

La numérisation des documents anciens

et son application
au Centre Historique des
Archives Nationales (CHAN):
Archim

Définitions

- Numérisation des documents anciens : obtention de bitmaps (images en mode points) au format TIFF, GIF ou autre.

Cette opération nécessite:

- un environnement (des références, une base de données/un corpus)
- un accès aux informations (des indexations, des transcriptions).

■ Document ancien:

D 'après *Le Petit Robert* est ancien « ce qui existe depuis longtemps, qui date d'une époque bien antérieure ».

Il s'agit donc d'un document produit au cours des siècles passés.

Ex: des imprimés mais aussi des manuscrits (registre de baptêmes, actes notariés, manuscrits d'écrivains, cadastres napoléoniens), partitions musicales, recueils de cartes, gravures...

Les aspects politiques : une incitation de l'État

- 1995 mise en œuvre d'un plan ministériel de numérisation portant sur:
 - les fonds documentaires ou iconographiques appartenant à l'État
 - les fonds conservés par les collectivités territoriales

- 1998 : naissance de la notion de « données publiques essentielles »
 - 23 août 1999 : réunion d'un comité stratégique de numérisation chargé de:
 - définir des contenus culturels essentiels
 - présenter des règles de coopération des établissements publics avec leurs partenaires.

Les aspects juridiques

- **Le droit d'auteur** : Les documents anciens, étant tombés dans le domaine public, sont librement exploitables.
- **Le droit de reproduction et de diffusion** relatif à l'exploitation de la base et des données par les internautes.
(Protection des bases de données par la loi n°98 – 536 du 1^{er} juillet 1998.)
 - Pour protéger les images contre une exploitation commerciale, on ne propose sur le web qu'une version dégradée de faible résolution en format JPEG.
 - Important "copillage numérique".

Objectifs de la numérisation

- Sauvegarder les originaux d'une détérioration graduelle liée à une fréquence de consultation importante et en diminuer la communication « numériser pour sauvegarder ».

Débat: le document numérisé peut-t-il se substituer à l'original ?

- Faciliter l'accès au patrimoine archivistique

-Mettre à disposition des publics présents ou distants, amateurs ou chercheurs des images comme substituts des originaux et
-démultiplier les possibilités de consultation en rendant possible l'accès simultané à un même document.

- Multiplication des copies

Offrir la possibilité de reproduire aisément des documents qui n'existent la plupart du temps qu'en un seul exemplaire

- Aider et susciter la recherche

- Alléger la charge du personnel liée à la gestion du fonds physique en supprimant les tâches répétitives.

Les aspects techniques

- La numérisation peut se faire en :
 - noir et blanc : une seule couleur – 1 bit par point image
 - niveaux de gris : 8 bits par point image
 - couleur : 3x 8 bits par points

- 2 sortes de numérisation:
 - Numérisation en mode texte à l'aide d'un logiciel OCR (reconnaissance optique des caractères)
 - Numérisation en mode image: l'exemple d'Archim
<http://www.culture.fr/caran/archim>

L 'exemple de la base image Archim

PRESENTATION D 'ARCHIM

- Conçue en 1995, Archim est une banque d'images numériques des documents d'archives conservés au centre historique des archives nationales (CHAN).
- Elle permet l 'accès sur place et à distance de documents originaux fragiles, précieux ou difficilement communicables et présentant un intérêt national.

Archim contient au 5 mars 2003 environ

- 3536 notices documentaires
- 1182 images numériques.

CONTENU DE LA BANQUE D 'IMAGES

- Des ensembles de pièces constituant des collections connues comme telles.

Exemples: images et moulages de sceaux anciens, manuscrits à peinture, traités originaux d'Ancien Régime

- Des ensembles archivistiques sériels provenant de versements des administrations centrales ou de dépôts.

Exemple: les 62 atlas dits de Trudaine (cartes des routes royales) versés par le ministère des travaux publics

- Des ensembles constituant des dossiers thématiques

Sont en cours de numérisation: des affiches de la 2e guerre mondiale et des album-photos d 'œuvres d 'art.

PRODUCTION DES IMAGES NUMERIQUES

La numérisation est effectuée en interne ou est confiée à des prestataires après appel d 'offre

- Plusieurs catégories d'images sont définies (images de documents aquarellés, de manuscrits, de textes sur parchemins) en fonction des supports, des dimensions et des caractéristiques colorimétriques des documents originaux.
- La numérisation peut être réalisée à partir du document original, d'une photographie ou d'un microforme. La numérisation à partir d'un microfilm est de qualité esthétique moindre, étant donné que le microfilm est composé à partir de nuances de gris.
- Matériel : caméra mobile, scanner à tambour virtuel, scanner à plat et graveur de cd-rom.

LE CAHIER DES CHARGES INDIQUE:

■ Pour l'opération de numérisation

- La nature des documents traités** : photos, imprimés, microformes

- La nature des équipements requis** : modèles et capacités des

scanners utilisés

- Les effectifs sollicités**: nombre d'opérateurs pour le scanning, l'indexation et la préparation des documents

- Les capacités de production** : nombre de documents/jour, durée d'immobilisation du fonds, périodicité des enlèvements, durée de l'opération et date de fin de l'opération

- Les coûts financiers.**

■ Pour chaque image

-Le format : dimensions formats standard (A4, A3) ou non standard

-La couleur : noir et blanc (N&B), niveau de gris ou couleur

-Le volume : nombre d'unités (images, liasses de plans) et nombre d'éléments par unités (pages, feuillets)

-La qualité du fonds : bonne, moyenne, mauvaise qui témoigne de l'état physique des documents et de la qualité des contenus.

■ Pour la gestion électronique des documents (GED)

-**GED d'archivage** : assurer une conservation à long terme des documents

= numérisation à haute résolution (300 à 3000 dpi)

-**GED de consultation en réseau local** : garantir une consultation de qualité tout en gardant des délais d'affichage respectables de l'ordre de quelques secondes

= numérisation à résolution moyenne (300 à 600 dpi soit une définition de 2000 x 3000 pixels)

-**GED de diffusion sur le web** : numérisation à faible résolution (100 dpi soit 200x200 pixels) permettant des temps de transmission acceptables sur le réseau téléphonique. On prévoit généralement 2 ou 3 niveaux d'affichages à l'écran : plein écran, 1/4 d'écran, vignettes.

ORGANISATION DE LA CHAINE DE PRODUCTION

- **Deux équipes** : -photographes et ingénieurs pour la prise de vue et la numérisation
 - documentalistes et conservateurs pour l'analyse documentaire et l'intégration de données.

- **Rédaction d'un protocole de travail** :
 - 1) les documents d'archives à traiter sont décrits avant d'être photographiés et/ou numérisés, afin que le photographe rejette un document non identifié, oublie de photographier certains éléments de la pièce ou que l'ingénieur nomme de façon incorrecte l'image numérique ;
 - 2) les phototypes d'archives à traiter sont décrits avant d'être photographiés.

Protocole de traitement :

- **Sélection des documents** évalués en fonction de leur nature et de leur état
- **Rédaction d'un inventaire** précis des documents à numériser
- **Traitement éventuel des images** : dépoussiérage, réparation, restauration
- **Préparation** : (chaque document est muni d'un code-barres), dégrafage, massicotage s'il est nécessaire d'apprêter le document pour la numérisation
- **Numérisation haute résolution**
- **Génération des différents formats:**
 - TIFF = format de stockage
 - JPEG, GIF = formats de diffusion
 - Plein écran (768x512)
 - Vignettes (128x192)

- **Indexation** = (phase la plus longue) rédaction d 'une notice bibliographique pour chaque document.

Description dans l 'ordre suivant: 1) le fonds
2) la série
3) l 'article.

Une image non décrite=une image perdue.

- **Utilisation de la norme XML: DTD EAD** (Document Type Definition Encoded Archival Description) afin d 'unifier les notices et d 'échanger avec les bases de données utilisant ce même protocole.

- Génération du fichier d'index

- Création de liens

pour reconstituer le document dans son intégralité car chaque entité produit un fichier.

Exemple : **Livres**

Répertoire	AAA
Pages	p0001.TIF
	p0002.TIF
	p0003.TIF

Indexation d 'une image



- Référence: ZZ999999
- Fonds: CP
- Série; versements des ministères ; Travaux publics
- Cote: F/14/*8481
- Type support: feuille
- N° support: 01
- Analyse: Atlas de Trudaine pour la généralité du Hainaut-Cambrésis.
- "Ensemble et sur ce cahier sont les divers chemins qui conduisent de la ville de Valenciennes, capitale du Haynaut autres villes circonvoisines sur le territoire de ce département. Fait par nous inspecteur des ponts et chaussées du haynaut, le 24 mars 1746. Havez. Du chapitre 1 au 12e de l'inventaire Havez. » "Nord, Jemmape. R
- Dates: 1745-01-01
1780-12-31
- Accès: original interdit pour mauvais état; tirage photographique en couleur consultable en salle de lecture du CARAN
- Reproduction: photographie ou impression d'image numérique
- Support: papier à dessin
- Forme: dans un livre relié
- Couleur: N
- Technique: manuscrit; encre; crayon
- Type de document: recueil; atlas; document écrit
- Noms de personnes: TRUDAINE, Daniel-Charles (intendant des finances - directeur des Ponts et Chaussées); HAVEZ (M.) (inspecteur des Ponts et Chaussées)
- Noms de lieux: FRANCE; HAINAUT-CAMBRESIS, GENERALITE DE (France); HAVEZ (M.) (inspecteur des Ponts et Chaussées)
- Mots clés: Ancien Régime français; XVIIIe siècle; généralité (subdivision administrative d'Ancien Régime); aménagement du territoire; voie de communication; route; route royale; directeur des Ponts et Chaussées; intendant des finances; inspecteur des Ponts et Chaussées

- Elaboration d 'une seule grille polyvalente mise en conformité avec la norme ISAD-G (General International Standard Archival Description) pour la production des notices documentaires
- Saisie des métadonnées associées définies par le Dublin Core Metadata Element Set pour l 'indexation des pages du site web
- Contrôle : comptage systématiques des pages de chaque lot et contrôle par rapport aux compteurs indiqués par le logiciel de scanning.
- Vérification des images et de l 'indexation pour s 'assurer de la qualité de la numérisation
- Génération des supports numériques
- CD-Rom " brut de scan " : Production d 'une matrice considérée comme une image de sauvegarde définitive
- CD-Rom de diffusion: Production d 'images " filles" à partir de la matrice

Les avantages de la numérisation

- **Diffusion plus large et moins onéreuse** qui permet des reproductions nombreuses, même en couleur et une diffusion via le serveur Web de collections rares et précieuses à des publics distants avec des temps d'affichages relativement rapides.
- **Aide à la recherche** avec l'accès à un grand nombre d'images en succession rapide. Idéalement chaque chercheur pourra, même par copie, par capture d'images, constituer ses propres dossiers de facsimilés numériques, très maniables et structurés.
- **Netteté des images**
La numérisation permet de très forts agrandissements, supérieurs à ceux d'une photographie classique: visualisation des détails avec zoom.

- **Restauration virtuelle d 'un document** : paramétrage des couleurs, des contrastes qui peut faire ressortir un texte difficile à lire, nettoyage de tâches.

Image originale



Image numérisée



Les limites de la numérisation

- Courte durée de vie: -des supports numériques
-des appareils de lecture.
- Faible espace de stockage
 - Transmission de fichiers trop lourds sur le web impossible.
 - Images à haute définition ne sont encore aisément exploitables qu'en local sur CD-Rom (Or jongler avec plusieurs CD-Roms pour comparer des corpus différents n'est pas plus commode que de jongler avec des volumes in-folio).
- Lisibilité des écritures à l'écran

Ergonomie liée à la dimension limitée de l'écran souvent inférieure à celle du document: défilement horizontal lorsque le document dépasse en largeur rend la lecture difficile et multiplie le risque d'erreurs de repérage visuel.

Bibliographie

Ouvrages

- André Jacques, Chabin Marie-Jeanne, *Les documents anciens*
- Buresi Charlette, Cédelle-Joubert Laure, *Conduire un projet de numérisation*
- Kattnig Cécile, *Gestion et diffusion d'un fonds d'images*
- Jacquesson Alain, Rivier Alexis, *Bibliothèques et documents numériques: concepts, composantes techniques et enjeux*

Sites Internet

- Le médiéviste et l'ordinateur