

enssib

école nationale supérieure des sciences de
l'information et des bibliothèques

Projet Professionnel Personnel de
bibliothécaire : dossier d'aide à la décision

Développement du fonds angevin numérisé de la bibliothèque universitaire d'Angers

Monique ETIENNEY

Sous la direction de Monsieur Christian DUCHARME,
maître de conférences associé à l'ENSSIB, tuteur pédagogique

2000

Conseiller
professionnel

Geneviève DEBLOCK
Conservatrice
adjointe à la direction
Bibliothèque du CNAM

Correspondant
d'établissement

Valérie NEVEU
Conservatrice
adjointe à la direction
SCD d'Angers

Introduction

De nombreuses institutions entreprennent de numériser des fonds patrimoniaux. La numérisation est un outil de mise en valeur des documents primaires, elle permet de fournir des moyens nouveaux à la recherche scientifique, de mieux diffuser les collections auprès du grand public en proposant en ligne des corpus d'ouvrages rares et disséminés. Parallèlement, la numérisation contribue à la sauvegarde et à la conservation des documents patrimoniaux.

C'est dans cette optique que se situe le projet de numérisation du fonds angevin de la bibliothèque universitaire d'Angers.

L'objectif du présent rapport est de faire le point sur la pratique de numérisation actuelle et au regard d'expériences plus avancées sur le sujet, de proposer des solutions d'amélioration de la procédure.

1. Le projet de numérisation du fonds angevin

1. Présentation de la Bibliothèque Universitaire d'Angers

1.1. Petit historique

En 1080, l'université, appelée Ecole d'Angers a déjà une grande renommée à travers le Royaume de France. C'est en 1364 que Charles V crée l'Université – Droit – Médecine et Théologie ; elle attire déjà plus de 1000 étudiants au milieu du XVème siècle lorsque s'ouvre la faculté des Arts. La Révolution abolit l'université française en 1793, mais Angers demeure un carrefour universitaire et intellectuel important. En 1971, avec la fédération des divers collèges, facultés et instituts déjà existants, l'Université est reconstituée sous sa forme moderne. Pluridisciplinaire, elle accueille 17 000 étudiants répartis sur 5 Unités de Formation et de Recherche, 3 Instituts Universitaires Professionnalisés, 1 Institut Universitaire de Technologie.

Le Service Commun de la Documentation comprend 3 sites principaux localisés à Angers et 1 site à Cholet. Celui de Belle-Beille, opérationnel dans ses nouveaux locaux depuis 1989, abrite les collections de sciences, lettres et sciences humaines. Les sites de Montclair et de Saint-Serge sont respectivement dédiés aux disciplines de médecine-pharmacie et de droit, sciences politiques, économie, et gestion. Les collections du site de Cholet sont liées aux 1ers cycles en droit, lettres et économie.

1.2. La politique documentaire

Un premier contrat universitaire (1992-1995) a permis à l'Université d'Angers de se doter d'une bibliothèque informatisée, lieu de développement des technologies numériques de l'information. L'un des objectifs du second contrat (1996-1999) portait sur la poursuite de l'informatisation et la mise en réseau des ressources documentaires numérisées.

Parallèlement au développement des collections liées aux enseignements et aux activités de recherche, la bibliothèque de Belle-Beille constitue des fonds spécialisés et fait de nombreuses acquisitions rétrospectives, (achats ou dons), principalement dans les domaines de la littérature et de l'histoire.

1.3. Le fonds local de la section lettres et sciences humaines

La bibliothèque rassemble en particulier une documentation sur l'Anjou et les livres anciens : elle possède une belle collection de plusieurs milliers de livres du XVIème au XIXème siècles, et même un incunable de 1497.

La bibliothèque renoue ainsi avec l'héritage de l'université de l'Ancien Régime. Elle se définit comme une bibliothèque patrimoniale.

Environ 4000 ouvrages et documents précieux sont conservés dans un magasin particulier, la réserve. Ils bénéficient d'un traitement et d'un conditionnement adaptés. La consultation de ces documents se fait uniquement sur place et sur autorisation.

1.4. Les usagers

Le public de la bibliothèque est essentiellement composé d'étudiants, de chercheurs ou d'enseignants. Régulièrement et de façon croissante, les étudiants et les enseignants en lettres ou en sciences humaines sollicitent des ouvrages du fonds local de la réserve : cartulaires, œuvres littéraires ou documents d'histoire.

Des demandes de prêt entre bibliothèques sont également enregistrées, notamment en provenance de l'étranger. Des chercheurs ou des étudiants de Montréal ont souhaité des ouvrages relatifs au fonds littéraire ou historique local.

1.5. Les projets du prochain plan quadriennal

Pour le prochain contrat quadriennal (2000-2003), les objectifs de la bibliothèque universitaire sont :

- d'améliorer les services et l'accueil offerts aux différents usagers, notamment par la poursuite de l'enrichissement des collections papier ou numériques,
- d'insérer la bibliothèque dans le réseau local et national,
- de mettre l'information scientifique et technique, la plus moderne à la disposition des lecteurs,
- de valoriser les ressources de la bibliothèque.

Dans le cadre du réseau local, la bibliothèque universitaire projette de diffuser elle-même des documents en ligne, avec la numérisation d'ouvrages spécifiques aux collections de la bibliothèque :

- littérature grise, - thèses et mémoires soutenus à l'Université d'Angers, thèses d'exercice de médecine et pharmacie, thèses et mémoires des fonds spécialisés -,
- manuscrits littéraires contemporains, - sous réserve de l'accord des ayants droit -,
- fonds patrimoniaux, (ouvrages essentiellement).

Pour réaliser cette numérisation, des moyens tels que l'acquisition d'un important matériel de stockage des données sont envisagés.

Ainsi, la numérisation des thèses, des fonds spécialisés ou anciens, est un projet qui s'inscrit naturellement dans la préparation du prochain plan quadriennal (2000-2003).

2. Le fonds angevin numérisé de la Bibliothèque universitaire d'Angers

2.1. Description du fonds à numériser : évaluation quantitative et qualitative

Répondant à des demandes précises, la Bibliothèque Universitaire d'Angers numérise et diffuse en ligne des éditions anciennes d'histoire locale (XIXe-début XXe s.) Elle souhaite mettre ainsi à la disposition des chercheurs des documents peu accessibles en dehors des bibliothèques de la région. A terme, environ 1000 documents de la réserve seront concernés.

Ces documents, tombés dans le domaine public peuvent être librement enregistrés, copiés ou imprimés. Ils sont reproduits en mode image, au format PDF¹. Le logiciel permettant de les visionner, Acrobat Reader -, est distribué gratuitement par Adobe. Quelques titres sont actuellement disponibles en ligne.

Afin de consolider cette collection numérique naissante, la numérisation sera étendue à l'ensemble des ouvrages du fonds local, livres de droit et possédant un intérêt patrimonial.

Les objectifs de cette numérisation sont multiples :

améliorer la conservation des documents originaux, parfois fragiles, en créant des supports de substitution ;
rendre accessibles les documents par la diffusion et la consultation à distance ;
valoriser le fonds local et de ce fait participer à la promotion du patrimoine angevin.

Au niveau régional, le projet contribue à la valorisation du patrimoine tout comme le projet de numérisation entrepris sur le site de la bibliothèque municipale de Lisieux².

Afin de constituer une collection cohérente, la sélection des documents s'effectue selon des critères :

thématiques, -littérature et histoire locale -,
éditoriaux, - date et lieu d'édition -,
matériels, - état physique, dimensions -...

2.1.1. Les critères thématiques

Il s'agit de documents des fonds anciens et patrimoniaux sur la littérature et l'histoire locale, sur support papier. Ils appartiennent au fonds de la réserve.

¹ PDF : Portable Document Format, ou format de document transférable, utilisé pour la description de pages en ligne.

² Site de la bibliothèque électronique de Lisieux : <http://www.bmlisieux.com>

Au même titre que les autres ouvrages, ces documents sont catalogués au format UNIMARC et indexés selon RAMEAU dans la base de la bibliothèque. Ils sont accessibles sur le catalogue de la bibliothèque, présent sur le site web de la BU via une passerelle Z 3950 : <http://buweb.univ-angers.fr/>.
Après la mise en ligne des documents, il sera possible, grâce à cette passerelle, de créer des liens hypertextes entre le document numérique et sa notice bibliographique.

2.1.2. Les critères éditoriaux

Afin de distinguer parmi les 4000 documents de la réserve, les documents prioritaires à numériser, une liste des documents de la réserve consultés en 1999 a été établie d'après les statistiques des consultations. A partir de cette liste, on a effectué une recherche croisée sur les titres, les lieux d'édition mentionnant Angers, Anjou, angevin*..., et les dates d'édition (antérieure à 1914). Les résultats ont fait émerger 110 documents. Dans le but de ne retenir que des documents librement exploitables¹, tous les documents dont le décès des auteurs était postérieur à 1930, ou ne pouvait être contrôlé, ont été exclus.

Sur les 110 documents, 29 s'avèrent libres de droits.

Les différentes vérifications ont pu être effectuées entre autre grâce à l'ouvrage "10 siècles de littérature angevine" de Georges Cesbron.

Puis deux autres tris ont été opérés :

l'un sur les ouvrages édités à Angers et publiés avant 1931,

l'autre sur ceux édités hors Angers, publiés avant 1931, avec des titres mentionnant les mots Angers, Anjou, angevin* ...

N'ont été retenus que les documents libres de droit.

2.1.3. Chiffrage des documents à numériser

Estimation du nombre de documents ou nombre d'unités

Après vérification des doublons issus de ces trois requêtes, **174 documents** dont les 24 prioritaires ont été sélectionnés sur un total de 420 documents référencés.

Estimation du nombre d'entités élémentaires : la pagination

L'évaluation du nombre total de pages à numériser porte sur un échantillon de 40 documents : la moyenne est de 315 pages par ouvrage, (page titre, et planches hors texte incluses). Globalement, la numérisation du fonds angevin concerne donc un total estimé de **13000 pages**.

2.1.4. Caractéristiques générales

¹ Voir paragraphe III.A.2b : « Aspects juridiques ».

La moyenne des formats est de 24 cm, (formats 12°, 8°, 4°), ce qui est adapté au format standard A4 et qui permettra une meilleure lisibilité des pages finales à l'écran. Les ouvrages sont imprimés en noir et blanc, la majorité est reliée, certains sont brochés. Le **contenu est essentiellement textuel** : les illustrations sont rares. A l'observation, l'état de conservation de l'ensemble de ces ouvrages est bon, toutefois, les pages des documents les plus anciens sont jaunies.

2.2. Le contexte technique

2.2.1. L'infrastructure informatique

La bibliothèque est informatisée depuis 1992 avec le logiciel Dynix (version V 152). Le catalogue est commun à tous les sites. Des accès internet permettent de consulter à distance les ressources documentaires de la bibliothèque universitaire :

- catalogue de la bibliothèque, (accès au catalogue via internet par l'interface Z39.50),
- bases de données bibliographiques proposées via un lecteur de cédéroms ou via internet,
- près de 1200 revues en texte intégral,
- fonds littéraires spécialisés.

L'architecture générale du réseau local est de type client-serveur en étoile. Les PC sont reliés par des câbles RJ45. Le protocole de transmission est le FTP (File Transfert Protocol) par TCP/IP (Transmission Control Protocol/Internet Protocol).

Le système d'exploitation est :

- windows 95 pour tous les PC,
 - Free BSD pour le serveur web, équivalent à Linux (ou Unix gratuit).
- Matériellement, la connexion sur l'internet est réalisée par fibres optiques.

2.2.2. Le matériel de numérisation existant

Actuellement, la bibliothèque dispose de 2 appareils.

Un scanner à plat : EPSON GT-9500 à une passe RVB, format A4. Ce scanner permet de numériser des documents en couleur. La sauvegarde des fichiers numérisés peut se faire momentanément en local sur disque dur avant transfert sur d'autres supports de stockage. Cependant, la taille des fichiers TIFF¹ obtenus est énorme : on est en TIFF groupe² II ou III, (la compression est médiocre).

Cet appareil convient à la numérisation d'entités élémentaires: il n'est pas adapté pour le fonds angevin.

¹ TIFF = Tagged Image File Format : format propriétaire standard proposé par défaut par les logiciels de numérisation. Voir paragraphe III.A.1.e

² Voir le paragraphe sur la compression : III.A.1.d..

Un photocopieur scanner Xerox Document Centre 230, noir et blanc à une passe, format A4 ou A3. Il peut être paramétré : résolution, taille de l'image... La résolution¹ conseillée est de 300 dpi². Le mode de compression utilisé par le photocopieur scanner est de type groupe IV multi-pages, c'est la compression la plus élevée actuellement en TIFF. Le poids moyen des fichiers images TIFF obtenus oscille entre 30 et 50 Ko. Equipé d'un passe feuilles, il permet de numériser en masse des documents de formats A4. Il a déjà été utilisé pour numériser des thèses ou des ouvrages divers.

2.2.3. Les inconvénients du photocopieur scanner

Si cet appareil convient bien à la numérisation des feuilles volantes, en revanche il est peu adapté à celle des ouvrages reliés : en effet il permet de numériser les ouvrages à plat, mais dans le cas des documents épais, une ouverture trop prononcée risque de détériorer la reliure, de plus la courbure du papier déforme les caractères de l'original.

Une photocopie initiale des pages du document à numériser est nécessaire, car le processus de numérisation est lent, l'appareil est alors monopolisé pour cette activité alors qu'il est l'unique photocopieur de la bibliothèque à usage interne. Comme le processus de numérisation page à page est lent, - manipulation des pages, délai de numérisation pour chaque page-, l'opérateur effectue actuellement une photocopie initiale du document intégral : la liasse de copies est alors prête à être traitée mécaniquement et numérisée en continu. Le gain de temps est significatif, mais la qualité des images obtenues est altérée par cette étape intermédiaire.

2.3. Les étapes de la numérisation actuelle jusqu'à la mise en ligne sur le serveur

La procédure de numérisation se déroule comme suit :

la numérisation du document en TIFF;
la conversion au format PDF et le travail du document sous Adobe Exchange ;
la mise en place sur le serveur avec une page HTML³ de présentation et de liens.

2.3.1. La numérisation

Pour chaque document à numériser :

- on pratique une photocopie préalable de toutes les pages,
- on crée un répertoire sur le photocopieur scanner,
- le document papier photocopié est numérisé,

¹ Voir notion de résolution paragraphe III.A.1c.

² La résolution est mesurée en dpi (dots per inches) ou ppp (points par pouce).

³ HTML = Hypertext mark-up langage. Voir paragraphe III.C.2

- l'image électronique est ensuite convertie dans une série de fichiers TIFF, (un pour chaque page du document papier),
- chaque fichier TIFF est lui-même enregistré dans le répertoire global du document.

Par transfert de type FTP, par TCP/IP, les répertoires sont stockés sur un espace d'archivage temporaire sur un serveur dédié du réseau : l'ordinateur "graveur".

2.3.2. Stockage en vue de la conservation

Cet ordinateur a la possibilité de récupérer les fichiers TIFF et de les graver sur des supports optiques :

- cédéroms, CD-ROM (Compact Disc Read Only Memory, d'une durée de vie de 10 à 20 ans, accessible en lecture uniquement),
- CD-R, (CD réinscriptibles jusqu'à 650 Mo, version effaçable du CD).

Cette étape permet d'assurer une sauvegarde sans encombrer l'espace mémoire du serveur.

2.3.3. Stockage en vue de l'exploitation/consultation

Conversion au format PDF

Avant de pouvoir être mis en ligne, les fichiers numérisés obtenus ou "bruts de scan" sont à retravailler. Le format de restitution choisi est le format PDF : il décrit la disposition de la page, mais pas son contenu. Bien que format propriétaire, il est recommandé dans le dossier technique « A propos de la numérisation », diffusé par le ministère de l'Education Nationale et de l'Enseignement Supérieur, car il permet de conserver l'apparence de la mise en page originale.

A partir de n'importe quel poste du réseau équipé du logiciel Acrobat, il est possible grâce à un mot de passe d'accéder au serveur de stockage, d'importer les fichiers TIFF et de les convertir au format PDF à l'aide du logiciel Acrobat Exchange. L'espace d'archivage temporaire du serveur graveur dédié est alors libéré pour de nouveaux fichiers.

Balilage et remodelage des fichiers sous Adobe Exchange

Cette étape est déjà réalisée dans sa totalité pour les thèses. Par contre, faute de temps, elle n'a pas encore été appliquée aux documents numérisés du fonds angevin.

Pour faciliter l'accès de l'utilisateur au document on pratique :

- un balilage des chapitres,
- une saisie en mode texte de la table des matières ou du sommaire,
- une création de liens entre cette table ou ce sommaire et les différents chapitres.

Parallèlement on fractionne le fichier global du document en sous-parties à l'intention des publics distants qui possèdent des ordinateurs peu puissants ou tributaires du réseau téléphonique : en effet la trop grande taille des fichiers peut ralentir la vitesse de transfert du document, ou empêcher totalement l'accès au document.

Création des pages HTML

La création de liens permettra soit un accès au document total, soit aux chapitres souhaités.

La page de présentation est normalisée : l'opérateur reprend un modèle en changeant seulement les variables. Il suffit de recopier n'importe quel code source d'un ouvrage déjà numérisé.

Ce code comprend :

- les références bibliographiques,
- la cote attribuée par la BU d'Angers,
- le sommaire de l'ouvrage avec les liens vers les différents chapitres,
- des métadonnées selon le Dublin Core¹.

3. Nécessité d'évoluer pour développer la numérisation du fonds angevin

3.1. Les problèmes de stockage générés par la numérisation

Les fichiers PDF sont stockés sur un serveur de la BU dédié aux documents en ligne.

Le poids moyen des fichiers PDF des quelques documents du fonds angevin déjà en ligne est d'environ 2 à 3 Mo pour les sous-parties, et de 8 à 12 Mo pour les fichiers globaux qui correspondent à l'intégralité des documents.

Si l'on extrapole à l'ensemble du futur fonds, on peut objectivement supposer qu'avec une moyenne de 25 à 30 Mo pour un document en ligne, sous-parties et document intégral compris, le volume mémoire nécessaire au fonds angevin oscillera autour de 4500 à 5500 Mo.

Cet espace mémoire sera-t-il compatible avec la capacité totale du serveur, ou faudra-t-il envisager une extension de la mémoire ou éventuellement un nouveau serveur ?

¹ Métadonnées : informations sur les informations. Voir le paragraphe III.C.3 : « utilisation des métadonnées pour la description du contenu ».

3.2. Un processus de numérisation actuellement limité

Comme le montrent les constats précédents, les procédés de numérisation actuellement utilisés au SCD d'Angers, n'autorisent que des travaux ponctuels qui permettent de répondre au coup par coup à des demandes individuelles de documents. Les raisons majeures à cet état de fait sont :

- l'inadaptation du matériel à la numérisation de livres reliés en nombre,
- les problèmes de stockage,
- le manque de disponibilité du personnel.

La numérisation est actuellement prise en charge en plus de ses activités par un conservateur, ou bien réalisée par un moniteur étudiant, - cas de la numérisation des annales d'examens de médecine -.

De plus, la photocopie préalable des documents originaux altère la qualité des documents numérisés obtenus.

Actuellement, pour un ouvrage de taille moyenne, l'ensemble des étapes incluant la numérisation et la mise en ligne, est estimé à une bonne journée pour un utilisateur confirmé, - témoignage du moniteur étudiant chargé de numériser les thèses des fonds spécialisés l'été 1999 -.

La décision de développer la numérisation afin d'offrir des corpus cohérents en ligne doit s'accompagner d'une réflexion sur les moyens dont le SCD doit se doter pour faire aboutir concrètement ce projet.

L'exploitation des diverses expériences rencontrées par des établissements relativement similaires, qui ont pu engager un programme de numérisation sera très précieuse à cette réflexion.

II. Approche de cas extérieurs

Il ne s'agit pas de viser à l'exhaustivité des exemples, mais de simplement cibler ceux rencontrés ou relatés dans la littérature qui peuvent servir de référence par un aspect commun avec le cas d'Angers.

1. Une expérience innovante, le Conservatoire Numérique des Arts et Métiers

1.1. Présentation du CNUM

La bibliothèque, le centre d'histoire des techniques, (CDHT), du conservatoire national des arts et métiers et le centre d'études et de recherches en informatique (CEDRIC) ont entrepris de créer une bibliothèque numérique sur l'histoire des sciences et des techniques : le conservatoire numérique des arts et métiers, (CNUM), en ligne depuis le 1^{er} février 2000. La bibliothèque du CNAM conserve, en France, le fonds documentaire le plus riche d'ouvrages du XIXe siècle dans le domaine de l'industrie et des techniques. La numérisation permet de concilier la mise en valeur de ce fonds et la protection des documents originaux contre les manipulations en complément des autres interventions de conservation. Le CNUM qui intéresse la recherche et l'enseignement en histoire des sciences et des techniques, en épistémologie et en didactique, contribue à renforcer une des missions du CNAM : la diffusion du savoir et la reconnaissance du patrimoine scientifique et technique francophone. Par ailleurs, le CNUM participe au programme de numérisation pour l'enseignement et la recherche (PNER), confié par le ministère de l'éducation nationale, de la recherche et de la technologie, à la Fondation Maison des sciences de l'homme¹.

1.2. Un travail de partenariat entre bibliothécaires, informaticiens et chercheurs

Les choix éditoriaux sont effectués par le CDHT et la bibliothèque qui sélectionnent un ensemble cohérent et représentatif de livres et de périodiques. La bibliothèque² assume ensuite la préparation scientifique et technique des documents. L'équipe « Multimédia et Interaction Homme-machine » du CEDRIC³ apporte son concours dans l'établissement, le stockage et la diffusion des documents numérisés, il joue un rôle de conseiller technique dans le choix des prestataires et il assure la maintenance du site. Le premier projet a bénéficié d'une collaboration de la bibliothèque de l'institut de France, de la bibliothèque interuniversitaire de médecine, de la médiathèque d'histoire des sciences de la cité des sciences et de l'industrie pour l'élaboration d'une bibliographie en complément des documents numérisés. Ce projet constitue une

¹ Site <http://www.msh-paris.fr/pner>

² Responsable du projet CNUM : Madame Geneviève DEBLOCK, conservatrice.

³ Responsable du projet CNUM : Monsieur Pierre Cubaud, maître de conférences.

opération test dans la mise au point d'outils de recherche pour la consultation de textes anciens et pour expérimenter de nouvelles techniques d'interaction dans les bibliothèques numériques.

1.3. Le premier corpus en ligne et son exploitation

Le premier thème retenu concerne l'électricité et le magnétisme au XVIII^e siècle. L'ensemble des 42 textes consultables en ligne représente 22 000 pages et 405 planches numérisées en mode image à partir de microfilms de conservation, en noir et blanc, au format TIFF, de groupe IV, selon une résolution de 400 dpi. Pour ces ouvrages précieux du XVIII^e, le microfilmage préalable, assure une sauvegarde sur un support stable, (actuellement, on attribue aux microformes conditionnées selon les normes adéquates, une durée de survie de 100 à 200 ans). Le CNAM a opté pour des garanties maximales en effectuant un microfilmage en négatif, dédié à la sauvegarde, et un en positif pour la consultation de routine. Pour les ouvrages plus récents, la numérisation est directe. Environ 400 pages de tables des matières et de légendes des illustrations ont été saisies et numérisées en mode texte, ce qui permet des recherches plein-texte. Un glossaire relatif aux différentes orthographes rencontrées pour certains termes permet d'affiner les recherches. La navigation au sein des volumes peut se faire soit séquentiellement à travers les images, soit en précisant une page ou un numéro de planche, soit encore par l'intermédiaire des tables des matières dont les titres de paragraphes donnent accès directement à la page correspondante.

En complément des textes diffusés, une bibliographie et un dépouillement de périodiques et de recueils d'articles, permettent d'élargir la recherche. Livres et périodiques sont localisés dans cinq des principales bibliothèques parisiennes (BnF, CNAM, Bibliothèque de l'Institut, BIUM, Cité des sciences et de l'industrie) et les ouvrages ou périodiques directement consultables soit par le CNUM soit par Gallica, sont signalés comme tels.

1.4. Le contexte informatique

L'architecture du site se veut simple et conviviale : elle reprend dans une large mesure celle mise en place pour la bibliothèque universelle de l'association des bibliophiles universels, - l'ABU¹ -, également hébergée au CEDRIC depuis 1994.

Pour garantir l'accès au CNUM au plus grand nombre d'internautes, quelle que soit la configuration de leur ordinateur, le serveur a été testé sur des configurations de micro-ordinateurs assez anciens (produits aux environs de 1995). Les outils les plus récents du web (Javascript, PDF, etc.), ont été volontairement délaissés. En revanche, la taille et donc le temps de transmission de l'affichage des fichiers issus de la numérisation ont été réduits au maximum, (fichiers GIF² de taille moyenne 62 Ko).

Une aide en ligne destinée au lecteur précise la configuration conseillée, décrit l'organisation du serveur et rappelle les droits d'usage.

La machine serveur est un compatible PC sous Linux. Le serveur Web utilisé est Apache. L'architecture, définie pour supporter jusqu'à un millier de volumes pourrait être dupliquée si ce seuil venait à être dépassé.

¹ <http://cedric.cnam.fr/ABU/index.html>

² GIF : Graphics Interchange Format, format de transfert des graphiques. Ce format d'image permet un bon affichage, mais il est limité à 256 couleurs. Il utilise une technique de compression sans perte. Il convient mieux que le format JPEG pour obtenir des lignes nettes, des images noir et blanc et divers tons de gris.

1.5. Projets...Appel à prestataires

Prenant en compte les enseignements de la première expérience, le CNUM projette actuellement de travailler sur une interaction plus riche :

- hiérarchisation plus fine des tables des matières, (de 2 à 4 niveaux, formes abrégées ou formes développées, avec ajout des résumés analytiques des articles),
- mise en place d'un nouveau moteur de recherche, (restriction de la recherche à un ouvrage ou à un auteur, indexeur, classement des résultats par ordre d'importance),
- téléchargement intégral des volumes dès que l'autorisera le débit de la connexion du CNAM à l'internet,
- expérimentation de nouvelles techniques d'interaction dans les bibliothèques numériques, basées en particulier sur l'image 3D.

L'objectif du CNUM est de se construire par étapes thématiques : un nouveau corpus de numérisation est en projet.

Pour ce corpus, une demande de devis auprès de différents prestataires a été formulée : numérisation en mode image, (tramage en format TIFF de groupe IV, en 400 dpi et éventuellement format JPEG¹ en niveaux de gris pour les pages contenant des planches photographiques), saisie de textes pour les tables des auteurs et des légendes des illustrations.

Les sociétés suivantes ont été contactées : Arkhênum, l'ACRPP², Xerox, Jouve, Flash copie. Au fur et à mesure des entretiens avec les partenaires, le cahier des charges se trouve affiné, notamment pour ce qui concerne la mention des rubriques "qualité³".

1.6. Les enseignements de l'expérience du CNUM

L'expérience du CNAM est intéressante sur différents points qui peuvent être reproduits dans des situations similaires :

- réel travail de partenariat : entre bibliothécaires, informaticiens et chercheurs, partenariat alimenté par des réunions régulières ;
- collaboration avec d'autres institutions, sollicitation d'avis extérieurs auprès d'experts ;
- choix techniques :
numérisation du texte en mode image, saisie des tables en mode texte,
option de sous-traiter la numérisation, (perçue comme plus économique en temps et en budget qu'une numérisation en interne),
souci de convivialité et de simplicité du site, (dans la sélection des tailles, formats, et temps de transmission des fichiers),
évolutivité du contexte informatique ;
- développement des outils de consultation :

¹ JPEG : Joint Photographics Expert Group, Groupe mixte d'experts en photographie. Format de compression à qualité contrôlée. Convient mieux que le format GIF pour les images réelles comme les photographies couleurs.

² ACRPP est l'Association pour la conservation et la reproduction photographique de la presse

³ Voir paragraphe III.B.5 : « étapes d'un projet de numérisation en sous-traitance ».

- recherche plein-texte par les tables, des matières et des illustrations, hiérarchisées par niveaux : formes développées ou abrégées,
- valeur ajoutée apportée par l'élaboration d'un glossaire et d'une bibliographie ;
rappel des droits d'usage, notés dans l'aide en ligne,
mise en place de statistiques de consultation.

Par ailleurs, disposant de moyens limités, - budgets, personnels -, le CNAM a choisi délibérément d'entreprendre la numérisation par tranches successives. Cette progression par palier permet d'évaluer le site au fur et à mesure et d'affiner les outils de navigation proposés.

2. La numérisation à la Bibliothèque Interuniversitaire de Médecine de Paris¹

2.1. L'exposition virtuelle « les frontispices des livres de médecine »

La BIUM, Bibliothèque interuniversitaire de médecine de Paris, présente en ligne une exposition virtuelle, composée essentiellement de documents iconographiques : "100 frontispices de livres de médecine du XVIe siècle au début du XIXe siècle".

L'exposition comprend :

- 1256 pages web,
- 663 images,
- 27 animations pour les pages inter-titres et les génériques.

Les documents originaux ont été préalablement photographiés puis numérisés : scanner de type Plustek OpticPro 9630. Les pages web ont été réalisées avec le logiciel Microsoft Front Page 98 et le code HTML. Elles intègrent quelques fonctions avancées comme les feuilles de style et quelques éléments de Javascript.

La numérisation et la mise en ligne de ces documents de prestige ont été réalisées entièrement à la BIUM, de manière discontinue sur 10 mois, conjointement par :

- un conservateur du service d'histoire de la médecine, (textes et sélection des images),
- un conservateur du service informatique, (conception et réalisation technique, traitement des images),
- deux photographes du service de reproduction.

2.2. Le corpus de textes classiques

¹ Les informations suivantes sont issues d'un entretien avec Monsieur Henri Ferreira-Lopez, conservateur au service de l'histoire de la médecine de la BIUM.

Dès septembre 2000, le site de la BIUM projette de diffuser en ligne 5 grandes éditions précieuses d'Hippocrate, soient environ 11 000 pages. Pour cette bibliothèque numérique, le choix est de numériser les documents en mode image, au format JPEG non compressé, (taille des fichiers de 800 Ko à 1 Mo maximum, alors que les mêmes fichiers numérisés au format TIFF sont environ 10 fois plus gros). La résolution est de 200 dpi.

La numérisation est assurée par un prestataire extérieur, le matériel de numérisation utilisé est le Digibook.

Le coût global est de 55 000.00 FF TTC.

2.3. Le service de reproduction de la BIUM

Le service de reproduction est sollicité par les lecteurs de la BIUM pour produire des photographies, des microformes ou des photocopies numériques à partir de documents textuels ou illustrés. Parmi les demandes de ce type, 1/5 émanent d'usagers extérieurs via les bibliothèques du CNAM, de l'Institut ou de l'Académie de médecine. Une centaine de documents sont ainsi traités mensuellement.

Le service de reproduction est équipé d'un matériel en « leasing¹ »: une photocopieuse numérique minolta PX 3000, (coût mensuel : 4 000.00 FF TTC).

L'inconvénient de ce matériel est qu'il est incapable de numériser en 256 niveaux de gris : les images sont inexploitable sur internet, car la qualité est trop mauvaise : c'est un appareil utile uniquement pour les photocopies des livres anciens, (il préserve les reliures).

Deux photographes sont chargés des travaux de microfilmage et de numérisation, cette équipe de travail est perçue comme bien insuffisante pour répondre aux besoins.

Les documents numérisés sont fournis sous formes de photos ou micro-formes. La facturation actuelle : 2.00 FF l'impression par page numérisée est actuellement encore trop faible pour équilibrer les frais engagés par le leasing.

2.4. L'utilisation d'un appareil photo numérique pour la numérisation

Parallèlement, la BIUM projette de mettre en place un service de numérisation à la carte pour la rentrée de septembre-octobre 2000. L'objectif serait de répondre à des demandes ponctuelles et non à des demandes de numérisations systématiques de corpus rétrospectifs. Actuellement en période d'essai par rapport à ce projet, l'équipe du service de reproduction teste les différentes étapes des procédures de numérisation par un appareil photo numérique récemment acquis. Il s'agit d'un appareil de la marque « Olympus », (référence C3030).

L'équipement est complété par :

¹ Leasing : système de location avec achat à prix préférentiel en option au terme d'une période déterminée, dans le cas présent, il s'agit de la troisième année sur 5 ans.

- une colonne de fixation de l'appareil,
- un plateau compensateur, (utilisé pour rétablir une surface de niveau homogène pour les ouvrages de volume épais),
- un ordinateur-graveur, (pour stocker les données sur cédéroms)
- une imprimante,
- une carte mémoire de 64 Mo, (valeur suffisante pour stocker temporairement jusqu'à 80 pages A4).

Le coût d'investissement n'a concerné que l'appareil et la carte mémoire : le service possédait déjà le reste du matériel, (soient en terme de budget : 10 000.00 et 20 000.00 F TTC).

La capacité estimée de l'appareil est de 50 clichés à l'heure, ce qui améliore les délais de numérisation d'un ouvrage.

Cependant la durée du traitement des images nécessaire, (retouche, cadrage...), réalisé via le logiciel de type Micrografx Picture Publisher, (concurrent équivalent du logiciel Photoshop), n'est pas encore bien quantifiée : +/- 2 minutes par image ?

Cet équipement est bien loin de rivaliser avec un appareil de type Digibook capable de réaliser 400 clichés à l'heure, mais dont le coût reste dissuasif, (environ 800 000.00 F TTC). L'acquisition d'un tel équipement n'est envisageable que pour des budgets conséquents : grosses structures ou associations d'établissements.

Le document numérisé pourra être imprimé sur papier photographique ou ordinaire, ou gravé sur cédérom, (grâce à l'ordinateur-graveur), selon la demande du lecteur. La facturation et le mode de paiement de ces différentes prestations sont encore à l'étude. Il est à noter que le service de production de cédérom est très apprécié par les usagers dans les quelques établissements qui le proposent déjà, comme l'Ecole Française d'Athènes ou certaines bibliothèques anglo-saxonnes. En effet, c'est un support fiable, peu encombrant pour une capacité de stockage des données élevée et téléchargeable gratuitement sur internet : la consultation est rendue possible grâce à un logiciel visualiseur d'images, (de type ACDsys.com¹ par exemple), dont la plupart des ordinateurs de laboratoires universitaires ou de recherche peuvent être équipés.

2.5. L'intérêt de l'exemple de la BIUM

L'exemple de la BIUM est intéressant par les différents aspects qu'elle expérimente :

- la mise en ligne de documents numériques,
- le système du leasing pour l'acquisition à terme d'un matériel au budget conséquent ;
- le choix de sous-traiter la numérisation pour les corpus conséquents ;
- l'acquisition et l'usage d'un nouvel appareil photo numérique, qui, à l'observation, se révèle :
 - o peu coûteux, sous réserve d'un équipement complémentaire déjà acquis,
 - o rapide et simple dans son utilisation en comparaison d'un scanner bureautique,

¹ www.acdsys.com

- lié à un travail de retraitement des images qui, lui, nécessite des compétences spécifiques ;
- la mise en place d'un projet de numérisation à la carte, (avec production de cédéroms ou de documents imprimés).

III. Propositions pour le fonds angevin

1. Les paramètres à déterminer

La numérisation entraîne une suite de procédures : production, stockage, conservation, installation en ligne. Les formats qui conditionnent ces étapes sont à définir selon la qualité des résultats attendue. Les paragraphes suivants font état des diverses recommandations préconisées, (sites ministériels, experts...).

1.1. Revue des aspects techniques

1.1.1. Principe de la numérisation

Schématiquement dit, la numérisation consiste en un processus de conversion d'une information analogique, continue, en une information digitale, numérique, exploitable par l'informatique. Elle peut s'appliquer au texte, au son, à l'image fixe ou animée. Dans le cas d'un document textuel et /ou graphique, le codage produit par la numérisation établit une correspondance entre les caractères linguistiques et /ou graphiques à saisir, et leur enregistrement numérique au niveau de la machine. La représentation numérique d'un texte peut se faire selon deux modalités.

1.1.2. Mode image et mode texte

Le mode image : le texte est représenté sur un mode photographique. Ce type de document est obtenu par numérisation directe du document grâce aux scanners bureautiques, aux scanners de livres, aux appareils photographiques numériques. La surface du document est balayée à l'aide de cellules photoélectriques et analysée selon un damier fictif recouvert de points élémentaires appelés pixels (matérialisés par de très petits carrés blancs, noirs ou de couleur). Le pixel est codé suivant une unité de mesure binaire : le bit, contraction de BInary Digit, dont la valeur est 0 ou 1. Le pixel peut être représenté par un ou plusieurs bits selon le degré de gris ou de couleur désiré : mode bitonal, niveaux de gris, ou couleurs.

Propriétés des variantes du mode image :

	Bitonal	Niveaux de gris	Couleurs
--	---------	-----------------	----------

Codage du pixel	sur 1 bit correspondant au noir ou au blanc	sur plusieurs bits : 8 bits donne 256 gris différents	sur 24 bits en général chaque pixel correspond à 1 des 3 couleurs Rouge Vert Bleu
Volume des fichiers	relativement faible	volume proportionnel au nombre de bits	24 fois plus volumineux qu'en mode bitonal
Caractéristiques générales	idéal pour les documents textuels récents Peu adapté pour préserver l'aspect des documents anciens	présERVE correctement l'aspect des documents anciens	adapté pour les documents photographiques couleurs

La numérisation en mode image produit un fac-similé du document original.

Cette procédure :

est simple à réaliser et d'un coût relativement faible,
mais elle génère cependant des fichiers encombrants,
elle ne permet pas de recherche documentaire à l'intérieur du document.

Le mode texte : chaque caractère a sa représentation unique sous forme numérique. Ce type de document en mode texte ou caractère est obtenu soit par saisie directe par des outils de type traitement de texte, soit par reconnaissance optique de caractères, (OCR), à partir d'un document en mode image. La saisie au clavier est coûteuse en temps et en main-d'œuvre. Délocalisée dans les pays à faible coût de main-d'œuvre la saisie manuelle peut être très compétitive, cependant ce type de sous-traitance soulève des questions déontologiques. Les techniques d'OCR, (taux d'exactitude de 99.5% pour les logiciels haut de gamme), ne sont pas encore en mesure de concurrencer cette saisie manuelle.

Le mode texte :

ne permet pas de conserver la présentation initiale du document,
possède un coût élevé,
permet la recherche plein-texte, et la navigation au sein du document.

1.1.3. Résolution, taille de l'image et compression

La résolution exprime la précision de la numérisation, elle est définie par un nombre de pixels par unité de longueur de la structure à numériser¹. Elle dépend de la qualité du

¹ La résolution est mesurée en dpi (dots per inches) ou ppp (points par pouce).

document initial, (support, caractères, encre, images, couleurs...), et du type de l'usage souhaité : feuilletage rapide et signalétique ou consultation de recherche.

Les préconisations de résolution sont les suivantes :

300 dpi pour une bonne lisibilité, (taux choisi par la BnF pour la base Gallica),
400 dpi pour les caractères originaux, très petits ou de mauvaise qualité,
600 dpi pour l'archivage et notamment pour les manuscrits.

Le poids d'une image est tout d'abord lié à sa taille en pixels, et par conséquent à la résolution qui a été utilisée lors de la numérisation .

La compression

Les documents numérisés en mode image occupent beaucoup de place. On diminue le volume de stockage initial en le compactant. Les images (documents textuels ou illustrations) sont alors codées selon des procédés de compression.

Certaines méthodes de compression sont utilisées selon l'avis du UIT-T (anciennement CCITT, ou Comité Consultatif International Télégraphique et Téléphonique) : CCITT groupe III, ou CCITT groupe IV.

Par exemple, une page de taille A4 contenant du texte numérisé à 200 dpi occupe lorsqu'elle n'est pas compressée un volume de 500 ko. Une fois compressée selon le G4, cette page occupe alors un volume généralement compris entre 20 et 50 ko. De même une page A4 dont le texte est numérisé à 300 dpi passe de 1 Mo à 70 Ko.

Pour les documents textuels en noir et blanc, le mode CCITT groupe IV est recommandé. C'est celui utilisé actuellement à la bibliothèque.

Pour les images fixes noir et blanc, couleur, on utilise la norme JPEG, mais la compression s'accompagne d'une certaine dégradation. Sur le site du CNUM, le format JPEG est envisagé pour la numérisation des planches photographiques.

1.1.4. Les formats de stockage des fichiers

Il est conseillé de compresser les fichiers images pour la diffusion des données, mais à l'inverse d'en garder une version non compressée, un "brut de scan", pour la sauvegarde et la conservation, et pour être en mesure de les convertir ultérieurement en d'autres formats au fur et à mesure de l'évolution de ces derniers avant que le matériel utilisé ne s'avère obsolète.

Les experts de la BnF conseillent :

d'utiliser des supports normalisés, assez répandus, connus, décrits avec une espérance de vie très confortable (cf. norme ISO);
et de stocker en parallèle les données sur un format indépendant de l'application qui l'aura généré, donc de dissocier conservation et exploitation.

Le format graphique de référence en matière de pérennité et de portabilité est le format TIFF¹, conçu au départ pour lutter contre les formats propriétaires, le format TIFF a été conçu dans un esprit d'universalité. Ce format n'engendre pas de perte, mais ne permet qu'une faible compression.

1.1.5. Les formats de diffusion

A partir des fichiers bruts stockés pour la conservation, il est possible d'obtenir des versions différentes en vue d'exploitations distinctes. La diffusion sur l'internet exige de compresser fortement les fichiers dont le poids initial est impropre à la consultation à distance.

Il est conseillé de privilégier les images à faible résolution et les formats de données normalisés sur le web. Les deux formats standards en la matière sont le JPEG pour les photographies numériques et les images couleurs, et le GIF pour les images en niveaux de gris.

L'IRHT² propose de conserver les données non compressées dans un format standard comme le format TIFF et de ne compresser que les fichiers destinés à la consultation dans le format JPEG avec un taux de compression de 10. L'IRHT a choisi 3 niveaux de consultation correspondant à une vignette (8 Ko), une imagerie (30 Ko), une image plein écran (100 Ko).

La bibliothèque municipale de Valenciennes a également retenu le format TIFF pour ses fichiers images qui pèsent en moyenne 25 Mo. Un second fichier est consacré à la diffusion sur écran. Le format JPEG réduit la taille de l'image à 1,75 Mo.

1.1.6. Les supports de conservation et de diffusion

Le stockage des fichiers « bruts de scan » peut se faire sur différents supports. Le support de conservation se caractérise par sa durée de vie et sa pérennité opérationnelle. Le CD-ROM, limité à une capacité de 640 Mo, ou la bande DLT, qui offre une capacité allant jusqu'à 80 Go après compression, sont les plus employés.

Le CD-R reste le support d'échange le plus approprié en attendant la stabilisation du DVD-R qui apportera une capacité de 4 Go.

1.2. Les aspects juridiques

Dans le cadre de la liberté d'expression et sous réserve de respecter les lois sur la propriété littéraire et artistique, ainsi que les lois sur la responsabilité civile, pénale ou professionnelle liée à la diffusion d'informations, les oeuvres de l'esprit peuvent circuler librement sur l'Internet.

¹ TIFF, acronyme anglais de Tagged Image File Format, format de fichier image balisé, rassemble une série de tags (balises), qui indiquent chacun une caractéristique spécifique de l'image : colorimétrie, compression, taille de l'image... Ce format est doué d'une grande flexibilité, aptitude à s'adapter à différentes plates-formes, différents types d'images, tailles et résolutions.

² IRHT : Institut de Recherche et d'Histoire des Textes.

1.2.1. Les lois sur la propriété intellectuelle

Elles créent au bénéfice de l'auteur un monopole d'exploitation qui se décompose en :

un droit moral perpétuel, inaliénable et incessible : il s'agit du droit au nom (droit à la signature de l'oeuvre), du droit de divulgation (seul l'auteur peut décider de porter son oeuvre à la connaissance du public) et du droit au respect de l'oeuvre, à son intégrité ;
- un droit patrimonial qui implique qu'aucune reproduction et aucune représentation de l'oeuvre ne peuvent être faites sans l'accord préalable de l'auteur ou de ses ayants droit, et que l'auteur ou ses ayants droit a droit à une rémunération proportionnelle aux recettes tirées de l'exploitation de son oeuvre.

Durée des droits patrimoniaux

pour l'oeuvre d'un seul auteur : jusqu'à 70 ans après son décès, majoré le cas échéant des années de guerre en France,

pour une oeuvre de collaboration : 70 ans après la mort du dernier auteur,

pour une oeuvre collective : 70 ans après la publication, (la date de publication est déterminée par tout mode de preuve de droit commun, et notamment par le dépôt légal).

Les oeuvres tombées dans le domaine public peuvent être librement exploitées lorsque la durée des droits patrimoniaux a été respectée.

Pour le fonds angevin, on a exclu les oeuvres protégées par le droit d'auteur¹ : le droit de reproduction et de diffusion impliquent une autorisation des titulaires des droits d'exploitation et de représentation, ce qui entraînerait un important travail de recherche inenvisageable actuellement compte-tenu des faibles moyens en temps disponibles pour le projet de numérisation. Dans le cas des auteurs secondaires, préfaciers, illustrateurs, dont la vérification n'aurait pu être réalisée, si d'éventuels ayants droit se faisaient connaître, la solution serait la négociation ou le retrait du document, - à l'identique de ce qui se pratique dans d'autres établissements -.

1.2.2. Rappel des droits d'usage à noter en avertissement

A priori, lorsque l'oeuvre est tombée dans le domaine public, les utilisateurs sont en droit de reproduire l'oeuvre mise en ligne. Cependant, si un utilisateur reproduisait l'intégralité ou une partie substantielle du fonds documentaire lui-même, ou en faisait des extractions répétées, y compris à des fins non commerciales, il serait en infraction avec la nouvelle législation sur les bases de données (loi n° 98-536 du 1^{er} juillet 1998), et passible de sanctions pénales.

Mises à disposition d'un large public, les images numériques ne sont pas à l'abri d'une déformation ou d'une réutilisation induite. Il faut avertir clairement les internautes que les droits d'exploitation de ces fichiers sont contrôlés. La Bibliothèque Universitaire d'Angers est titulaire du droit d'auteur sur la base créée.

¹ Voir paragraphe I.B.1b.

2. Plusieurs options possibles pour développer la numérisation

En préliminaire, il conviendrait de régler les problèmes de stockage générés par la numérisation. A terme, le 1^{er} corpus numérisé en ligne occuperait environ 5500 Mo sur l'espace mémoire du serveur.

Cela peut être résolu par :

une extension de la mémoire,
l'acquisition d'un nouveau serveur.

Le choix d'une numérisation en interne ou confiée à un prestataire sera fortement lié à l'étude :

→ des différents coûts :

- budget d'investissement,
- budget de fonctionnement,
- crédits en personnel,
- crédits en formation si nécessaire,

→ des différentes procédures,

→ des moyens humains à engager,

→ mais aussi de la qualité finale souhaitée pour le fonds à numériser.

Tout d'abord, voici un aperçu du matériel actuellement disponible.

2.1. Les outils de numérisation disponibles sur le marché

Les scanners de livres ou scanners à livre ouvert

Leur vaste surface de numérisation assure le traitement des grands formats. Lors du traitement, le livre est ouvert, texte dirigé vers le haut, le dispositif de numérisation se trouvant au-dessus. Certains appareils intègrent des outils très performants de correction automatique de la rectitude des pages grâce à des algorithmes informatiques : les images obtenues sont totalement plates.

La méthode d'analyse d'images par les scanners de livres peut se faire soit :

par l'intermédiaire d'un dispositif à tête de numérisation fixe, à éclairage fixe sur les côtés, exemple du scanner de livre à tête fixe Zeutschel, (principe apparenté à celui du microfilmage) ;

à l'aide d'un dispositif mobile qui balaie la surface du document comme un scanner bureautique, exemple du scanner de livre i2S à balayage ;

par un dispositif d'ouverture réduite pour reliure fragile, exemple du scanner de livre Zeutschel avec prisme pour document à ouverture réduite.

Les scanners les plus performants qui correspondent davantage aux besoins des bibliothèques sont très coûteux : seules les grandes institutions peuvent s'en équiper.

La numérisation peut s'effectuer également à l'aide d'un caméscope ou d'un appareil-photo numérique. Cependant, un retraitement des images est nécessaire pour obtenir une bonne qualité.

Pour les utilisations plus professionnelles, on peut se diriger vers des caméras numériques tri-CCD, mais avec des budgets élevés.

2.2. L'acquisition de matériel

→ En location

Cette solution engage le SCD dans des coûts de fonctionnement assez conséquents : 4 000.00 TTC mensuels en « leasing » pour un photocopieur numérique Minolta de type PX 3000, auxquels s'ajoutent :

les crédits en personnels responsables du fonctionnement et de la maintenance,
les crédits en formation.

→ Par achat

Pour information, l'acquisition d'un scanner de livre avoisine les :

800 000.00 F TTC, pour un scanner de type i2S,
180 000.00 F TTC, pour un scanner à tête fixe,
1 000 000.00 F TTC, pour un scanner avec prisme.

→ Par achat ou location de matériel en commun.

L'achat ou la location de matériel partagé avec d'autres établissements est une solution à envisager avec prudence. L'intérêt évident est de pouvoir acquérir son propre matériel. Cependant l'utilisation et la maintenance ultérieures doivent être bien définies. Chaque partenaire doit s'engager par rapport à ses "droits et devoirs" : participation aux coûts d'investissement et de fonctionnement, respect des périodes d'attribution du matériel, respect de l'intégrité du matériel...

2.3. Poursuite de la numérisation en interne et acquisition de matériel complémentaire

Acquisition d'un photocopieur

Si la numérisation est poursuivie en interne, selon la procédure décrite au 1^{er} chapitre, on peut améliorer cette procédure par l'acquisition d'un photocopieur dédié à la photocopie préalable des documents avant numérisation : du même type que les photocopieurs qui permettent de poser le livre ouvert à 90° et de reproduire les pages sur une surface comprise jusqu'aux bords de l'appareil.

Cette solution serait certes économique, mais n'apporterait qu'une très faible progression :

libération du photocopieur scanner pour d'autres fonctions, manipulation des documents réduite aux photocopies préalables à la numérisation, d'où une meilleure protection des originaux.

Cependant :

la qualité des images numérisées ne serait pas améliorée : la photocopie initiale des documents originaux entraîne forcément une dégradation, le photocopieur scanner serait toujours monopolisé le temps que les liasses de photocopies soient numérisées, (environ 1 heure pour un document de 300 pages).

Acquisition d'un appareil photo numérique

L'expérience de la BIUM montre qu'une estimation préliminaire des coûts, des besoins en personnels qualifiés, (formation), des besoins en temps, est nécessaire avant de se lancer dans un tel projet.

L'acquisition d'un appareil photo numérique peut être une solution réalisable. En effet, le SCD possède déjà un ordinateur graveur, et une imprimante.

Les investissements sont représentés par les coûts de l'appareil, d'une colonne de réglage, d'une carte mémoire, et d'un plateau compensateur. La fourchette actuelle des prix oscille autour de :

10 000.00 F TTC pour un appareil photo numérique, (d'une capacité d'environ 50 clichés à l'heure, marques de type Olympus ou Nikon),

3 000.00 F TTC pour la colonne,

2 000.00 F TTC pour une carte mémoire de 64 Mo.

Par contre le prix d'un plateau compensateur de qualité reste élevé. Indispensable pour les ouvrages très épais, cet outil peut éventuellement être considéré comme accessoire si l'épaisseur moyenne des ouvrages observée dans le corpus se révèle relativement faible.

La durée du retraitement des images doit être évaluée pour prendre en compte les besoins en personnel et en formation. Ce travail cumulé avec celui de la mise en ligne, est à prendre en compte dans les coûts de maintenance. Il peut facilement faire l'objet d'un emploi à mi-temps.

L'expérience de la numérisation des thèses des fonds spécialisés d'Angers montre que la prise en charge de cette activité par des moniteurs étudiants se révèle trop ponctuelle pour assurer un suivi régulier de qualité, (trop aléatoire dans la durée).

2.4. Sous-traitance de la numérisation

A l'heure actuelle, c'est la solution la plus appropriée aux établissements qui n'ont pas les moyens d'investir dans du matériel de numérisation perfectionné ou qui, bien que déjà propriétaires d'un matériel, ne peuvent mobiliser un personnel suffisant pour entreprendre et achever la numérisation d'un corpus conséquent. Le cas du SCD d'Angers se rapprocherait le plus de ce dernier exemple.

L'hypothèse d'une sous-traitance de la numérisation mérite d'être développée car elle pourrait s'appliquer au cas particulier du projet de développement du fonds numérisé angevin.

La sous-traitance entraîne une série d'étapes successives.

2.5. Les étapes d'un projet de numérisation en sous-traitance

Une grande partie de ces étapes de numérisation sera traitée en interne.

En préambule à la phase de préparation on s'orientera sur :

- la recherche de financements : subventions éventuelles, partenariats,
- la prévision des coûts incluant la préparation des documents, le paiement des droits de reproduction, le transport et l'assurance.

2.5.1. Préparation du train de numérisation

Elle comprend :

la définition du type et du format des documents originaux,
l'analyse de l'état des documents,
la vérification des restrictions légales d'accès,
le traitement pièce par pièce.

Ce dernier comprend le comptage des pages, feuillets dépliant, illustrations, la description très précise des différentes tables, (des matières, des illustrations...), la notation de recommandations particulières.

L'étude du corpus à numériser met parfois en évidence des lacunes. On peut alors procéder à la localisation de ces manques et à la recherche d'autres exemplaires pour compléter. Cette étude doit être également l'occasion de contrôler des notices catalographiques, et de signaler les ajouts ou les modifications éventuels à effectuer, (par exemple, indexation RAMEAU des mots matière).

Ensuite il convient de définir le choix de la meilleure forme de numérisation pour les différentes parties, (modes, formats), de préciser les instructions de traitement des images afin de préparer le cahier des charges.

2.5.2. Elaboration du cahier des charges

Ce cahier des charges comportera en particulier les recommandations correspondant aux spécificités rencontrées lors de la préparation du train de numérisation. Parfois les prestataires proposent des prestations d'assistance à la rédaction du cahier des charges. Le ministère diffuse en ligne des recommandations sous la rubrique "A propos de la

numérisation"¹, il peut fournir sur demande un dossier très détaillé, "Fonds documentaires des bibliothèques universitaires. Cahier des charges de numérisation", rédigé par le cabinet de consultant Parker Williborg. Dans tous les cas, une demande de conseils auprès d'établissements qui ont déjà entrepris une telle expérience est bienvenue, ne serait ce que pour connaître des sociétés de service qui assurent la numérisation.

Il est à noter que la consultation des dossiers relatifs à la numérisation du CNUM a été très précieuse à la rédaction des paragraphes suivants.

2.5.2.1. *Aspects techniques*

Les questions techniques aborderont les points suivants :

- choix du mode de numérisation : mode texte/mode image (noir et blanc, niveaux de gris, couleurs),
- détermination de la taille physique et de la résolution des images (en pixels ou dpi),
- détermination des formats pour l'acquisition numérique (GIF, PDF, TIFF, JPEG), en fonction du poids des fichiers²,
- définition des instructions de traitement des images obtenues (cadrage, retouche...),
- définition du balisage structurel qui déterminera les entrées pour la saisie en mode texte,
- énoncé des règles de nommage des répertoires ou des images :
lien entre le document imprimé original et le fichier électronique correspondant,
lien entre le n° de page de l'ouvrage physique, le type de pagination, (arabe, romain, non paginé...), et son n° d'ordre de numérisation.

2.5.2.2. *Aspects juridiques, contrôle qualité et recommandations diverses*

→ questions juridiques :

respect des obligations légales, (loi informatique et liberté, droits d'auteur...), engagement du prestataire à détruire tous les fichiers images en sa possession sur simple demande du client, ou à ne procéder à d'éventuelles utilisations ultérieures de ces fichiers qu'avec l'accord du client;

→ questions d'ordre déontologique :

signalement précis du lieu de saisie des fichiers textes, (sous-traitance éventuelle en France ou à l'étranger ...);

→ contrôle qualité des fichiers images :

correction des taches,

¹ Site <http://www.sup.adc.education.fr/bib/acti/Num/>

² Voir paragraphe III A.

correction des courbures naturelles du texte dues à la reliure,
correction de la rectitude des pages, redressement sur chaque image,
contrôle de la lisibilité image par image,
contrôle de la continuité bibliographique image par image ;

→ contrôle qualité des fichiers texte :

présentation d'un échantillon de saisie,
mention du taux d'erreur et de la qualité de saisie ;

→ délai de validation : possibilité de modifier les résultats, si détections d'anomalies,
dans un délai raisonnable après la livraison, selon la référence du cahier des charges ;

→ conditionnement, transport: à charge du prestataire ou de la bibliothèque ?

→ conditions et frais d'assurance ;

→ délai d'exécution et date de disponibilité, immédiate ou différée : à préciser ;

→ procédures de contrôle, appliquées selon un plan assurance qualité validé par le client et mentionnant notamment :

la conservation de l'intégrité du document et de sa reliure,
le respect des délais,
le respect des consignes et des engagements de confidentialité.

2.5.3. Le choix du prestataire

Une fois le cahier des charges défini, une demande de devis auprès de différents prestataires peut être formulée. Pour son propre projet, le CNAM a sollicité les sociétés Arkhênum, l'ACRPP¹, Xerox, Jouve, flash copie. L'ADDNB² cite également ICG, La Comète, CICL.

L'analyse comparative des devis permet ensuite de choisir le prestataire. Selon les travaux demandés, la fourchette des prix est très variable en fonction des prestations fournies ou demandées.

Cependant le critère du prix ne doit pas être le seul paramètre déterminant. Il est nécessaire de croiser entre eux les critères techniques, qualitatifs et budgétaires pour effectuer la sélection définitive.

Puis on procède à la constitution et à l'envoi des lots à la numérisation.

2.5.4. Retour de la numérisation et réception des lots en interne

¹ L'ACRPP est l'Association pour la conservation et la reproduction photographique de la presse

² L'ADDNB est l'Association pour le développement des documents numériques en bibliothèques, son site : <http://addnb.org/fr/docs/numproj.htm>

Lorsque la recette du lot numérisé est effectuée, on procède à :

la vérification de l'état matériel des documents,
la relecture,
la réalisation des procédures de contrôle qualité.

Si la saisie des tables des matières en mode texte est effectuée en interne, elle peut être réalisée parallèlement à la phase de numérisation.

Les phases de travail ultérieures concernent les traitements informatiques à réaliser sur le site : production de métadonnées, conception et réalisation des interfaces de consultation, dépôt sur le serveur d'exploitation, mises à jour des pages statiques. Ce travail ne peut être improvisé, s'il n'est pas possible de le confier à des informaticiens, les bibliothécaires responsables doivent suivre une formation de base minimum.

3. Rendre l'information accessible

3.1. Le lien avec la notice catalographique

Le système de gestion de bases de données de la BU d'Angers s'appuie sur le logiciel Dynix lié à une interface web (passerelle Z39.50) qui permet l'accès au catalogue de la bibliothèque via internet.

Comme tous les documents de la bibliothèque, ceux du fonds angevin sont catalogués au format UNIMARC et indexés selon RAMEAU.

A l'occasion de la numérisation, un contrôle des notices est utile : vérification des champs "mots clés", modifications ou ajouts si nécessaire... Pour un document numérisé en ligne, l'accès par la notice est possible grâce à un lien entre la notice bibliographique de ce document et le fichier image correspondant. Ce lien est réalisé par le remplissage du champ 856 du format de catalogue UNIMARC

3.2. L'accès par la table des matières et l'encodage des documents

Concrètement, la création de liens entre les tables des matières et les pages concernées permet à l'internaute un accès direct à la partie du document qui l'intéresse plutôt qu'un passage obligatoire par le feuilletage systématique du document.

Cette étape suppose une saisie manuelle en mode texte des tables des matières existantes ou si elles font défaut, la création de tables. Plusieurs niveaux peuvent être proposés : versions développées ou abrégées, (exemple du site du CNUM). Pouvant être effectuée en interne, cette saisie exige beaucoup de minutie et de temps. Un important

travail de relecture est également nécessaire. Pour le fonds angevin, on peut estimer à 400 pages environ la saisie des tables correspondantes.

En offrant ainsi une valeur ajoutée au document brut, la présentation en ligne des documents s'oriente vers une véritable activité de publication.

Les formats de documents structurés.

Techniquement, cette opération de balisage du texte est réalisable grâce à l'emploi de formats spécifique à l'encodage des documents : les formats de documents structurés.

Les formats de description de pages numériques décrivent la forme sur la page, c'est-à-dire, la disposition d'un document mais pas son contenu ou sa structure. Ils sont utilisés pour une présentation pour une simple lecture d'images. C'est le cas du format PDF. Ces formats sont donc inaptes à décrire le contenu des documents. Pour cela il faut utiliser des formats spécifiques.

Les formats de documents structurés, établis sur des langages structurés de marquage de l'information décrivent les renseignements (contenu) et la structure, et non la disposition. Ce sont des outils et des plates-formes de traitement indépendants qui facilitent l'indexage automatique en décrivant les en-têtes, les chapitres, les paragraphes, les notes de bas de page, etc...

Parmi ces formats, on note les langages suivants.

- Le langage SGML¹ (Standard Generalised Markup Language) : langage standard généralisé de balisage. Un document en SGML comporte 3 éléments : la déclaration, (relatif au milieu de traitement requis), la définition de type de document ou DTD, (description de la structure et du contenu du document), et la suite de documents elle-même.
- Le langage XML² (Extensible Markup Language) : langage de balisage extensible. Il constitue un sous-ensemble réduit et simple du langage SGML.
- Le langage HTML (Hyper Text Markup Language) : langage de balisage hypertexte, directement issu du langage SGML, (HTML est une DTD SGML).

Si les deux premiers langages sont utilisés actuellement de manière exploratoire, par contre HTML est le langage le plus répandu : en effet, dans la majorité des cas, les visualiseurs web utilisent la norme HTML.

3.3. Utilisation des métadonnées pour la description du contenu

Les systèmes d'encapsulation des données nécessitent d'intégrer une DTD. Les métadonnées du Dublin Core, D.C., en sont un exemple.

¹ Pour exemple, le codage en SGML est utilisé sur le site de l'université du Kentucky : <http://www.iath.virginia.edu/blake/public/about/tech/>

² XML est utilisé dans le projet de numérisation de la bibliothèque de l'université de Göttingen : http://www.sub.uni-goettingen.de/java_home-e.htm

Créé sur l'initiative de l'OCLC¹, le D.C., est destiné à la description bibliographique des documents numériques, il peut être perçu comme "le plus petit dénominateur commun des formats de description". L'ensemble des métadonnées du Dublin Core consiste en 15 éléments relatifs à la ressource qui peuvent être classés en trois groupes :

- ceux qui concernent principalement son contenu, (titre, sujet, description, éditeur),
- ceux liés à sa propriété intellectuelle, (auteur, gestion des droits),
- ceux qui correspondent surtout à sa matérialisation, (identifiant et type de la ressource, format).

Le DC est directement implémentable en HTML à travers l'utilisation d'un balisage inséré dans l'en-tête du document HTML.

Au SCD d'Angers, son usage a été testé avec succès pour la mise en ligne de thèses littéraires.

Cependant il faut connaître la mise en garde du ministère² quant à la pérennité du DC : le ministère conseille d'avoir recours à une double description en utilisant un format MARC d'une part et un "header" ou en-tête en D.C. d'autre part et en utilisant une "passerelle" entre ces deux éléments.

3.4. Conception et réalisation des interfaces de consultation

Le graphisme spécifique au fonds sera élaboré dans le respect de la charte graphique du site général de la bibliothèque : choix des icônes relatives aux différentes rubriques, (aide, recherche plein-texte, catalogue).

Un soin particulier sera apporté à l'aide en ligne pour rappeler les droits d'usage et pour décrire :

- la configuration conseillée pour la consultation,
- les modes de navigation :
- par feuilletage,
- par interrogation plein-texte des tables, (mode de recherche simple ou expert).

Lors de l'évaluation du site, l'exploitation régulière des statistiques de consultation permettra d'affiner les outils de navigation en fonction des types de requêtes enregistrées.

4. Moyens humains et compétences requises

Dans la procédure de numérisation, les aspects techniques, bibliothéconomiques et informatiques s'interconnectent. Les différents exemples de projet de numérisation rencontrés ou relatés dans la littérature professionnelle, mettent en évidence la nécessité

¹ OCLC : Online Computer Library Center, (Dublin, Ohio).

² Dossier "A propos de la numérisation" : <http://www.sup.adc.education.fr/bib/acti/Num/notech.htm>

d'une mobilisation de compétences diversifiées. A défaut de pouvoir toutes les cumuler, il est nécessaire d'établir une concertation et un travail de collaboration réels entre les divers agents des services potentiellement impliqués : bibliothécaires, informaticiens. La répartition des tâches déterminées en fonction des compétences de chacun est à énoncer clairement.

L'avis d'experts extérieurs est également instructif.

Ainsi pour les questions spécifiques relatives au droit, la sollicitation de l'avis du service juridique de l'université peut s'avérer précieuse. Par exemple, au CNAM, pour une meilleure compréhension d'une clause de nature juridique, émise par un prestataire potentiel, (droit d'exploitation des données numérisées), le service juridique du CNAM a été consulté.

La mise en ligne des documents et les manipulations qui en résultent :

- balisage des fichiers numérisés,
 - encodage des documents,
 - mise en place des outils de consultation et de navigation,
- sont des activités qui ne s'improvisent pas et qui nécessitent une formation de base et le dégagement d'un créneau horaire spécifique pour l'agent responsable.

En effet, la numérisation d'un corpus conséquent ne peut être considérée comme une activité annexe, exécutée de façon discontinue, sinon la qualité du résultat serait menacée.

D'après les exemples observés, ce travail peut faire l'objet d'un mi-temps.

Pour exemple d'après le bilan de la numérisation sous-traitée du 1^{er} corpus du CNUM, on peut ébaucher le calendrier suivant :

- 6 mois entre la définition du cahier des charges et la numérisation effective,
- 2 mois à temps plein pour la saisie des tables et la relecture confiées à des vacataires,
- 2 mois à temps plein pour l'élaboration du graphisme et de l'interface de consultation, (travaux confiés à une contractuelle titulaire d'un DEA informatique),
- 1 ou 2 mois pour les procédures de contrôle qualité.

Conclusion

Lors de la 3^{ème} conférence internationale sur les bibliothèques numériques, "quels contenus pour les bibliothèques numériques ?", les débats ont fait émerger les diverses compétences liées aux activités inhérentes à la numérisation : elles sont de nature bibliothéconomique, informatique, juridique, et même éditoriale. En effet la numérisation sous-entend de nombreux travaux apparentés en partie aux activités d'édition : choix, tri, préparation, élaboration d'outils d'interrogation et d'appareils bibliographiques, rédaction de notices de présentation, relecture, installation en ligne, maintenance et mises à jour.

La décision de développer la numérisation du fonds angevin devra s'accompagner d'une réflexion sur les différents moyens à mettre en œuvre pour pouvoir réaliser cette opération et par conséquent pour mener à bien toutes les activités citées plus haut.

Bibliographie

1. Bibliographie

BODIN, Bruno, ROUX-FOUILLET, Jean-Paul. *La gestion électronique des documents*. Paris : Dunod, 1992. 190 p.

CESBRON, Georges. *Dix siècles de littérature angevine : des « scriptoria » du XI^e siècle à la récente modernité*. Angers : Presses de l'université d'Angers, 1985. 263 p.

DEMONET, Marie-Luce. « XVI^e siècle : textes électroniques sur la Renaissance », intervention du 9 juin 2000 au cours de la troisième conférence internationale sur les bibliothèques numériques organisée par la Bibliothèque nationale de France et la New York Public Library. *Quels contenus pour les bibliothèques numériques ?*

MINISTÈRE DE L'ÉDUCATION NATIONALE DE LA RECHERCHE ET DE LA TECHNOLOGIE. DIRECTION DE L'ENSEIGNEMENT SUPÉRIEUR, SOUS-DIRECTION DES BIBLIOTHÈQUES ET DE LA DOCUMENTATION. *Droits d'auteur et propriété intellectuelle : numérisation et bibliothèques. Document de synthèse*. Paris : MENRT / Alain Bensoussan, 1999. 14 p.

DUCHARME, Christian. *Du CD-ROM à la numérisation : développer les documents numériques en bibliothèque*. Villeurbanne : ifb, 1997. 172 p. (Boîte à outils).

FONDIN, Hubert. *Le traitement numérique des documents*. Paris : Hermes, 1998. 382 p.

FRESCHARD, Chantal, OKRET, Christine. « La documentation électronique : état des lieux et perspectives de développement dans les bibliothèques des établissements d'enseignement supérieur français », *Bulletin des Bibliothèques de France*, 1999, n° 4, p. 62-65.

GILMONT, Jean-François. *Le livre, du manuscrit à l'ère électronique : note de bibliologie*. 3^e éd. revue et augm.. Liège : Edition du CEFAL, 1998. 154 p. (Bibliothèque élémentaire du bibliothécaire).

HABERT, Benoît, FABRE, Cécile, ISSAC, Fabrice. *De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques*. Paris : Masson / InterEditions, 1998. 320 p. (Informatiques).

JACQUESON, Alain, RIVIER, Alexis. *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*. Paris : Cercle de la librairie, 1999. 377 p. (Bibliothèques).

"Les bibliothèques numériques". DUPOIRIER, Gérard, coordonnateur. *Document numérique*, vol. 2, n°3-4, 1998. Paris : Hermes, 1999. 299 p.

MADDALONI, Marie-Claude. « L'information numérique dans les BM. Etat des lieux », *Bulletin des Bibliothèques de France*, 1999, n° 4, p. 66-69.

Nouvelles technologies dans les bibliothèques. Sous la direction de Michèle ROUHET. Paris : Cercle de la librairie, 1996. 386 p. (Bibliothèques).

PRAX, Jean-Yves. *La gestion électronique documentaire : manager les flux d'information dans l'entreprise*. 2^e éd.. Paris : Masson / InterEditions, 1998. XV-208 p. (Informatiques).

VIGOUREUX Michel. *Exploitation d'images numérisées en vue d'une création de banque d'images dans le cadre du service de reproduction de la BnF*. [PPP]. Villeurbanne : Ifb, 1998. 48 p.

VOS, Andreas. *Dictionnaire de l'informatique et de l'internet*. Paris : Micro Application, 1998. 731 p.

2. Sites

Enquêtes, recensements sur les projets de numérisation

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPÉRIEUR. (Page consultée le 08.03.00). Enquête sur les ressources électroniques acquises par les SCD, SICD et les bibliothèques de grands établissements, [en ligne] < URL : <http://www.abes.fr/enquete.htm> >

FRANCE. MINISTÈRE DE L'EDUCATION NATIONALE DE LA RECHERCHE ET DE LA TECHNOLOGIE. DIRECTION DE L'ENSEIGNEMENT SUPÉRIEUR. SOUS DIRECTION DES BIBLIOTHÈQUES ET DE LA DOCUMENTATION. (Page consultée le 06.04.00). Enquête sur les projets de numérisation, [en ligne].<URL : <http://www.sup.adc.education.fr/bib/acti/Num/Enq/num1.htm>>

FRANCE. MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION. (Page consultée le 06.03.00). Bibliothèques numériques, [en ligne] < URL : <http://www.culture.fr/culture/int/index.html> >

MAISON DES SCIENCES DE L'HOMME. (Page consultée le 06.03.00). Les usages et les besoins des documents numériques dans l'enseignement supérieur et la recherche. [en ligne] < URL : <http://www.pner.org/html/activduprog/doc2bl.asp>>

Aspects juridiques

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE. Page consultée le 06.04.00). Quelques généralités sur le droit d'auteur, [en ligne] < URL : <http://www.sg.cnrs.fr> >

COMITÉ RÉSEAU DES UNIVERSITÉS. (Page consultée le 15.03.00). Questions sur les droits d'auteur, [en ligne] < URL : <http://www.cru.fr/droit-deonto/droit/autres.elements/index.htm>>

Recommandations ministérielles - conduite de projets

ASSOCIATION POUR LE DÉVELOPPEMENT DES DOCUMENTS NUMÉRIQUES EN BIBLIOTHÈQUES. (Page consultée le 05.04.00). Les étapes d'un projet de numérisation en bibliothèque, [en ligne] < URL : <http://addnb.org/fr/docs/numproj.htm>>

FRANCE. MINISTÈRE DE L'EDUCATION NATIONALE DE LA RECHERCHE ET DE LA TECHNOLOGIE. DIRECTION DE L'ENSEIGNEMENT SUPÉRIEUR. SOUS DIRECTION DES BIBLIOTHÈQUES ET DE LA DOCUMENTATION. (Page consultée le 06.03.00). A propos de la numérisation, [en ligne] < URL : <http://www.sup.adc.education.fr/bib/acti/Num/num2.htm>>

FRANCE. MINISTÈRE DE LA CULTURE. DIRECTION DU LIVRE ET DE LA LECTURE. (Page consultée le 05.04.00). Numérisation des bibliothèques, [en ligne] < URL : <http://www.culture.fr/culture/mrt/numerisation/fr/dll/index.htm>>

SETA, Frédérique. (Page consultée le 17.03.00). Le rôle de la numérisation dans la mise en valeur des fonds patrimoniaux : l'exemple de la bibliothèque interuniversitaire Cujas, [en ligne] < URL : <http://www.enssib.fr/bibliotheque/documents/dcb/seta.pdf>>

Aspects techniques - Standard-Normes

BOUDRY, C. (Page consultée le 17.03.00). Références et sites à consulter, [en ligne] < URL : http://www.ccr.jussieu.fr/urfist/image_numerique/home_image.htm>

CHARTRON, Ghislaine. (Page consultée le 14.06.00). Standard-Normes-Documents numériques, [en ligne] < URL : <http://www.ccr.jussieu.fr/urfist/presse/standard/coursintro.htm>>

CUBAUD, Pierre, TOPOL, Alexandre. (Page consultée le 14.06.00). A WWW-based digital library for antiquarian collection, [en ligne] < URL : <http://jasmin.cnam.fr> :8081/RR9909>

THIRION, Benoît. (Page consultée le 20.04.00). Dublin Core , qu'est-ce-que c'est ?, où en est-on ?
[en ligne] < URL : <http://www.ccr.jussieu.fr/urfist/presse/standard/dcurfist1/sld010.htm> >

Corpus numériques en ligne

ASSOCIATION DES BIBLIOPHILES UNIVERSELS. (Page consultée le 09.03.00).
[en ligne] < URL : <http://cedric.cnam.fr/ABU/index.html> >

BIBLIOTHÈQUE ÉLECTRONIQUE DE LISIEUX.(Page consultée le 03.04.00).
[en ligne] < URL : <http://www.bmlisieux.com/>>

BIBLIOTHÈQUE DE L'UNIVERSITÉ DE GOETTINGEN. (Page consultée le 06.06.00).
[en ligne] < URL : <http://www.sub.uni-goettingen.de/ssgfi/>>

BIBLIOTHÈQUE INTERUNIVERSITAIRE DE MÉDECINE. (Page consultée le 07.04.00). Exposition virtuelle, [en ligne] < URL : <http://www.bium.univ-paris5.fr/expo/> >

BIBLIOTHÈQUE NATIONALE DE FRANCE. (Page consultée le 07.04.00). Gallica 2000,
[en ligne] < URL : <http://gallica.BnF.fr/>>

CONSERVATOIRE NUMÉRIQUE DES ARTS ET MÉTIERS. (Page consultée le 09.03.00).
[en ligne] < URL : <http://cnum.cnam.fr> >

Tables des matières

INTRODUCTION	2
I. LE PROJET DE NUMÉRISATION DU FONDS ANGEVIN	3
1. Présentation de la Bibliothèque Universitaire d'Angers	3
1.1. Petit historique.....	3
1.2. La politique documentaire.....	3
1.3. Le fonds local de la section lettres et sciences humaines.....	4
1.4. Les usagers.....	4
1.5. Les projets du prochain plan quadriennal.....	4
2. Le fonds angevin numérisé de la Bibliothèque universitaire d'Angers	6
2.1. Description du fonds à numériser : évaluation quantitative et qualitative	6
2.1.1. Les critères thématiques.....	6
2.1.2. Les critères éditoriaux.....	7
2.1.3. Chiffrage des documents à numériser.....	7
2.1.4. Caractéristiques générales.....	7
2.2. Le contexte technique.....	8
2.2.1. L'infrastructure informatique.....	8
2.2.2. Le matériel de numérisation existant.....	8
2.2.3. Les inconvénients du photocopieur scanner.....	9
2.3. Les étapes de la numérisation actuelle jusqu'à la mise en ligne sur le serveur	9
2.3.1. La numérisation.....	9
2.3.2. Stockage en vue de la conservation.....	10
2.3.3. Stockage en vue de l'exploitation/consultation.....	10
3. Nécessité d'évoluer pour développer la numérisation du fonds angevin	11
3.1. Les problèmes de stockage générés par la numérisation.....	11
3.2. Un processus de numérisation actuellement limité.....	12
II. APPROCHE DE CAS EXTÉRIEURS	13
1. Une expérience innovante, le Conservatoire Numérique des Arts et Métiers	13
1.1. Présentation du CNUM.....	13
1.2. Un travail de partenariat entre bibliothécaires, informaticiens et chercheurs	13
1.3. Le premier corpus en ligne et son exploitation.....	14
1.4. Le contexte informatique.....	14
1.5. Projets...Appel à prestataires.....	15
1.6. Les enseignements de l'expérience du CNUM.....	15
2. La numérisation à la Bibliothèque Interuniversitaire de Médecine de Paris 16	
2.1. L'exposition virtuelle « les frontispices des livres de médecine ».....	16
2.2. Le corpus de textes classiques.....	16
2.3. Le service de reproduction de la BIUM.....	17

2.4.	L'utilisation d'un appareil photo numérique pour la numérisation	17
2.5.	L'intérêt de l'exemple de la BIUM	18
III.	PROPOSITIONS POUR LE FONDS ANGEVIN	19
1.	Les paramètres à déterminer	19
1.1.	Revue des aspects techniques.....	19
1.1.1.	Principe de la numérisation.....	19
1.1.2.	Mode image et mode texte	19
1.1.3.	Résolution, taille de l'image et compression	20
1.1.4.	Les formats de stockage des fichiers.....	21
1.1.5.	Les formats de diffusion.....	22
1.1.6.	Les supports de conservation et de diffusion	22
1.2.	Les aspects juridiques.....	22
1.2.1.	Les lois sur la propriété intellectuelle.....	23
1.2.2.	Rappel des droits d'usage à noter en avertissement.....	23
2.	Plusieurs options possibles pour développer la numérisation	24
2.1.	Les outils de numérisation disponibles sur le marché	24
2.2.	L'acquisition de matériel.....	25
2.3.	Poursuite de la numérisation en interne et acquisition de matériel complémentaire.....	25
2.4.	Sous-traitance de la numérisation	26
2.5.	Les étapes d'un projet de numérisation en sous-traitance.....	27
2.5.1.	Préparation du train de numérisation	27
2.5.2.	Elaboration du cahier des charges.....	27
2.5.2.1.	<i>Aspects techniques.....</i>	<i>28</i>
2.5.2.2.	<i>Aspects juridiques, contrôle qualité et recommandations diverses</i>	<i>28</i>
2.5.3.	Le choix du prestataire	29
2.5.4.	Retour de la numérisation et réception des lots en interne	29
3.	Rendre l'information accessible	30
3.1.	Le lien avec la notice catalographique	30
3.2.	L'accès par la table des matières et l'encodage des documents.....	30
3.3.	Utilisation des métadonnées pour la description du contenu.....	31
3.4.	Conception et réalisation des interfaces de consultation.....	32
4.	Moyens humains et compétences requises	32
	CONCLUSION.....	34
	BIBLIOGRAPHIE	35
	TABLES DES MATIÈRES	38