisation en prévoyant les besoins et les utilisations futures. (chapitre 3.2)

- 4. Les formats d'enregistrement sans pertes de données doivent être utilisés pour constituer une banque de fichiers source
VRAI. A partir de ces fichiers maîtres, des copies avec pertes de données, plus petites, peuvent être effectuées afin, par exemple, de mettre les documents en ligne. (chapitre 4)

- 5. Le choix des formats d'enregistrement doit seulement se faire en fonction du type de documents numérisés et des besoins de l'organisme
FAUX. Si le format choisi est incompatible avec les applications courantes la numérisation est un échec. (chapitre 4.3.1)

> Solution n°4 (exercice p. 102)

- 6. La mise de photographies en ligne requiert que l'on offre aux internautes des documents de la plus haute qualité possible, donc comprenant une excellente définition
FAUX. La mise en ligne sans contrôle des documents de haute qualité n'est pas recommandée, car il est pénalement très risqué de faire des reproductions en faisant fi des droits d'auteur. (chapitre 4.3.1)

- 7. Le format PDF, bien que populaire, est à déconseiller puisqu'il s'agit d'un format propriétaire
FAUX. Il est vrai que l'assemblage des fichiers requiert l'achat d'un logiciel spécifique. Cependant, le logiciel de lecture (Acrobat Reader) est distribué gratuitement et demeure aisément disponible. (chapitre 4.3.2.2)

- 8. Les formats les plus usuels sont les formats libres
VRAI. Ces formats ont l'avantage d'être lus au moyen de toutes les applications. (chapitre 4.4.1)

- 9. De tous les supports de stockage, ce sont les supports informatiques qui ont la plus longue longévité
FAUX. En fait, la micrographie et la matérialisation de ces informations les mettent à l'abri de l'effacement par erreur, de l'incompatibilité informatique et de la fragilité des supports d'enregistrement informatiques. (chapitre 4.5)

> Solution n°5 (exercice p. 102)

1. Vous mettez en ligne la copie numérique de lettres du XIIIe siècle que les internautes pourront consulter et faire imprimer pour leurs besoins personnels	d. Vous optez pour la lisibilité tout en acceptant une certaine perte de qualité au niveau de la définition.
2. Vous mettez en ligne des aperçus des publications de votre organisme. Ils comportent à la fois du texte et des images	c. Vous optez pour une basse définition et la compression des documents. Vous offrez des documents lisibles, mais dont l'impression est de piètre qualité.
3. Vous numérisez un testament du XVIe siècle en très mauvais état	b. Vous optez pour la plus haute résolution possible.
4. Vous mettez en ligne des documents administratifs	a. Vous visez une portabilité (netite taille) permettant un

dans votre organisme	vous optez pour une résolution bitonale.
----------------------	--

Réponse a. C'est un document qui se retrouve sous la catégorie 1. On retrouve dans cette catégorie les documents administratifs utilisés dans les réseaux internes ou pour diffusion sur le Web. (chapitre 3.4)

Réponse b. C'est un document qui se retrouve sous la catégorie 4. Ce sont des documents sévèrement endommagés ou dédiés à être imprimés. (chapitre 3.4)

Réponse c. C'est un document qui se retrouve sous la catégorie 2. C'est souvent le choix fait par les services intranet. Les documents sont de piètre qualité pour l'impression mais tout à fait acceptable pour consultation à l'écran. (chapitre 3.4)

Réponse d. C'est un document qui se retrouve sous la catégorie 3. Les centres d'archives ou les bibliothèques peuvent rendre ainsi des documents disponibles en ligne. (chapitre 3.4)

> Solution n°6 (exercice p. 102)

1. Documents multipages fait de plusieurs formats	d. PDF
2. Diffusion sur le Web mais sans perte de données	c. PNG
3. Diffusion sur le Web	b. GIF
4. Diffusion d'images	a. JPEG

Réponse a. Créé pour permettre la diffusion sur toutes les plates-formes et en ligne des images numériques en couleur. (chapitre 4.2)

Réponse b. Réservé à l'usage sur le Web et n'est pas utile pour les activités de numérisation des archives mais seulement pour leur diffusion sur le Web. (chapitre 4.2)

Réponse c. Moins volumineux que le format GIF et de meilleure qualité que le format JPEG ; ce format sera capable bientôt de rentrer en compétition avec le format TIFF utilisé aux archives. (chapitre 4.2)

Réponse d. Le format PDF se prête à plusieurs utilisations, entre autres la compilation de textes avec d'autres fichiers images de tous les formats. (chapitre 4.2)

> Solution n°7 (exercice p. 102)

- 1. Une fois la numérisation d'un microforme terminé, on peut le jeter et utiliser la version numérique car elle est plus facilement accessible
FAUX. En fait, la copie numérisée du microforme reste une copie de travail, le microforme en lui-même est un meilleur médium de conservation. (chapitre 5.7)
- 2. Afin d'améliorer la visibilité lors de l'opération de numérisation, il est préférable d'installer le poste de travail près d'une fenêtre afin de bénéficier de la lumière du jour
FAUX. En fait, peu importe le type de numériseur car la lumière du jour peut causer des reflets ou des taches sur l'image numérisée. (chapitre 5.8)
- 3. Pour le traitement des images et leur enregistrement sur des supports physiques, il est préférable d'opter pour des unités de travail Mac
VRAI. Les Mac peuvent lire et enregistrer sur tous les supports de tous les systèmes d'exploitation, MAC, PC et UNIX, ce qui permet de les relire universellement. (chapitre 5.8)
- 4. Dans certains cas, il est utile de traiter l'image après sa numérisation
VRAI. Un tel traitement permet notamment d'enlever des bordures noires, ainsi que de redresser l'image et d'améliorer son rendu. (chapitre 5.10)
- 5. La mise en ligne de photographies, documents fort recherchés, nécessite des copies à haute définition
FAUX. En fait, les copies disponibles gratuitement pour les utilisateurs sont des copies à basse résolution. La numérisation en très haute qualité reste dans les serveurs du centre d'archives ou sur des supports de sauvegarde comme le CD-ROM.

> Solution n°8 (exercice p. 103)

- | | |
|-------------------------------------|--|
| <input type="checkbox"/> | 6. Lors de la numérisation d'un négatif sur plaque de verre, il faut s'assurer que celui-ci soit bien plaqué contre la vitre de l'appareil afin d'éviter un bris
<i>FAUX. Cela pourrait provoquer un effet de ventouse qui risquerait d'entraîner le bris du document au moment de le retirer de la vitre. (chapitre 7.2)</i> |
| <input checked="" type="checkbox"/> | 7. Lors de la numérisation de photographies couleur, il y a parfois lieu d'effectuer des retouches
<i>VRAI. Les caractéristiques du support entraînent, avec le temps, l'altération des couleurs. Les retouches permettent, dans une certaine mesure, de restituer l'aspect original. (chapitre 7.4.2)</i> |
| <input checked="" type="checkbox"/> | 8. La numérisation des gravures se fait prioritairement en mode bitonal
<i>VRAI. Ce mode évite les altérations dues au mode en niveaux de gris, comme la transformation des zones franches entre le noir et le blanc en flous gris. (chapitre 8.1 et 8.2)</i> |
| <input checked="" type="checkbox"/> | 9. Le format TIFF est recommandé pour la sauvegarde des archives
<i>VRAI. Il possède notamment l'avantage de conserver la gamme complète des couleurs ou de niveaux de gris.</i> |
| <input type="checkbox"/> | 10. Le format TIFF est recommandé pour la mise en ligne d'images, le téléchargement d'images et leur transmission par courriel
<i>FAUX. Le format TIFF ne permet pas de réduire le nombre de couleurs de pixels ; le fichier demeure volumineux, ce qui ralentit le téléchargement.</i> |

> Solution n°9 (exercice p. 104)

Exercice

- | | |
|----------------------------------|---|
| <input type="radio"/> | Il est recommandé de numériser le document physique dans la plus grande des résolutions adaptée à l'usage que l'on en fera. |
| <input type="radio"/> | Il est recommandé d'utiliser le mode RVB (rouge vert bleu) pour la reproduction de tous documents en couleurs |
| <input checked="" type="radio"/> | Il est recommandé de tenir compte de la taille du document pour déterminer le mode colorimétrique à utiliser |

Exercice

- | | |
|----------------------------------|--|
| <input type="radio"/> | Textes et images pouvant être consultés sans exigence graphique |
| <input type="radio"/> | Ouvrages ou documents en téléchargement |
| <input type="radio"/> | Documents iconographiques numérisés pour la sauvegarde ou la reproduction imprimée |
| <input checked="" type="radio"/> | Documents iconographiques pour le Web |

Exercice

Solution des exercices

Le format TIFF

Le format PDF

Le format JPEG

Le format GIF

Le format PNG

Glossaire

Cloud ou nuage

Le cloud ou l'informatique en nuages en français, est une technologie qui permet de mettre sur des serveurs localisés à distance des données de stockage ou des logiciels qui sont habituellement stockés sur l'ordinateur d'un utilisateur, voire sur des serveurs installés en réseau local au sein d'une entreprise

Fichier

Ensemble de fiches ou d'enregistrements informatiques contenant des informations de même nature.

GAMUT

gamme de couleurs, spectre caractéristique.

Manuscrit

Texte écrit à la main ou, par extension, élaboré par son auteur à l'aide d'une machine à écrire ou d'un ordinateur.

Numérisation

Procédé électronique de reproduction d'un document d'archives sous forme de document électronique.

passé-bas

Filtre passe-bas: dispositif qui limite la résolution à une valeur supérieure limite. Tout ce qui est plus grand en résolution sera recalculé, et une nouvelle image en résolution plus basse sera incorporée au fichier final.

Type de document

Catégorie de documents distingués en raison de critères communs, matériel (par exemple: aquarelle, dessin) ou fonctionnels (par exemple: journal, livre de compte, main-courante, registre de délibération...).

Bibliographie

[Pour la numérisation] Conseils à voir dans le chapitre 12 de la deuxième section du présent module.

[Pour le microfilmage] *La conservation entre microfilmage et numérisation. Actes des journées patrimoniales organisées par la BNF tenues à Sablé les 8 et 9 novembre 1993.*

[Pour le microfilmage] LEISINGER, ALBERT H. *La microphotographie aux Archives.* Bruxelles, ICA, 1975 (traduit de l'anglais par Christian GUT).

[Pour le microfilmage] *Recueil de normes françaises. Supports d'informations, micrographie.* AFNOR, 5ème édition, 1992.

[Pour le microfilmage] ROPER, MICHAEL. *Directives pour la préservation des microformes.* Paris, édition française préparée par Georges WEILL., ICA, Etudes n° 2., 1990.

[Pour le microfilmage] *SCOM. Méthodologie d'emploi de la micrographie.* Paris, ministère de l'Economie et des Finances, direction du Budget, 3ème édition, 1983.

[Pour le microfilmage] FAVIER, JEAN, NEIRINCK, DANIELE. *La pratique archivistique française.* Paris, Archives nationales, 1993, p.507-512.

[Pour le microfilmage] KORMENDY, LAJOS, KEENE JAMES A., POWELL TED F., WEILL, GEORGES. *Manual of Archival Reprography.* Munich, Londres, New-YORK, Paris, ICA Handbooks Series, volume 5. K.G., SAUR, 1989.