

Numérisation et valorisation des archives historiques de Châtillon-sur-Chalaronne dans l'Ain

Vincent BAAS

Sous la direction de Frank Le Bourgeois¹ et Isabelle TURCAN²

1 Maître de conférences, LIRIS, INSA

2 Professeur à la Faculté des Lettres de l'Université Nancy 2, laboratoire ATILF, UMR 7118, Nancy 2 : responsable de l'équipe « Métalexigraphie historique » et du programme de numérisation du fonds ancien

Remerciements :

Nous remercions tout d'abord Monsieur Frank Le Bourgeois, tuteur de notre stage, de nous avoir proposé de participer au projet de numérisation des archives de Châtillon.

Nous remercions Monsieur Noël Ravassard, maire de Châtillon-sur-Chalaronne, pour l'intérêt qu'il porte au projet de numérisation et les facilités mises en place dans l'accès et la consultation des archives. De même, nous remercions les membres du conseil municipal et le personnel de la mairie qui ont accompagné notre intervention d'échanges fructueux et amicaux.

Nous remercions Monsieur Jean-Louis Maret, président de l'association Saint-Guignefort, et tous les membres de l'association, pour leur engagement passionné dans la valorisation du patrimoine, l'impulsion du projet de numérisation et le travail accompli. Grâce à la prise en charge de nos déplacements depuis Lyon, ils nous ont permis une participation plus concrète et donc plus efficace au travail de numérisation.

Nous devons souligner le plaisir que nous avons eu à travailler avec Caroline Saret, étudiante en histoire de l'art médiéval : nous avons pu apprécier son travail de numérisation et transcription des manuscrits.

Nous remercions particulièrement Isabelle Turcan, sans laquelle notre intervention n'aurait pu avoir lieu dans les mêmes conditions. Sa disponibilité, sa rigueur intellectuelle, ses remarques toujours constructives, nous ont permis de donner une autre ambition à ce stage. Nous lui sommes reconnaissants d'avoir effectué une relecture attentive et exigeante de ce rapport.

Résumé :

Les archives historiques de Châtillon-sur-Chalaronne constituent un fonds important porteur d'une valeur historique et linguistique. Le projet de valorisation de ces archives passe par une consultation numérique reposant sur différentes étapes : la numérisation, le traitement, l'indexation puis la mise en place d'une stratégie de diffusion auprès du grand public ou de la recherche.

Descripteurs :

A/ liste normalisée¹

Analyse et indexation des documents, Archives municipales, édition électronique, numérisation.

B/ descripteurs adaptés au corpus de Châtillon²

Documents manuscrits anciens, valorisation documentaire.

Abstract :

The historic archives of Châtillon-sur-Chalaronne constitute a mattering resource possessing a historic and linguistic value. The project of valorization of these archives passes by a numerical consultation resting on various stages: digitalization, treatment, the indexing then the installation of a strategy of diffusion near general public or research.

Keywords :

A/ list standardized LCSH

Abstracting and indexing, Digitalization, Electronic publishing, Municipal government.

B/ descriptors adapted to the corpus of Châtillon

Old handwritten documents, documentary valorization.

Toute reproduction sans accord express de l'auteur à des fins autres que strictement personnelles est prohibée.

¹ Les descripteurs sont issus de la liste d'autorité Rameau de la Bibliothèque Nationale de France. Les descripteurs en langue anglaise sont les équivalents LCSH, Library of Congress Subject Headings, vedettes-matière de la bibliothèque du congrès américain.

² Proposés par Isabelle Turcan.

Sommaire

INTRODUCTION.....	7
LA PRÉSENTATION DES ARCHIVES HISTORIQUES	9
1. LE CONTENU ET L'ÉTAT DE LA COLLECTION	9
2. L'INVENTAIRE ET LE CLASSEMENT DES ARCHIVES HISTORIQUES.....	10
3. LES ARCHIVES HISTORIQUES DE CHÂTILLON : ENJEU PATRIMONIAL, OUTIL DE CONNAISSANCE LINGUISTIQUE ET DÉVELOPPEMENT D'APPLICATIONS	12
4. L'ACCÈS AU CONTENU DES ARCHIVES HISTORIQUES	13
4.1. <i>Les archives manuscrites</i>	13
4.2. <i>La transcription diplomatique</i>	14
4.3. <i>La traduction</i>	14
LA NUMÉRISATION DES ARCHIVES	15
1. LES ACTEURS DU PROJET DE NUMÉRISATION	15
1.1. <i>L'association culturelle Saint-Guignefort</i>	15
1.2. <i>L'intervention de Caroline SARRET, étudiante en histoire de l'art médiéval</i>	15
1.3. <i>Notre intervention</i>	16
2. LES DISPOSITIFS DE NUMÉRISATION	17
2.1. <i>Le plan de numérisation</i>	17
2.2. <i>Les paramètres de numérisation</i>	18
2.2.1 Le format de fichier JPEG	18
2.2.2 Une résolution de 400 dpi.....	19
2.2.3 Une palette de couleurs de 24 bits en mode RVB.....	19
2.2.4 La vitesse de numérisation	19
2.2.5 Autres paramètres de numérisation utilisés	20
2.2.6 Récapitulatif des formats utilisés	20
2.3. <i>La manipulation des archives</i>	20
2.4. <i>Que faut-il numériser dans un document d'archive ?</i>	21
2.4.1 Les éléments matériels	21

2.4.2	Les éléments non-textuels	23
2.4.3	Les éléments ajoutés.....	23
2.5.	<i>Comment nommer les fichiers d'archives ?</i>	25
2.6.	<i>Les fichiers renommés automatiquement</i>	26
2.7.	<i>Le stockage des fichiers d'archives</i>	27
2.8.	<i>Numérisation physique et première indexation</i>	29
3.	LE SCANNER, LE PILOTE ET L'INTERFACE DE NUMÉRISATION	31
3.1.	<i>Le scanner, un Epson perfection 1200S</i>	31
3.2.	<i>Le pilote du scanner, Epson Twain</i>	31
3.3.	<i>L'interface de numérisation</i>	32
4.	L'ÉVALUATION DES TEMPS DE NUMÉRISATION	34
	LE TRAITEMENT NUMÉRIQUE DES ARCHIVES	35
1.	POSSIBILITÉS TECHNIQUES ET CONDITIONS D'UTILISATION	35
2.	LE LOGICIEL DE TRAITEMENT D'IMAGES BOOK RESTORER.....	36
2.1.	<i>Correction des conditions de numérisation</i>	36
2.1.1	La correction géométrique et correction d'éclairage	36
2.1.2	Le redressement des images	37
2.2.	<i>Transformation de l'image et conversion de couleurs</i>	37
3.	ANALYSE DE LA STRUCTURE PHYSIQUE DES FOLIOS.....	38
	L'INDEXATION DES ARCHIVES.....	39
1.	PROBLÉMATIQUES RENCONTRÉES DANS LES PROJETS DE NUMÉRISATION DES MANUSCRITS ANCIENS	39
2.	L'EAD, UNE DTD POUR LA DESCRIPTION DES FONDS D'ARCHIVES	41
3.	ORGANISATION DES ARCHIVES ET MÉTADONNÉES	43
3.1.	<i>Métadonnées et la notion de mot-clé métalinguistique</i>	44
3.2.	<i>D'une indexation de base au contenu</i>	44
3.3.	<i>La pièce d'archive comme unité de base d'indexation</i>	45
4.	DE L'INTÉRÊT D'UN THÉSAURUS.....	48
5.	DES INFORMATIONS CONTEXTUELLES COMPLÉMENTAIRES COMME PREMIÈRE PHASE DE VALORISATION.....	49
5.1.	<i>Une chronologie universelle et locale</i>	49

5.2.	<i>Des lexiques au Dictionnaire Universel [...] de Trévoux(1704-1771)</i>	50
5.2.1	Des manuscrits aux lexiques.....	50
5.2.2	Le corpus des dictionnaires anciens.....	51
5.2.3	Des manuscrits au <i>Dictionnaire de Trévoux</i>	52
5.2.3.1	Un dictionnaire de la Dombes qui s'intéresse à la Bresse.....	52
5.2.3.2	Un dictionnaire du XVIII ^e siècle qui s'intéresse au droit féodal et au droit coutumier	53
5.3.	<i>Représentation du modèle de valorisation des archives à l'aide du Dictionnaire de Trévoux</i>	54
	L'INTERFACE DE CONSULTATION DES ARCHIVES.....	57
1.	LA CRÉATION D'UN CÉDÉROM DE DÉMONSTRATION	57
2.	LES DÉVELOPPEMENTS ENVISAGÉS	58
	CONCLUSION.....	60
	BIBLIOGRAPHIE RÉCAPITULATIVE.....	61

Introduction

Les archives historiques de la ville de Châtillon-sur-Chalaronne constituent une collection patrimoniale dont les premiers folios remontent au XIV^e siècle. Ces archives sont le témoignage de la vie quotidienne des habitants de l'agglomération, de l'évolution du pouvoir féodal, de la formation d'une identité au travers de changements de tutelles et découpages territoriaux. Elles nous renseignent sur les évolutions de la langue écrite et parlée à Châtillon au cours des différentes périodes, des mélanges de parlers régionaux, de la constitution progressive d'une langue française de référence dès le XV^e siècle.

Notre intervention se situe dans le cadre d'un projet de numérisation déjà en cours de ces archives historiques. Elle répond à plusieurs objectifs s'inscrivant dans l'ensemble des étapes d'une chaîne de numérisation :

- les modalités de numérisation
- le traitement automatisé des images
- les choix d'indexation
- la valorisation d'une première partie de la collection d'archives et la création d'un cédérom

En dehors de ces aspects, le contenu de ce mémoire a une visée pratique. Il doit permettre aux acteurs du projet de numérisation de faire un premier bilan, d'envisager une normalisation des procédures et de permettre un déploiement de notre méthodologie à plus grande échelle, sur l'ensemble de la collection d'archives³.

Pour mener à bien ces étapes, le stage a été placé sous la double responsabilité scientifique de Frank Le Bourgeois (LIRIS, INSA, Lyon 1) pour ce qui concerne le traitement automatisable des documents numérisés et d'Isabelle Turcan pour ce qui concerne les modalités de transmission et de valorisation des contenus des documents numérisés, via différentes procédures d'indexation.

³ Ce n'est pas le lieu de développer ici une réflexion sur la notion de *collection* ; cf. I.Turcan, *Numérisation et valorisation des collections d'ouvrages anciens du laboratoire ATILF, CNRS, UMR 7118, Nancy 2* in *Les institutions culturelles et la numérique/ Cultural institutions and digital technology*, Ecole du Louvre ICHIM Paris 2003 (actes sur Cdrom et sur www.ichim.org).

Pour mener à bien la rédaction de ce rapport nos sources ont été de quatre ordres :

- sources bibliographiques imprimées traditionnelles
- sources éditées sur support électronique
 - o sur cdrom
 - o sur le Web
- sources dactylographiées
- colloque de la *semaine du document numérique*⁴ à la Rochelle et entretiens particuliers.

⁴ La *semaine du document numérique* s'est déroulé du 21 au 25 juin 2004. Dans le cadre de ce colloque, nous avons assisté aux tutoriaux, sessions, tables rondes consacrés au thème *Numérisation et patrimoine*.

La présentation des archives historiques

1. Le contenu et l'état de la collection

Les archives historiques couvrent une période qui s'étend du XIV^e au XVIII^e siècle. La collection est évaluée à 13 000 folios manuscrits, recto et verso, ce qui correspond environ à 27 000 images numériques à réaliser. Ces manuscrits ont été répartis en deux ensembles concernant deux périodes, la période ancienne de 1273 à 1790 et la période révolutionnaire de 1790 à 1804⁵. Ces archives rendent compte de l'évolution de la ville dans son fonctionnement interne (administration communale, organisation du travail, du commerce), dans ses relations avec les autres entités territoriales (la Savoie, la France, la Bresse, la Dombes) et fournissent des informations nombreuses sur l'histoire de la langue et des sociétés qui l'ont fait vivre.

Les manuscrits sur lesquels nous avons travaillé reposent essentiellement sur l'activité des conseils de syndics et plus précisément les comptes de syndics, instance communale qui a vu le jour à Châtillon dès la moitié du XIV^e siècle constituée d'agents exécutifs des décisions de l'assemblée générale des habitants et ayant pour mission la comptabilité et la gestion des finances de la ville⁶.

Ces comptes de syndics sont organisés de façon uniforme dans leur structure logique (préambule, recettes, dépenses, conclusion) et spatiale (l'apurement du compte est aligné sur la droite du manuscrit). Cette organisation répétitive a son importance dans la mesure où elle permet d'envisager un repérage automatisé des

⁵ Cf. l'inventaire établi par J.C Mercier, archiviste au CDG de l'Ain en mai 2003, consultable en mairie à Châtillon-sur-Chalaronne.

⁶ La mise en place d'un conseil de syndics s'est faite à Châtillon-sur-Chalaronne quatre-vingt ans après l'instauration d'une charte de franchise instituant la ville en « communauté » et acte d'indépendance par rapport au pouvoir central et féodal. (cf. MOREL M. Octave. *Une petite ville forte en Bresse sous la première domination savoyarde; la vie à Châtillon-en-Dombes d'après les comptes de syndics 1375-1500*. Bourg en Bresse : Imprimerie « Courrier de l'Ain », 1925, pp. 50-64.

différents éléments du manuscrit, ce qui pourra en faciliter le traitement et, éventuellement, la transcription systématique.

Dans l'ensemble, la collection est en bon état de conservation. Cependant nous avons constaté que certains manuscrits étaient déchirés, troués ou étaient endommagés soit par le feu, soit par des taches d'humidité. Les archives historiques sont constituées quasi exclusivement de documents textuels calligraphiés de différentes mains. Nous n'avons repéré dans le corpus sur lequel nous avons travaillé aucun autre élément d'illustration (dessins, enluminures) ou d'habillage calligraphié du texte (lettrines).

2. L'inventaire et le classement des archives historiques

L'inventaire des archives a été réalisé dans les années 1920-1930 par M. Octave Morel⁷ archiviste du département de l'Ain. Depuis 1998 les manuscrits de Châtillon-sur-Chalaronne font l'objet d'une mission auprès des archives départementales de l'Ain et le plan de classement des archives est conforme aux normes archivistiques en vigueur pour les archives communales avec une organisation thématique puis chronologique. L'organisation se présente de la façon suivante :

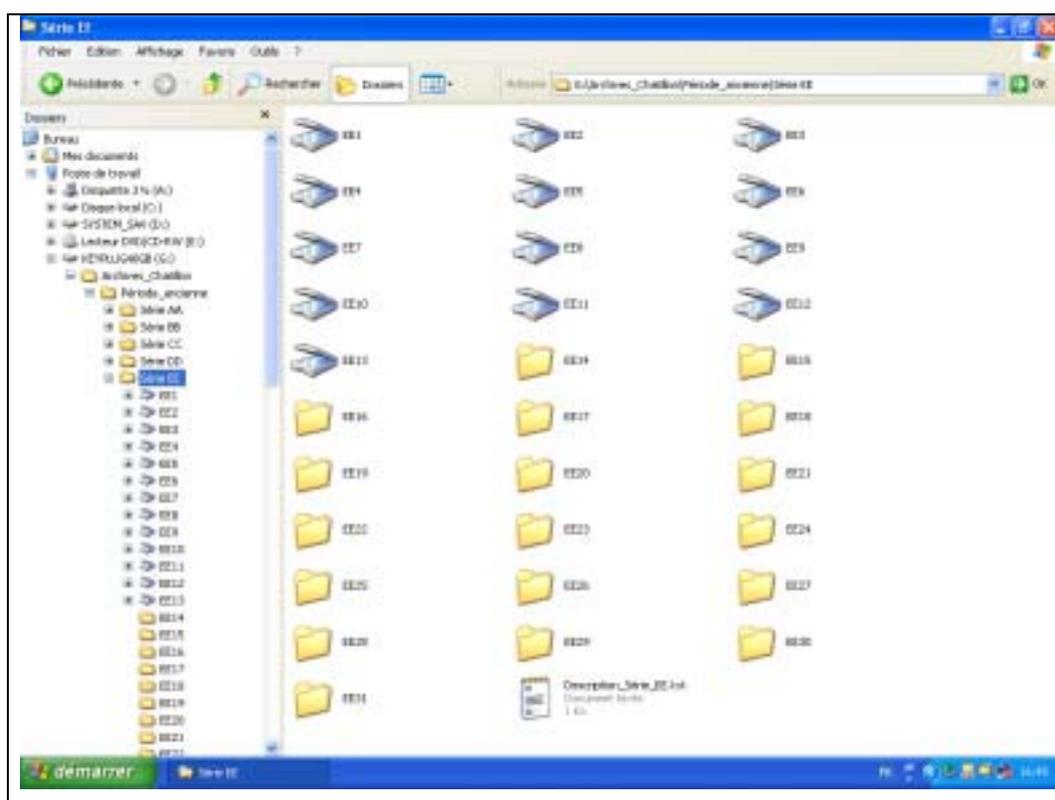
- Un premier niveau recense 4 périodes (période ancienne (1273-1790), période révolutionnaire (1790-1804), période moderne (1805-1982), période contemporaine (1983-...)).
- Un deuxième niveau constitué de séries reprend les différents aspects de la vie de Châtillon désignés par ce que nous considérons comme des *mots clés*⁸ (Ex : « Série CC : finances, impôts et comptabilité »).
- Un troisième niveau établit une chronologie dans les thèmes du niveau supérieur (Ex : « CC1 : comptes des syndics, 1374-1375 »).

⁷ Il est aussi l'auteur de l'histoire de Châtillon avec une étude systématique des comptes de syndics entre 1375 et 1500, mentionnée *supra* note 6.

⁸ Cf. sur le plan terminologique la distinction entre *mots-clés thématiques* et *mots-clés métalinguistiques*, TURCAN, Isabelle. *Introduction au Grand atelier Historique de la langue française*, Cdrom publié chez Redon en 2001, diffusion Le Robert.

A ce niveau nous avons affaire à une collection de manuscrits reliés en cahier ou regroupés en feuilles volantes. Certains de ces manuscrits ont été annotés postérieurement à leur création, certainement par les archivistes. Cette atteinte à l'intégrité de l'archive nous renseigne cependant sur le contenu (thème plus précis que l'intitulé de la sous-série, date rapportée au jour), tout en nous renseignant sur la vie du manuscrit de siècles en siècles, ce qui oriente forcément nos choix de traitement numérique.

L'une de nos premières activités a été de créer une arborescence de dossiers destinée à recevoir les images numérisées. Cette arborescence reprend la structure du classement des archives (emboîtement de dossiers), permet le renseignement sur l'état d'avancement de la numérisation (indication du nombre d'images numérisées, du poids représenté par les images) et constitue la base pour une exploitation ultérieure sur cédérom ou sur un site web.



* L'icône d'un scanner permet de faire apparaître visuellement les sous-séries numérisées. Dans chaque dossier est présent un fichier texte qui décrit sommairement le contenu de la série, de la sous-série ou de la pièce dans son unité.

3. Les archives historiques de Châtillon : enjeu patrimonial, outil de connaissance linguistique et développement d'applications

La numérisation des archives historiques de Châtillon répond à trois objectifs immédiats :

- Porter à la connaissance des habitants de la ville la richesse de leur patrimoine par une mise en relation des textes avec des éléments matériels encore existants ou dont l’empreinte est encore visible, tel le bâti (portes de la ville, remparts, église). Cet objectif s’appuie aussi sur une démarche pédagogique qui vise à favoriser une réappropriation ancrée dans l’histoire d’une communauté et à encourager le souci collectif de la préservation du patrimoine châtillonnais.
- Offrir aux chercheurs en linguistique et plus précisément en lexicographie historique un corpus de textes représentatif de la période du moyen-français permettant l’analyse de l’évolution d’une composante de la langue française jusqu’à présent peu étudiée et mal connue faute de diffusion des documents, en particulier pour les parlers régionaux, les écarts entre la langue vernaculaire, écrite et parlée, et la langue française officielle⁹.
- Evaluer les outils de traitement automatique des images élaborés au sein du laboratoire LIRIS de l’INSA en vue de faciliter leur restauration, transcription et indexation.

Au delà de ces objectifs immédiats, la numérisation des archives va procurer un corpus documentaire important pour des recherches en sciences humaines et sociales. L’importance de la collection permet d’envisager des recherches dans de nombreux domaines historiques relatifs au patrimoine : histoire des textes manuscrits ; histoire des royaumes de France et de Savoie ; histoire locale ;

⁹ Cet aspect de la recherche est mené par Isabelle Turcan dont le travail sur les grands dictionnaires du XVII^e et XVIII^e siècle et plus précisément le *Dictionnaire universel [...] de Trévoux* permet de mettre en perspective la langue utilisée à Châtillon. La constitution de lexiques de termes repérés dans les archives et la confrontation avec les grands dictionnaires permet de repérer les évolutions et les formes linguistiques retenues. Comme nous le verrons cette démarche sera formalisée au moment de la valorisation des archives par une mise en relation de leurs contenus avec les articles du *Dictionnaire universel [...] de Trévoux*, qui est un des premiers sous l’Ancien Régime à avoir accordé de l’intérêt aux parlers régionaux (dans la continuité de J.Nicot, 1606, A.Furetière, 1690 et G.Ménage, 1650 et 1694).

histoire des idées et des hommes ; histoire de l'art et des techniques, de la langue, du droit, etc..

Cette multiplicité des axes de recherche doit être prise en compte tout au long de la chaîne de numérisation afin de permettre des accès diversifiés aux archives¹⁰. Cette prise en compte s'applique particulièrement aux paramètres de numérisation et d'indexation des contenus en fonction des publics visés.

4. L'accès au contenu des archives historiques

En dehors d'un public averti l'accès au contenu des archives manuscrites ne peut être immédiat. Ces difficultés tiennent à la fois

- à la nature des archives : archives manuscrites parfois difficiles à déchiffrer
- à l'hétérogénéité des langues utilisées : latin plus ou moins classique, parlers régionaux appartenant à la famille du franco-provençal (Dombes, Bresse, Lyonnais, Val d'Aoste, etc..¹¹), français de référence en particulier pour le vocabulaire juridique.
- au vocabulaire concret utilisé, par exemple les termes techniques liés aux pratiques artisanales anciennes et aux métiers traditionnels.

4.1. Les archives manuscrites

La première difficulté repose sur le déchiffrement des écritures. La multiplicité des scripteurs dans la rédaction des manuscrits freine la reconnaissance manuelle des écritures. Pour pallier cette difficulté, l'objectif serait d'automatiser la reconnaissance des *images de mots*¹² dans les manuscrits. En établissant une correspondance entre l'image de mots et les mots de la transcription, il est possible de constituer une base d'images de mots manuscrits multi-scripteurs utilisable comme aide à la transcription.

¹⁰ En dehors même des usages qui ne sont pas envisagés. L'examen des projets de numérisation existants montre à quel point le « succès » de consultation d'archives peut dépendre de demandes « non-savantes ». Voir la demande sur les recherches généalogiques.

¹¹ Recherches d'Isabelle Turcan en cours.

¹² Terminologie utilisée par Frank Lebourgeois et Isabelle Turcan dans leurs travaux respectifs.

4.2. La transcription diplomatique

La transcription diplomatique consiste à restituer aussi fidèlement que possible le contenu et l'organisation d'un texte. Cette opération nécessite la codification et le balisage du texte pour marquer aussi bien les caractéristiques calligraphiques du texte (par exemple l'impossibilité de déchiffrer une lettre, un mot,...) que la structure physique du texte (espace entre les mots, saut de ligne, décroché,...). La transcription diplomatique pour être complète suppose à la fois la reproduction d'un texte et sa description, même de façon très codifiée sous forme de protocole¹³.

4.3. La traduction

La traduction consiste à rendre le texte accessible en français moderne à partir de la transcription diplomatique. Cette opération simple en apparence soulève cependant un certain nombre de questions :

- Que faire avec les termes qui n'ont pas d'équivalents en français moderne ou recouvrent une réalité différente selon les époques ?

Pour répondre à ces interrogations, il est prévu dans l'application finale une mise en relation avec des ressources externes qui pourraient s'organiser selon trois niveaux :

- L'accès à partir de la traduction à un lexique qui donnerait une définition minimale de façon binaire : le mot de la transcription = le mot en français de référence
- Pour ce même mot et, chaque fois que cela est possible, indiquer l'historique, la période de son apparition reconnue par les lexicographes dans la langue (qui peut être antidatée par les manuscrits), ses transformations formelles et ses variantes sémantiques.
- Renvoyer au mot tel qu'il apparaît et est décrit dans les dictionnaires sous l'Ancien Régime, du *Thresor* de J.Nicot au *Dictionnaire universel [...] de Trévoux*.

¹³ Travaux en cours dont le cdrom de démonstration donnera l'essentiel des choix adoptés.

La numérisation des archives

1. Les acteurs du projet de numérisation

1.1. L'association culturelle Saint-Guignefort

Le lancement du projet de numérisation s'est fait à l'initiative de l'association culturelle Saint-Guignefort, association loi 1901 à but non lucratif, qui vise à valoriser les différents éléments patrimoniaux de Châtillon-sur-Chalaronne et à porter à la connaissance du public et des chercheurs la collection des manuscrits tout en la préservant : l'objectif patrimonial est de faire connaître les archives anciennes, mais conjointement d'en limiter ou même éviter la consultation directe par souci de sauvegarde du patrimoine écrit ancien¹⁴.

Dans ce cadre une première série d'archives a été numérisée par deux membres de l'association¹⁵ à partir d'une thématique particulière, l'ensemble des manuscrits relatifs à la rénovation de l'église Saint-André à partir de la fin du XIV^e siècle. Ce premier travail de numérisation dont la finalité est la publication d'un ouvrage¹⁶ a réuni une équipe complète pour mener à bien la valorisation du contenu des archives (numérisation, transcription, traduction, rédaction d'un lexique, photographies, dessins de bâtiments). Cet exemple montre à l'évidence que la valorisation d'archives historiques implique outre la numérisation, une chaîne de traitement avant d'aboutir à une valorisation et ceci indépendamment du support choisi, papier ou numérique.

1.2. L'intervention de Caroline SARRET, étudiante en histoire de l'art médiéval

L'intervention de Caroline Sarret se situe dans une démarche proche de celle de l'association de Saint-Guignefort. A partir d'un sujet de recherche portant sur le

¹⁴ Pour une présentation complète et synthétique du projet de numérisation voir :

TURCAN Isabelle, LE BOURGEOIS Frank. *La numérisation des archives historiques de la municipalité de Châtillon-sur-Chalaronne dans l'Ain*. 2004. La Rochelle.

¹⁵ MONNIER.J et MARET J.L qui ont numérisé les comptes de syndics CC1 et CC2 en utilisant un scanner CANON SCANGEAR CS 7.0.0.

¹⁶ « Ils ont construit l'église de Châtillon-les-Dombes », parution prévue pour le second semestre 2004.

patrimoine bâti du vieux Châtillon¹⁷ et la gestion du patrimoine défensif à partir du XIV^e siècle et jusqu'au XVIII^e siècle, elle a participé au programme de numérisation dans le cadre d'une convention de stage : outre la numérisation des manuscrits correspondant au corpus documentaire défini pour son mémoire de maîtrise¹⁸, elle a contribué à la transcription et à la traduction commentée d'un certain nombre d'archives. Cet objectif de recherche, placé sous la responsabilité scientifique d'Isabelle Turcan, supposait une numérisation sélective et l'attention s'est portée sur deux séries d'archives : la série DD comprenant les biens communaux, eaux, forêts, travaux publics et la série EE comprenant les affaires militaires. Ce travail de numérisation s'est déroulé sur les mois de mai, juin et juillet 2004.

1.3. Notre intervention

Notre intervention dans son principe diffère de celles citées précédemment :

- Elle s'inscrit dans le cadre de la convention de recherche passée entre la mairie de Châtillon, l'association Saint Guignefort et le LIRIS (coordonnée par Isabelle Turcan représentant le laboratoire ATILF)
- Elle se donne comme objectif la définition d'un cadre de numérisation, d'indexation, de valorisation, pour servir conjointement une diffusion de qualité de documents patrimoniaux et le travail de recherche monographique.

Cet objectif spécifique impose une attention particulière dans les choix de protocoles techniques afin de permettre une valorisation numérique de la totalité du corpus. Il envisage l'ensemble de la chaîne de numérisation et des choix spécifiques doivent intervenir dès le déclenchement du processus de numérisation physique. Ces choix sont consécutifs à plusieurs interrogations fondamentales :

- Faut-il envisager un plan de numérisation ?
- Quels paramètres de numérisation choisir (format, résolution, nombre de couleurs, vitesse de numérisation) ?
- Quelles précautions prendre dans la manipulation des archives ?

¹⁷ L'étude de Caroline Sarret porte sur le patrimoine bâti qui comprend notamment : les remparts avec tours, les courtines, les portes, fossés et systèmes hydrauliques d'approvisionnement, etc...

¹⁸ Sous la direction de Mr Reveyron, Université Lumière Lyon II.

- Que doit-on numériser dans un document d'archive au cours de l'opération de recadrage ?
- Comment nommer les fichiers numériques d'images d'archives ?
- Selon quels principes stocker les fichiers numériques d'archives ?
- Faut-il indexer les critères physiques des archives ?
- Au premier stade d'enregistrement de l'image numérisée quels éléments de description du contenu de l'archive indexer ?

2. Les dispositifs de numérisation

Les dispositifs détaillés ci-dessous sont ceux que nous avons appliqués dans la numérisation des archives et que nous préconisons pour l'ensemble du corpus. Les principes de ces dispositifs visent à organiser, collecter un certain nombre d'informations au moment de la numérisation afin d'éviter un retour à l'archive physique et de permettre une consultation numérique sans déperdition d'un trop grand nombre d'informations. Il faut toujours garder en mémoire que l'image numérique n'est qu'une représentation d'un document original. Ces dispositifs concernent la production de l'image d'archive valorisée entre autres par les métadonnées qui peuvent lui être associées.

2.1. Le plan de numérisation

Il est fondamental que le plan de numérisation se fasse dans le respect du classement des archives tel qu'il existe aujourd'hui. Ce plan de classement procure une architecture globale des archives, les renseigne par un premier niveau d'indexation (thème, date, noms des syndics) et sa normalisation permet d'envisager une interopérabilité avec d'autres archives communales.

Il est aussi impératif de respecter ce plan de classement afin de conserver des possibilités de correspondance entre le document d'archive physique et son image numérisée. Pour ce faire, nous considérons que l'entité de base repérable de ce classement est la sous-série¹⁹. Une sous-série peut contenir différents types d'archives (feuillet, cahier assemblé ou non) à l'unité ou en plusieurs exemplaires,

sur un thème commun et dans une continuité chronologique. Il est fortement conseillé de numériser l'intégralité du contenu d'une sous-série et ceci dans l'ordre de classement opéré par l'archiviste. En dehors de cette contrainte, les séries et les sous-séries peuvent être numérisées selon l'exigence de recherche monographique envisagée en nommant comme nous le verrons par la suite les dossiers et fichiers conformément au classement des archives.

2.2. Les paramètres de numérisation

Les paramètres techniques de numérisation choisis doivent permettre de répondre aux usages envisagés quel que soit le public visé, aussi bien dans le premier traitement numérique des archives que dans leur consultation sur écran.

2.2.1 Le format de fichier JPEG

Ce format de fichier est fondé sur une méthode de compression dite « psychovisuelle ». Elle agit par réduction de l'information de couleur mais les changements demeurent imperceptibles à la vision humaine. En réduisant le poids des images, cette méthode de compression est particulièrement adaptée aux images qui sont destinées à la diffusion sur le Web. En revanche ce format provoque une perte d'information irréversible sur les fichiers image par réduction des informations spatiales inutiles et redondantes. Cet aspect peut cependant s'avérer problématique pour les images destinées à être océrisées²⁰ mais dans le cas présent ce traitement n'est pas envisageable, la reconnaissance de caractères restant impossible sur des images contenant du texte manuscrit de mains hétérogènes. Si le format JPEG est bien compatible avec les traitements envisagés dans l'aide à la transcription et non basés sur le traitement fin des caractères (séparation recto/verso, séparation texte/fond, repérage de métadonnées), on peut souligner les avantages qu'aurait présenté un format TIFF : éviter la compression destructive au moment de la numérisation, éviter le cumul des compressions en JPEG et réserver ce format pour la seule valorisation finale.

¹⁹ Exemple : dans la série CC(finances, impôts et comptabilité), CC14 (comptes des syndics, 1449-1450) correspond à la sous-série.

²⁰ L'océrisation consiste à reconnaître les caractères dans un fichier image et permettre ainsi sa conversion en fichier texte.

2.2.2 Une résolution de 400 dpi

La résolution donne l'indication du nombre de points ou pixels²¹ présents sur une surface donnée. Cette dimension s'exprime en pouce et correspond à 2,54 cm. Il est à noter que la résolution (en dpi ou ppp) lie le nombre pixels d'une image à ses dimensions réelles. Pour la numérisation des archives la résolution retenue est de 400 dpi ce qui signifie que dans un carré de 2,54 cm x 2,54cm nous avons 160 000 pixels. Cette résolution peut paraître élevée au regard de la résolution affichée par les navigateurs internet (72 dpi) mais il faut considérer que cette définition est une résolution destinée à la fois au traitement des images et à leur conservation²². Il est clair que pour la diffusion des images nous adapterons la résolution aux contraintes et aux besoins envisagés (poids des images, possibilité de zoom).

Autre élément qui a dû être pris en compte sans être déterminant, les capacités du scanner utilisé. Avec une résolution de 1200 dpi, le format de compression JPEG choisi doit correspondre à un multiple de 2, 3, 4,... Avec un multiple de 3, la résolution passe à 400 dpi ($1200/3=400$ dpi par suppression de 2 pixels sur 3 qui peuvent être inutiles et redondants).

2.2.3 Une palette de couleurs de 24 bits en mode RVB

Le codage en 24 bits²³ est celui utilisé par la plupart des écrans d'ordinateurs actuellement car, avec plus de 16 millions de couleurs, il permet une restitution particulièrement fidèle des nuances de couleurs. Le mode RVB pour rouge, vert, bleu, fonctionne sur la synthèse de ces trois couleurs et permet une gamme de couleurs très étendue. Dans la consultation des archives, cette palette de couleurs procure un confort visuel, esthétique et facilite l'accès à un nombre d'informations supplémentaires, textuels ou formels.

2.2.4 La vitesse de numérisation

La vitesse de numérisation dépend à la fois de l'électronique, de la mécanique du scanner et du type d'interface qui relie le scanner à l'ordinateur. Sur le scanner que

²¹ Le pixel est l'unité de base visible sur un écran. Pour les écrans couleur il est constitué des trois couleurs (rouge, vert, bleu) dont la variation d'intensité permet d'obtenir des milliers de couleurs différentes.

²² Une résolution trop faible peut interdire le traitement des images ou la réponse à des usages plus pointues. Par exemple, permettre le zoom sur les images jusqu'au point de pouvoir déterminer la texture du papier et sa composition.

²³ L'information est codée en binaire. Avec 1 bit on a 2 possibilités (noir ou blanc), avec 24 bits on a $256 \times 256 \times 256$ possibilités.

nous avons utilisé, la vitesse n'est pas un élément paramétrable et dépend des choix de résolution et formats. En revanche, l'interface choisie SCI (Small computer standard interface) permet, par rapport à une interface USB, d'accélérer le transfert des images en destination de l'ordinateur.

2.2.5 Autres paramètres de numérisation utilisés

Les premières numérisations réalisées par l'association Saint-Guignefort se sont faites selon plusieurs paramètres. Nous avons recensé l'utilisation des formats JPEG et TIFF avec des résolutions variant de 200 à 300 dpi. Plusieurs palettes de couleurs ont été utilisées, couleur en 8 bits ou niveaux de gris en 8 bits. Cette variété de formats²⁴, si elle n'a pas d'incidence pour les opérations de transcription et traduction, peut s'avérer plus problématique pour le traitement des images en lots et briser l'homogénéité au moment de la consultation. C'est pour cette raison que nous avons été conduit à refaire certaines images en couleurs avec l'objectif d'homogénéiser notre corpus.

2.2.6 Récapitulatif des formats utilisés

	Paramètres utilisés et préconisés (en caractère gras)
Format	JPEG
	TIFF
Résolution	200 dpi
	300 dpi
	400 dpi
Couleurs	Niveaux de gris, 8 bits
	Couleurs, 8 bits
	Couleurs, 24 bits

2.3. La manipulation des archives

Les archives comme documents papiers ou parchemins, dont les premiers remontent au XIV^e siècle, exigent des précautions particulières au moment de la manipulation. Le support de la plupart des archives de Châtillon est le papier tel

qu'il était fabriqué jusqu'à la moitié du XIX^{ème} siècle, à base de chiffon de lin, de chanvre puis plus tard de coton. A cet égard nous pouvons noter que le premier moulin à papier français est mis en service à Troyes en 1348²⁵, soit vingt six ans avant les premières archives des comptes de syndics de Châtillon Si les matériaux utilisés à cette époque sont bien adaptés à la conservation, des précautions doivent être prises en fonction de la nature des fibres du papier de chaque manuscrit. Certains manuscrits apparaissent comme durs mais peuvent casser si le pli d'origine du feuillet est contrarié. D'autres plus souples s'effrangent aux extrémités. Ce dernier phénomène semble le plus fréquent et est lié à la fréquence de manipulation mais aussi aux mauvaises conditions de stockage : ainsi dans un classeur inadapté quand le poids du manuscrit repose sur la tranche inférieure ou quand la taille de la sangle d'un classeur n'est pas adaptée aux dimensions du manuscrit ; à cela s'ajoute un mode de rangement qui comprime trop les documents. Cette érosion des manuscrits peut se constater à partir du relevé de leur dimension et de l'écart négatif qui apparaît au regard d'un format standard en fonction des typologies de documents conservés.

2.4. Que faut-il numériser dans un document d'archive ?

Cette question apparaît de façon évidente à partir du moment où nous admettons que la numérisation n'est qu'une représentation et qu'elle peut occasionner une perte d'informations ou un biais dans la perception matérielle du document²⁶. Prenons quelques exemples et voyons quels choix de numérisation peuvent être faits pour conserver certaines informations susceptibles d'être jugées à première vue non pertinentes par un non spécialiste et pour tenter de restituer dans son ensemble la page physique, celle de la perception directe, sans rogner les marges, ni les coins, même endommagés. Cf images ci-dessous en 2.4.1.

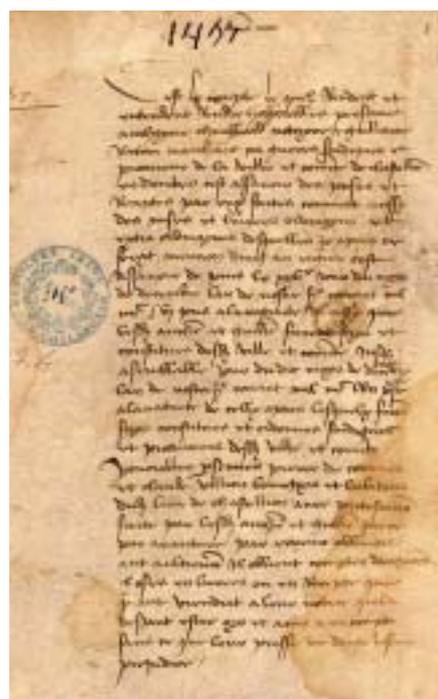
2.4.1 Les éléments matériels

Quand nous avons en main un document d'archives nous pouvons spontanément évaluer le nombre de folios et la masse physique de texte présent sur les pages. A

²⁵ GALLIMARD FLAVIGNY Bertrand. *Etre bibliophile, petit guide pratique*. Paris : Séguier, 2004, 235 p.

²⁶ Voir la réflexion sur les livres électroniques, le feuilletage proposé par certains sites web, comme tentatives de retrouver les « sensations » d'une version papier.

la numérisation, cette information risque de disparaître, sauf si l'image contient des éléments nous renseignant sur cet aspect. Dans les deux exemples de numérisation ci-dessous, nous pouvons comparer le même manuscrit numérisé dans des conditions différentes: le premier fait apparaître la tranche et le dos des manuscrits alors que le second est centré uniquement sur le texte.



Comme nous le voyons ces deux images donnent accès au texte de la même façon mais la seconde nous renseigne approximativement sur le nombre de folios, offre un meilleur rendu esthétique et semble mieux adaptée à une consultation finale par feuilletage. Cet exemple nous permet de mettre en évidence l'importance d'un cadrage adapté à l'exploitation du corpus envisagé. En raison de la fenêtre de numérisation limitée (format A4) et des formats de manuscrits divers, nous préconisons de privilégier avant tout et malgré tout l'intégrité du texte, mais dans la mesure du possible de préserver l'ensemble du feuillet contenant un élément textuel même minimal, telle une simple marque de foliotation, de faire apparaître la tranche et le dos des manuscrits. A l'extrême, on préconisera même de

numériser des feuillets vierges en particulier lorsqu'ils font partie d'une série foliotée²⁷.

2.4.2 Les éléments non-textuels

En dehors des éléments textuels d'autres signes particuliers apparaissent sur les manuscrits : tampons, cachets, marques d'imprimeurs et autres signes particuliers. Quand ceux ci sont intégrés au cœur du texte leur numérisation est systématique mais comment procéder quand ils figurent sur des folios vierges de tout élément textuel ? Un constat s'impose ; ces éléments sont susceptibles de nous donner de précieuses informations sur l'environnement économique et culturel de la ville de Châtillon : par exemple les marques d'imprimeurs sur le papier nous permettent d'avoir une information sur les lieux de production et l'état des relations commerciales avec l'apparition du papier en France à partir du milieu du XIV^e siècle. Il est donc important de les numériser.

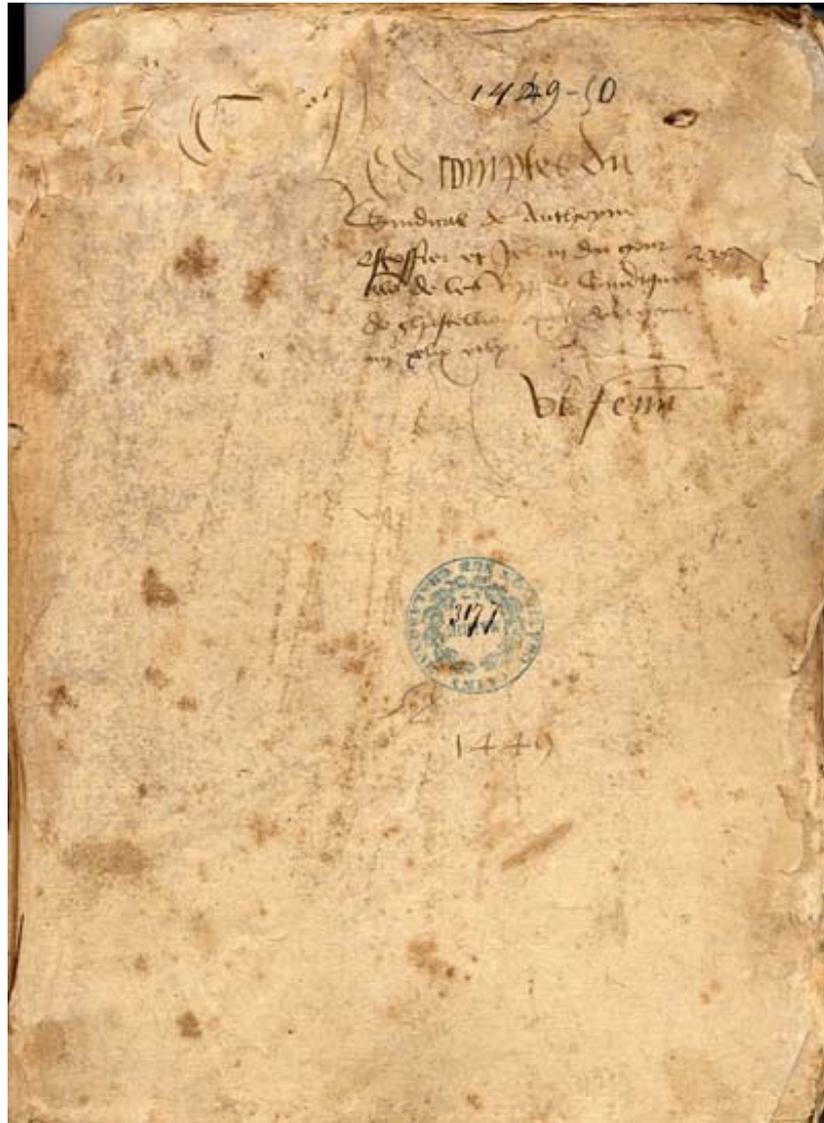
De même, certains folios vierges de tout texte portent une marque ostensible du scripteur. Le cas le plus fréquemment rencontré est celui d'un folio barré en croix sur toute sa dimension. Un tel folio doit être impérativement numérisé, il nous donne une indication précieuse de la volonté du scripteur de séparer, d'instituer une rupture entre plusieurs éléments textuels.

2.4.3 Les éléments ajoutés

Nous avons constaté également que des éléments textuels postérieurs à la production des archives figurent sur certains folios ou sur des feuillets joints. Ces éléments ont été pour la plupart inscrits au crayon par les premiers archivistes en charge de l'inventaire et mentionnent la date et le thème de l'archive. On est donc en présence de plusieurs stades différents d'annotations (qualité différente d'encre, mains identifiables,...), ce qui peut donner lieu à un souci de distinction matérialisée par un traitement particulier des images. Si l'effacement physique de ces annotations est exclu sur le manuscrit lui-même, il pourrait être envisagé dans une version de publication électronique ultérieure, de gommer numériquement ces

²⁷ A titre d'exemple voir dans la sous-série CC10 constituée en cahier, le folio 2 verso vierge de toute inscription. Nous pouvons émettre l'hypothèse que l'organisation spatiale du texte dans un document organisé procède d'une volonté délibérée de la part du scripteur ou des scripteurs. A ce titre, elle possède une signification qui peut nous renseigner sur le contexte du document et nous éclairer sur sa compréhension.

commentaires. En revanche et nous le verrons par la suite, une trace de l'existence de ces annotations doit être impérativement conservée comme élément d'information et premier niveau d'indexation pour toute recherche éventuelle sur la réception de ces manuscrits.



Manuscript: CC14_piece001_folio001recto

Nous pouvons voir le tampon apposé par l'archiviste ainsi que plusieurs indications de dates qui semblent provenir de scripteurs différents.

2.5. Comment nommer les fichiers d'archives ?

Au cours du processus de numérisation physique, le logiciel de pilote du scanner (EPSON TWAIN, EPSON Scan II32 Vers 2.35E) nomme automatiquement les images en utilisant dans l'ordre :

- le nom du dossier de destination choisi
- un numéro de fichier incrémenté²⁸ automatiquement de 0 à 99
- le format de fichier

Ce qui donne par exemple pour le nom de fichier DD101.jpg (DD1 pour le nom du dossier, 01 pour le numéro de fichier et .jpg pour le format).

Le nom de fichier généré permet de mentionner la sous-série d'archives dans le choix du nom de dossier, mais l'incrémentation à deux chiffres oblige, arrivée à 99 à créer un autre dossier ; de ce fait le nom de fichier généré ne rend pas compte de l'unité de base de l'archive qui peut être le cahier, le feuillet ou le folio. L'enjeu de cette désignation est de conserver un lien entre l'organisation physique des archives, y compris au sein d'une sous-série, et les images numérisées. Cette désignation n'a pas pour objet de décrire le contenu des manuscrits, puisque cette description sera assurée par les métadonnées, mais de conserver une image de l'organisation physique des archives qui reste notre élément référent.

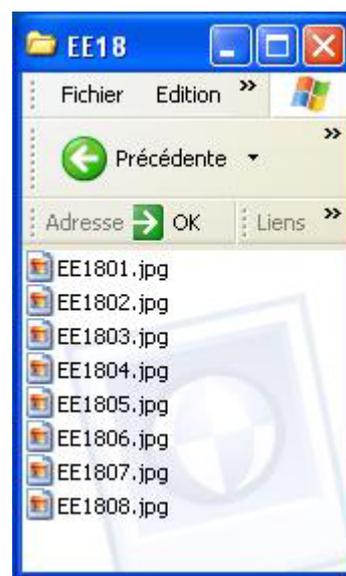
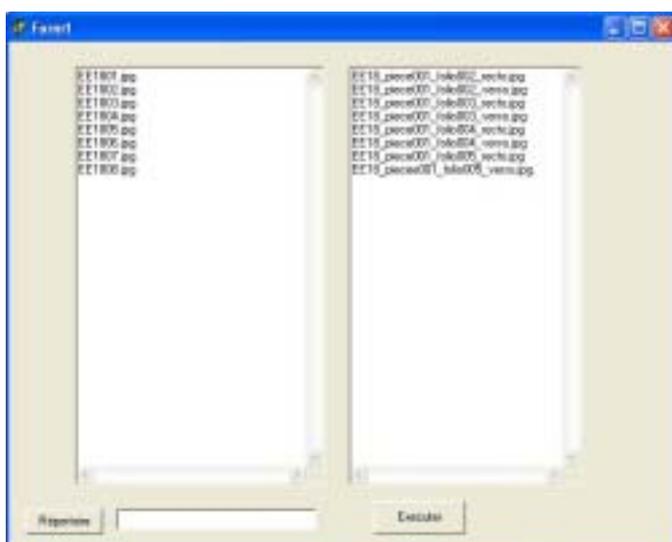
Quand plusieurs documents sont présents au sein d'une sous-série d'archives, nous avons choisi le terme de *pièce* pour les désigner, terme employé par les archivistes pour nommer la plus petite unité de description indivisible à la fois matériellement et intellectuellement (feuillet simple ou double, plusieurs feuillets reliés, cahier, registre,...). Cette opération d'inscription d'un nom de fichier définitif, comme d'autres opérations de description de l'archive, se fait en parallèle du processus de numérisation physique.

Par exemple, prenons le nom de fichier généré automatiquement suivant : DD101.jpg. Comme nom de fichier définitif nous indiquerons, DD1_pièce001_folio001_recto.jpg. Le nom de fichier définitif permet d'identifier l'unité de base (pièce) et le positionnement du folio (numérotation 001, indication du *recto* et du *verso*). Pour prendre en compte les lots d'archives volumineux,

l'incrémement pour les pièces comme pour les folios se fait sur trois chiffres. Pour ce faire nous utilisons un modèle de document au format Excel, document intermédiaire permettant d'autres exploitations dans la chaîne de traitement (renommer automatiquement les fichiers) et d'indexation (exporter les éléments de description de l'archive vers une base de données).

2.6. Les fichiers renommés automatiquement

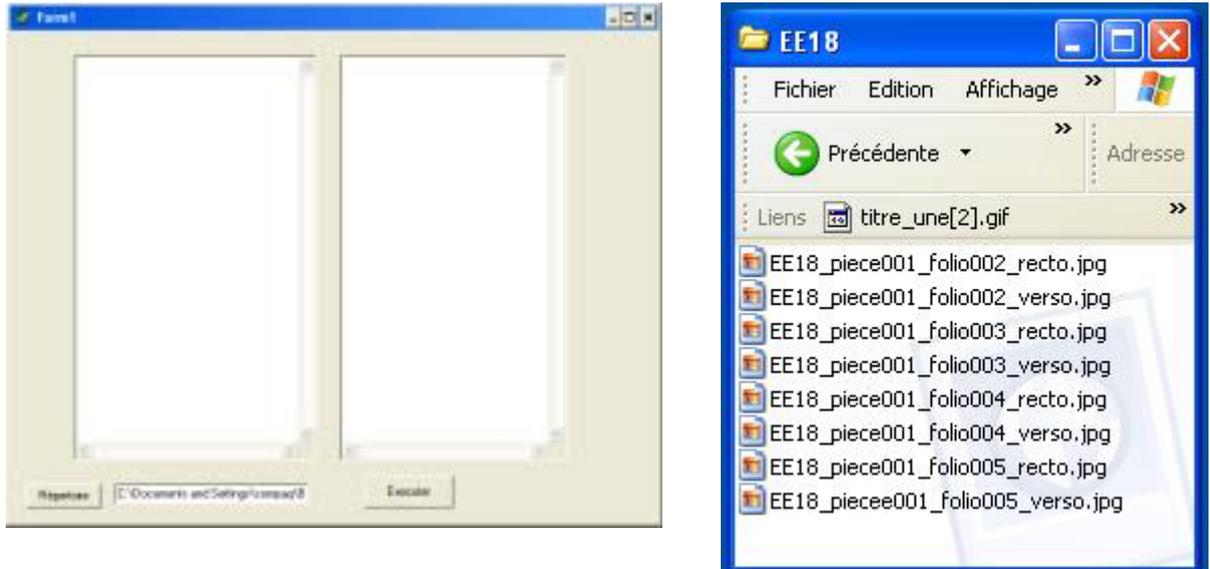
Dans le fichier Excel créé apparaissent deux colonnes, l'une avec le nom de fichier généré automatiquement, l'autre avec le nom de fichier définitif Cf image d'écran ci-dessous. A l'aide d'un script écrit en langage Pascal²⁹, après avoir sélectionné le dossier contenant les fichiers et exporté le contenu des deux colonnes, les fichiers sont automatiquement renommés.



Après exécution du programme, les fichiers sont renommés automatiquement comme on peut le voir à la page suivante:

²⁸ Incrémenter : Augmenter régulièrement la valeur d'une variable par une quantité constante prédéfinie (*incrément*), par exemple dans la boucle d'un programme. Définition à partir de : *Le grand dictionnaire terminologique*, [en ligne]. http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp (page consultée le 17 septembre 2004)

²⁹ Nous remercions Frank Lebourgeois d'avoir écrit à notre demande ce script qui peut être installé sur n'importe quel poste de travail. Deux fichiers issus d'un compilateur Pascal (rtl60.bpl et vcl60.bpl) doivent être copiés dans le répertoire de windows (C:/windows/system32).



2.7. Le stockage des fichiers d'archives

Le stockage des fichiers d'archives soulève la question de la conservation et de la pluralité des supports, du poids estimé des fichiers archivés, des capacités d'accès que permettent ces supports.

L'architecture que nous préconisons dans le circuit des fichiers d'images et leur conservation est la suivante :

- Au moment de la numérisation, les fichiers sont enregistrés sur le disque dur du poste de travail et sont nommés automatiquement par le pilote du scanner.
- Une fois qu'une sous-série a été numérisée, elle est alors copiée sur un disque amovible externe et les fichiers sur ce disque sont renommés à l'aide du script présenté au chapitre 2.6 et conformément au modèle décrit dans le chapitre 2.5.
- Les fichiers ainsi traités peuvent être enregistrés sur deux types de supports, un support magnétique qui sera le disque d'un serveur et un support optique (CD-R ou DVD-R). A partir de cet instant, les fichiers peuvent être effacés du disque dur du poste de travail comme du disque amovible.

Dans cette troisième étape, les fichiers sont organisés dans le serveur selon la classification de dossiers présentée dans le chapitre 2 et conforme au classement

physique des archives. Le serveur doit posséder une capacité suffisante : la capacité minimale est estimée à 300 GO, avec des dispositifs de sauvegarde interne. C'est à partir du serveur que se fera l'accès aux archives dans une optique de développement Web. En conséquence, les données devront être rapidement et facilement accessibles.

La conservation sur support optique devra respecter l'intégrité des séries ou sous-séries d'archives et ce critère est déterminant : il faut proscrire la conservation isolée d'un fichier d'image non intégrée dans sa série ou sous-série d'origine. Le tableau ci-dessous nous donne un aperçu des différents supports, de leur capacité de stockage et de la relation avec les besoins du programme de numérisation.

Supports	Capacité en GO	Equivalent CD-R	Unités de supports nécessaires³
CD-R ¹	0,65	1	434
DVD-R ² Standard	4,7	7	60
DVD-R ² Dernière génération	17	26	17

1 CD-R ou CD-WORM, ils sont inscriptibles une seule fois.

2 DVD-R standard ou de dernière génération, ils ne sont pas réinscriptibles et non effaçables.

3 Avec un nombre total de 27 000 fichiers, le poids total des fichiers est estimé à 282 GO. L'estimation d'un poids moyen de chaque fichier a été faite à partir des fichiers déjà numérisés et à partir d'un échantillon de 2942 fichiers provenant des séries DD et EE.

Les supports optiques (CD-R ou DVD-R) ont une durée de vie limitée³⁰ qui dépend des matériaux utilisés, de la vitesse de la gravure, des conditions de conservation. Pour les matériaux utilisés, il est conseillé de retenir les produits de type « Gold » qui se montrent à l'usage beaucoup plus fiables en abandonnant ceux de type « Silver »³¹. Il est possible de minimiser les risques au moment de la gravure en choisissant de le faire en basse vitesse (maximum 16X pour les CD et 1X pour les DVD). Quant aux conditions de conservation, il faut conserver ces supports à l'abri de la lumière, de l'humidité et des variations de températures. Une fois

³⁰ Durées de vie estimées entre 5 et 200 ans selon les matériaux utilisés.

l'ensemble de ces précautions prises, il est malgré tout judicieux de vérifier une fois par an la qualité des fichiers et de les recopier si besoin est.

Compte tenu de tous ces éléments nous préconisons plutôt la conservation sur des supports type DVD-R, qui avec des qualités de matériaux équivalentes, facilitent le stockage et la vérification des fichiers. Pour les opérations de transcription des manuscrits et compte tenu de l'utilisation intensive du support, il est préférable d'établir une copie qui, selon l'importance de la sous-série sera établie sur CD-R ou DVD-R.

Pour être complet sur le stockage des fichiers d'archives, il faut mentionner le temps important consacré au transfert des fichiers, de l'ordinateur au disque amovible, du disque amovible au serveur et celui consacré à leur gravure.

2.8. Numérisation physique et première indexation

Nous avons noté précédemment l'intérêt de l'utilisation d'un document type Excel, pour l'indication d'un nom de fichier définitif parallèlement au processus de numérisation physique. Outre cet élément, il nous semble judicieux d'intégrer dans ce document des éléments dont l'observation se trouve malaisée voire impossible dans une consultation numérique. Comme pour la problématique de la nomination des fichiers, il faut conserver à l'esprit que l'un des objectifs du projet de numérisation est la préservation des manuscrits. Cet objectif de préservation sera atteint si l'archive numérique, ses métadonnées, contiennent suffisamment d'informations et permettent d'éviter un retour à l'archive physique. Ces premiers éléments d'observation nous renseignent sur :

- la date et le thème des manuscrits tels qu'ils ont été signalés par l'archiviste (ce qui n'est pas systématique).
- la dimension des manuscrits, élément difficilement évaluable dans une consultation sur écran.
- l'état de conservation des manuscrits, en vue d'un traitement éventuel des images pour faciliter la transcription.

³¹ Il n'est pas nécessaire de se tourner vers des produits annoncés comme garantie sur plusieurs dizaines d'années, alors que l'évolution des matériels en rendra la lecture certainement impossible.

A ces indications, nous ajoutons une zone de commentaires permettant d'indiquer aussi bien :

- les conditions de numérisation (problèmes éventuels),
- les caractéristiques physiques des manuscrits (folio, feuillet, cahier, cahier relié,...),
- la logique de l'organisation des folios (dans une quittance, le folio contenant l'objet et la date de la quittance est situé généralement à la fin du document).

Dans une perspective de consultation numérique grand public, il serait préférable que ce folio apparaisse comme folio d'entrée, ce qui impose en revanche que l'information de cette modification de la structure physique soit mentionnée. Toute intervention sur les images des documents originaux doit faire l'objet d'une note de traçabilité qui doit être archivée et répond aux principes d'éthique numérique³². Il nous apparaît important d'insister sur l'intérêt pratique et pragmatique de ce journal qui est indispensable à toute opération de numérisation, qui doit accompagner quotidiennement les intervenants, aussi bien au moment même de l'utilisation du scanner que dans les opérations successives tout au long de la chaîne de numérisation.

Numérisation	Date	Folio	Caractéristiques	Observations	Statut	Autres	Commentaires
00001	2004	1
00002	2004	2
00003	2004	3
00004	2004	4
00005	2004	5
00006	2004	6
00007	2004	7
00008	2004	8
00009	2004	9
00010	2004	10
00011	2004	11
00012	2004	12
00013	2004	13
00014	2004	14
00015	2004	15
00016	2004	16
00017	2004	17
00018	2004	18
00019	2004	19
00020	2004	20

Image d'écran du journal de numérisation

organisé en colonnes pour les indicateurs choisis, les lignes correspondant aux folios numérisés ; certaines rubriques peuvent ne pas être renseignées.

³² Cf. Isabelle TURCAN, note infra 38

3. Le scanner, le pilote et l'interface de numérisation

3.1. Le scanner, un Epson perfection 1200S

Le scanner Epson Perfection 1200S est un modèle de scanner de bureau. Ses performances en résolution optique (1200 dpi x 2400 dpi), en échantillonnage couleur (24 bits), permettent d'acquérir des images en haute définition. En revanche, le format de sa vitre de numérisation (216mm x 297 mm, format A4) est un obstacle à la numérisation de documents dont les dimensions sont supérieures au format A4. Certains manuscrits ont des dimensions que le scanner ne peut pas prendre en compte (par exemple dans la sous-série EE14, les manuscrits ont des dimensions supérieures à la vitre du scanner 22 x 33 cm). Pour ces dimensions, nous numérisons les folios en deux parties (haut et bas) et nous les nommons en conséquence :

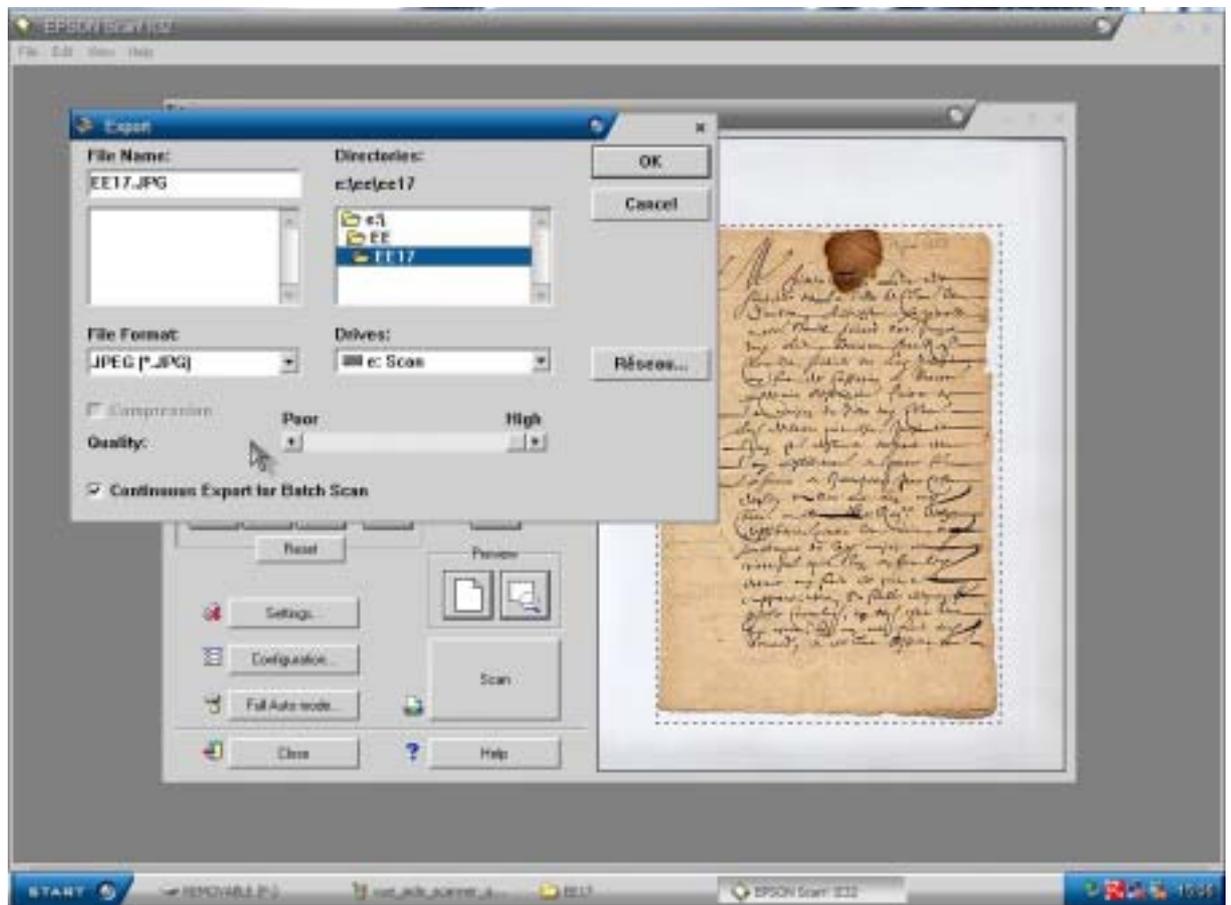
(Ex : EE14_pièce001_folio002_verso_hautpage).

A la différence des scanners dits « patrimoniaux » qui numérisent par le haut, l'Epson Perfection est un scanner dit « à plat ». Cette caractéristique nous oblige à prendre un soin particulier au moment de la numérisation. Il est impératif de respecter le pli d'origine des feuillets, de ne jamais les plier en sens contraire, et d'éviter d'accentuer la pliure sous le poids du battant du scanner. De même, à l'emplacement du manuscrit, entre la lampe de numérisation et le battant du scanner, nous constatons une certaine augmentation de température qui peut détériorer en cas d'exposition prolongée les encres et papiers. En conséquence, le manuscrit doit rester présent sur la vitre du scanner uniquement durant le temps nécessaire au processus de numérisation.

3.2. Le pilote du scanner, Epson Twain

Le pilote du scanner permet de créer une configuration personnalisée et un profil à partir de plusieurs paramètres. Les paramètres de configuration retenus sont ceux qui ont été préconisés dans le chapitre 2.2.5 sur les conseils de Frank Lebourgeois (format : JPEG, résolution : 400dpi, couleurs : 24 bits). Reste à paramétrer

manuellement au démarrage de la numérisation d'un lot, le niveau de compression des images : sur une échelle de 1 à 5, la qualité 5 dite « haute » a été retenue.



Un aperçu de l'interface du pilote du scanner

qui permet d'indiquer le nom générique des fichiers d'images, le dossier dans lequel il va être enregistré, le format des fichiers, la qualité de compression.

3.3. L'interface de numérisation

L'interface de numérisation telle qu'elle existe ne permet pas de visualiser le bon déroulement du processus de numérisation en cours ; par exemple comment vérifier qu'une image a été correctement numérisée et en un seul exemplaire ? Pour effectuer un contrôle en parallèle, il est conseillé de réorganiser la fenêtre de l'écran du bureau de l'ordinateur en faisant apparaître le contenu du dossier où sont enregistrés les fichiers d'images numérisées.

L'affichage doit se faire en mode miniature afin de permettre le rapprochement immédiat entre le manuscrit physique et l'image numérisée. Pour comprendre l'intérêt d'un tel dispositif, il faut imaginer l'ensemble des opérations à effectuer,

si, au moment de la consultation, on réalise qu'un manuscrit est manquant ou illisible. Cette interface de numérisation se présente de la façon suivante, avec un écran scindé en plusieurs sections comme on le voit ci-dessous :

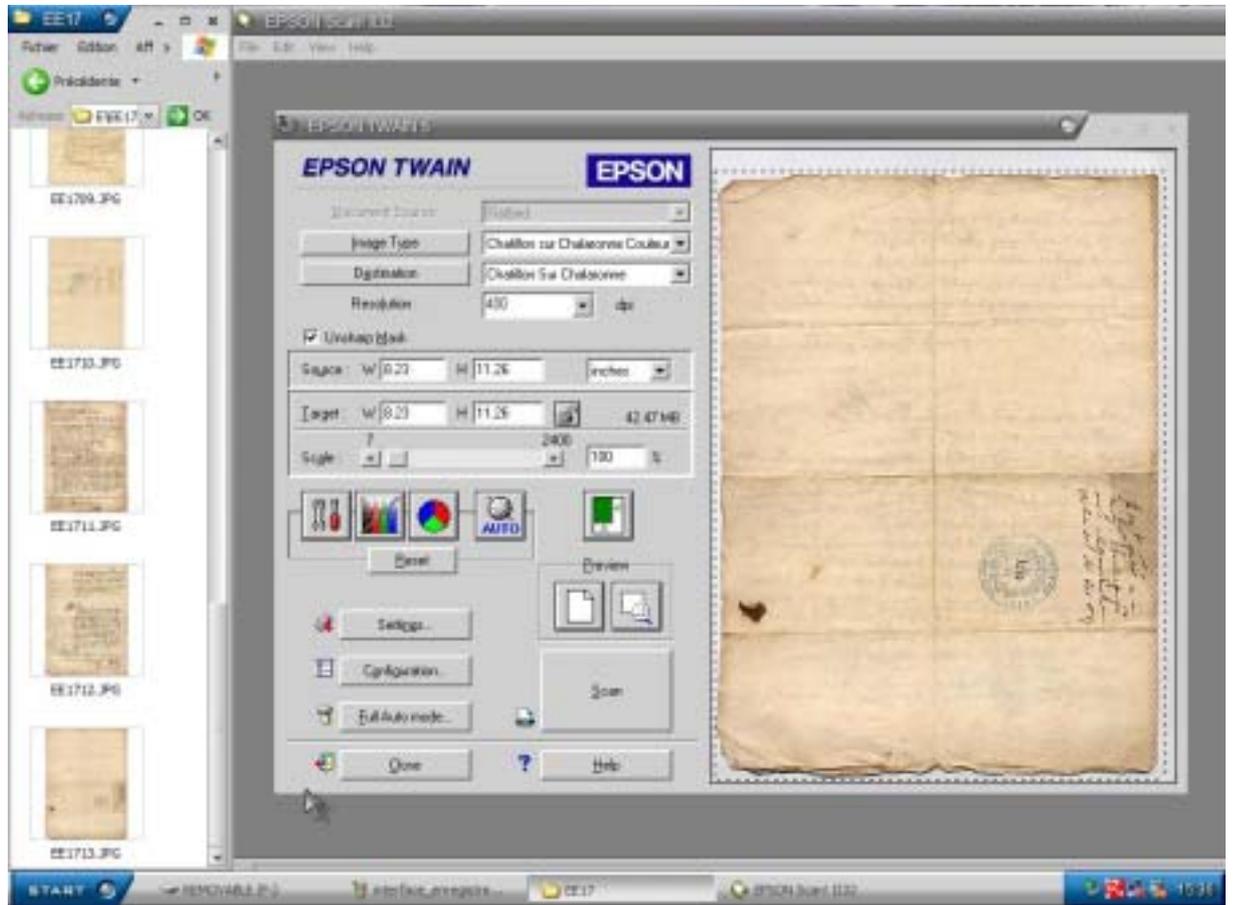


Image d'écran de notre choix d'interface de numérisation :

- à gauche, un écran déroulant correspondant au dossier de destination des images scannées permet de visualiser en icône l'ensemble des manuscrits d'une sous-série déjà numérisée ;
- à droite, l'image doit être mise en relation avec la section de gauche pour vérifier qu'il n'y a pas eu ni omission ni doublon ; elle permet d'apprécier la qualité de l'image.
- Au centre, la composante centrale permet une vérification supplémentaire du profil et des paramètres de numérisation.

4. L'évaluation des temps de numérisation

Avant le lancement d'un projet de numérisation, il est important d'évaluer le temps nécessaire à la numérisation d'une unité d'archive et de le rapporter au volume global à numériser. Notre pratique de la numérisation³³ nous permet d'estimer que le nombre moyen d'images numérisables est de l'ordre de 150 images³⁴ par journée de sept heures de scanner, et ceci en dehors de tout incident majeur. Si nous décomposons les opérations propres à la numérisation, ceci sans compter environ une heure consacrée aux opérations périphériques (lancement et installation de l'ordinateur, récupération des archives, création de nouveaux dossiers, nettoyage vitre de numérisation,...), nous obtenons le tableau suivant :

Opérations de numérisation	Temps estimé pour une image (en secondes)
Préparation archive	15
Prévisualisation	10
Redressement éventuel et nouvelle Prévisualisation ¹	5
Sélection cadre image	10
Numérisation	65
Rapprochement archive et image	5
Indexation	10
TOTAL	120

¹ Le temps indiqué est une moyenne. Certains manuscrits sont immédiatement cadrés alors que d'autres avec des dimensions importantes et du texte proche des bords de page, nécessitent plusieurs redressements et visualisations.

A partir de l'estimation d'un nombre de 150 images produites par jour et du volume du corpus à numériser estimé à 27 000 images, nous pouvons évaluer à 180 jours le temps nécessaire à la numérisation complète des archives historiques. En considérant qu'approximativement 25% de ce corpus a déjà été numérisé par les membres de l'association Saint Guignefort, par Caroline Sarret et par nous mêmes, il reste 20 250 images à réaliser soit l'équivalent de 135 jours.

³³ En raison de 2 jours par semaine sur 6 semaines entre mi-juillet et fin août 2004.

³⁴ Cette moyenne semble aussi avoir été celle des bénévoles de l'association de Saint-Guignefort pour les archives qu'ils ont numérisées avant la mise en place du programme ayant fait l'objet d'une convention officielle.

Le traitement numérique des archives

Dans ce chapitre nous aborderons le traitement numérique³⁵ des archives³⁶ qui est à différencier du traitement préalable à l'exploitation automatique des données et ceci dans la perspective de création d'une base de données.

1. Possibilités techniques et conditions d'utilisation

Le traitement numérique intervient sur la composante de base d'une image, le pixel, son intensité, sa position dans l'image et son modèle de représentation, niveaux de gris, couleurs. Les applications de traitement de l'image peuvent intervenir :

- comme correction des conditions physiques de numérisation, correction géométrique et d'éclairage, redressement ;
- comme transformation de l'image et de ses paramètres, conversion de couleurs;
- comme analyse de la structure physique des folios, distinction entre autres de zones textuelles et graphiques. Il est nécessaire d'utiliser ces applications avec discernement et selon le principe du respect de l'intégrité du document d'archive.

Dans l'utilisation de ces possibilités trois principes peuvent être envisagés³⁷ :

- la correction des conditions physiques de numérisation peut uniquement être appliquée à l'image du document si elle permet de tendre vers une conformité la plus fidèle possible de l'aspect physique du document d'origine, sans le trahir. Cette correction peut s'appliquer

³⁵ Pour une présentation des différents traitements numériques voir les cours en ligne de Frank Lebourgeois : <http://rfv6.insa-lyon.fr/flebourg/> (consulté le 13.09.2004)

³⁶ Non pas en tant que spécialiste de la programmation informatique mais en tant qu'expérimentateur soucieux de trouver les solutions les plus adéquats pour valoriser un corpus aussi particulier que celui des archives de Châtillon.

³⁷ Nous estimons préférable ici d'insister sur les principes sous-jacents à l'utilisation des potentialités des outils, plutôt que de les décliner sans prendre de position claire sur les risques afférents (par exemple, réaliser une très belle image qui ne corresponde plus à celle du document original).

indifféremment au moment de la numérisation physique, du traitement du contenu du document d'archive pour les opérations de transcription ou par l'utilisateur final au moment de la consultation.

- La transformation de l'image et de ses paramètres peut intervenir au seul stade de la transcription du document d'archive ou comme option clairement identifiée pour le consultant final. La notion de *numérisation en mode image amélioré* avancée par certains producteurs de contenus numériques suscite de nombreuses interrogations³⁸.

L'analyse de la structure physique des folios peut être utile à l'indexation des différents éléments constituant le document.

2. Le logiciel de traitement d'images Book Restorer

Le logiciel de traitement d'images Book Restorer est le résultat d'une collaboration entre le laboratoire LIRIS dans sa composante d'intégration d'outils logiciels pour le document numérique et la société I2S, fabricant de *scanners patrimoniaux*³⁹. Nous n'aborderons que les seules fonctionnalités de ce logiciel susceptibles d'être utilisées pour notre corpus d'archives. Pour la présentation de ces fonctionnalités nous utiliserons la typologie proposée au chapitre précédent.

2.1. Correction des conditions de numérisation

2.1.1 La correction géométrique et correction d'éclairage

La correction géométrique permet de supprimer la déformation des caractères qui intervient dans certaines conditions de numérisation; quand les manuscrits à numériser sont rassemblés sous la forme de plusieurs feuillets pour composer un cahier volumineux, le texte situé le long du bord intérieur du folio est susceptible

³⁸ TURCAN Isabelle. *Numérisation et édition électronique de dictionnaires anciens : éthique et pratique pour le respect des livres et de leurs consultants*, in Actes du colloque *Ethique numérique. Défis ou opportunités pour la compréhension des situations et des contextes de communication*. (Saint Cyr-sur-Mer, mai 2003), pp. 119-125.

³⁹ Les *scanners patrimoniaux* sont des scanners à plat avec une numérisation par le haut, dispositif permettant la préservation des documents et la numérisation des documents les plus divers dans leurs formats. Pour une étude complète et comparative des *scanners patrimoniaux* voir : ROUMIEUX olivier. *Scanners patrimoniaux, des matériels dédiés*. Guide pratique Archimag, 2004, Hors-série les scanners professionnels, pp. 74-90.

de subir une courbure qui se traduit au moment de la numérisation par un écrasement des caractères.

De même, la correction d'éclairage permet de pallier la sous-exposition localisée d'un document d'archive, à l'emplacement du pli des feuillets ou quand le manuscrit possède des dimensions à la limite de celles de la vitre de numérisation.

2.1.2 Le redressement des images

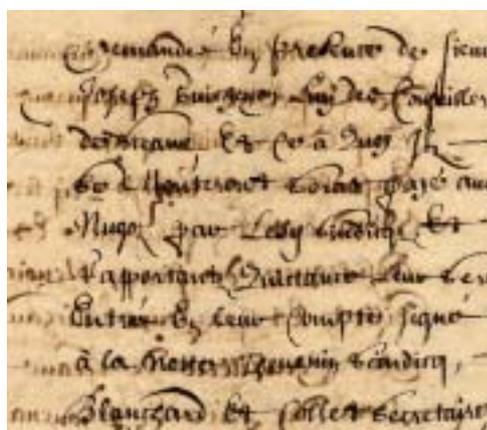
Le redressement permet de corriger l'inclinaison du manuscrit présente au moment de la numérisation. Ce défaut d'inclinaison par rapport à la fenêtre de numérisation se produit régulièrement quand les manuscrits organisés en feuillets, en cahier se déplacent sous le poids du battant du scanner et ceci pour les deux raisons de la fermeté du pli des feuillets et du manque de souplesse du papier. Plutôt que d'endommager le manuscrit en forçant le pli des feuillets, il est possible d'utiliser cette fonctionnalité. En revanche, l'opération de redressement provoque une perte d'information des données de l'image qui se manifeste sur le contour des caractères. L'utilisation de cette fonctionnalité doit donc rester exceptionnelle.

2.2. Transformation de l'image et conversion de couleurs

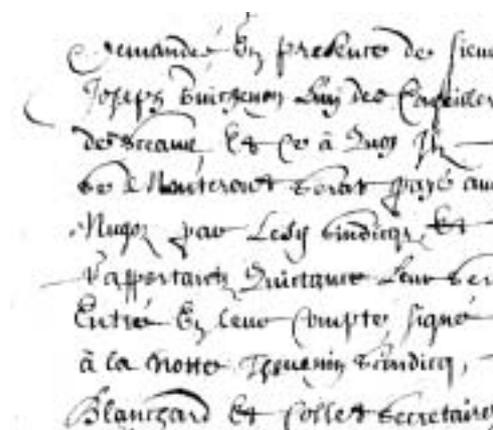
La conversion de couleurs permet de passer d'une image couleur à une image en niveaux de gris ou binaire (noir et blanc). La transformation en niveaux de gris d'une image couleur peut s'avérer très utile comme aide à la transcription. Si une image en niveaux de gris peut être préférée à une image en couleurs, c'est parce qu'elle permet de marquer les contrastes et d'approcher des conditions de lecture papier qui semblent plus adaptées à une utilisation prolongée ; d'autre part, la réduction du poids des fichiers d'image en facilite la manipulation dans les opérations d'ouverture et de fermeture. En revanche, cette opération entraîne une perte d'informations et provoque une dégradation de la lisibilité pour les images numérisées avec une résolution trop faible. Cette opération est déconseillée pour une consultation finale à moins qu'elle soit présentée à titre d'option complémentaire.

3. Analyse de la structure physique des folios

L'analyse de la structure physique de l'image d'un document permet d'identifier le texte des images. Cette analyse est souvent utilisée pour faciliter le repérage automatique des éléments de décoration ou d'illustration figurant sur un document. Le corpus des archives de Châtillon possède un nombre limité d'éléments graphiques et cette application de la segmentation nous sera peu utile. En revanche, à partir de ce principe de segmentation fondé sur le repérage des lignes propres aux éléments textuels, le laboratoire LIRIS à l'INSA développe actuellement des outils prometteurs dans l'aide à la transcription et à la compression des images⁴⁰. En séparant le recto et verso d'un document, les méthodes de segmentation utilisées permettent d'obtenir une image binaire, augmentant par là même la lisibilité du texte et autorisant une compression différenciée des différents éléments de l'image. En compressant peu le texte et en appliquant une forte compression en JPEG sur l'arrière plan, le poids de l'image se trouve fortement réduit.



Manuscrit numérisé en couleur



Segmentation LIRIS

⁴⁰ LE BOURGEOIS F., EMPTOZ H., TRINH E. *Compression et accessibilité aux images de documents numérisés*. Document numérique, 2003, vol. 7, n°3-4, pp. 103-125.

L'indexation des archives

1. Problématiques rencontrées dans les projets de numérisation des manuscrits anciens

Depuis le début des années 90, et après la réalisation de programmes de catalogues électroniques, de nombreux projets de numérisation de fonds patrimoniaux sont apparus aux Etats-Unis et en Europe⁴¹. Si l'émergence de l'Internet a permis d'envisager une consultation des documents à distance, la question centrale de ces différents projets repose sur les modalités d'accès aux documents dans leur intégralité et avec le souci de préserver leur authenticité. Cet accès aux documents, soulève deux types de questions que nous retrouvons dans tous les projets et nous renvoient aux interrogations concernant le traitement du corpus de Châtillon:

- Quelles métadonnées utiliser ? Doit-on s'orienter vers une DTD existante ou créer une DTD adaptée à la spécificité du corpus à décrire ? L'indexation doit-elle être ouverte ou contrainte ?
- Quels sont les outils les plus appropriés pour accéder aux documents ? Doit-on faire le choix d'une base de données relationnelles et/ou celui des normes SGML et XML ?

Pour illustrer ces différents aspects nous nous limiterons au projet européen MASTER⁴² qui traite des manuscrits anciens et qui est donc plus proche du corpus de Châtillon, alors qu'un projet comme DEBORA⁴³, similaire dans certaines de ses problématiques concerne uniquement les imprimés du XVI^e siècle.

Le projet MASTER, financé par l'union européenne a débuté en janvier 1999 et tente de définir, d'implanter une norme générale fondée sur XML pour la

⁴¹ Sur ces aspects nous pouvons souligner en France le rôle pilote de l'IRHT (Institut de recherche de l'Histoire des Textes) qui dès 1937, au moment de sa création, a entrepris le microfilmage de manuscrits anciens. Aujourd'hui les bases *Enluminures*, <http://www.enluminures.culture.fr/> et *Liber Floridus* <http://liberfloridus.cines.fr/> permettent d'accéder en ligne aussi bien aux images numérisées qu'à une description des manuscrits enluminés.

⁴² Manuscript Access through Standards for Electronic Records. <http://www.cta.dmu.ac.uk/projects/master/>

⁴³ Digital accEss to BOoks of the RenAissance. Pour une présentation complète du projet voir: DEBORA : projet européen no. LB 5608 A. Coordinateur R. Bouché, juin 2000. 179pages.

description des manuscrits en langue latine⁴⁴. Dans cette optique un groupe de travail international s'est constitué pour définir une DTD qui serait compatible avec les recommandations existantes de la TEI⁴⁵. La volonté commune de permettre une description complexe des manuscrits a semble-t-il suscité des débats importants sur l'orientation à envisager, entre le choix d'un système générique, souple, utilisant les normes internationales SGML, XML, et le souci de conserver une compatibilité avec les systèmes de bases de données. Comme réponse à cette interrogation, le choix s'est porté sur le maintien de deux systèmes de bases de données, l'une utilisant MARC et l'autre XML.

Une proposition de DTD MASTER inspirée de la TEI a vu le jour. Elle est constituée de six éléments :

- un élément <msIdentifier>, qui regroupe toute l'information nécessaire pour identifier de façon unique le manuscrit décrit à l'intérieur du dépôt où il est conservé;
- un élément facultatif <msHeading>, qui donne un titre fourni ou une brève entête pour le manuscrit décrit ;
- un élément facultatif <msSummary> qui fournit une notice minimale du manuscrit dans un format structuré spécifique; il peut être utilisé en remplacement d'une notice plus détaillée;
- un élément facultatif <msContents>, fournissant une liste de composants du contenu intellectuel d'un manuscrit ou d'une partie de manuscrit avec les descriptions bibliographiques et tout ce qui s'y rapporte;
- un élément facultatif <physDesc>, qui regroupe des informations sur tous les aspects matériels d'un manuscrit ou d'une partie de manuscrit (tels que son format, matériau, styles d'écriture utilisés, foliotation et hors-texte, reliure etc.);

⁴⁴ « Le projet est étroitement lié aux projets nord-américains sur le manuscrit EAMMS (Electronic Access to Medieval Manuscripts) et Digital Scriptorium, ainsi qu'avec la Text Encoding Initiative (TEI), entreprise internationale ». <http://www.tei-c.org/Master/Hermes/hermesFR.html>

⁴⁵ TEI (Text Encoding Initiative) est une DTD pour indexer et décrire un document textuel. Destiné au départ pour les sciences humaines, c'est un ensemble de balises basées sur le langage XML.

- un élément facultatif <history> , qui récapitule l'histoire connue d'un manuscrit ou d'une partie de manuscrit, et dans lequel origine et provenance sont distingués;
- un élément facultatif <additional> qui regroupe d'autres informations relatives au manuscrit, en particulier les informations administratives concernant son emplacement actuel, les pièces annexes qui lui sont associées, les reproductions disponibles etc.;
- un ou plusieurs éléments facultatifs <msPart> contenant chacun une description structurée pour chaque fragment de manuscrit, considéré maintenant comme partie intégrante du manuscrit décrit, bien que physiquement distincte à l'origine.

Nous pouvons constater que si cette DTD a été conçue pour répondre aux besoins d'indexation et d'accès aux manuscrits, elle ne correspond qu'imparfaitement à la spécificité du corpus de Châtillon, notamment en ce qui concerne l'organisation propre des documents d'archives au sein d'un inventaire. Dans le chapitre qui suit, nous examinerons le standard de DTD EAD qui semble plus proche de la spécificité du corpus de Châtillon.

2. L'EAD, une DTD pour la description des fonds d'archives

Toute indexation d'un fonds d'archives doit permettre aussi bien la description de la structure des archives que celle de leurs contenus. Pour répondre à cette exigence descriptive, les archivistes américains ont, au début des années quatre-vingt, élaboré le format MARC AMC (ARCHIVAL and Manuscript Control) spécifique aux archives et manuscrits. Il s'inspire du format MARC (pour MACHine Readable Cataloguing) utilisé traditionnellement dans les bibliothèques pour la description des documents dans un catalogue. Ce format placé sous le signe de l'échange de l'information bibliographique a donné naissance à une multitude

de formats nationaux⁴⁶ et l'établissement en 1977 de la norme internationale UNIMARC adapté au catalogage des livres anciens. Mais, si ce format MARC est adapté au recensement d'un fonds d'archives, en revanche il interdit le repérage d'un document particulier à l'intérieur de ce fonds. C'est en raison de cette impossibilité qu'a été élaboré l'EAD (Encoding Archival Description) avec une première version en 1996.

L'EAD est une DTD (Description Type Document) qui permet de rendre compte aussi bien de la structure d'un fonds d'archives que de son contenu. Cette DTD se présente sous la forme de balises qui ont été écrites pour fonctionner avec le SGML (Standard Generalized Markup Language)⁴⁷ tout en étant entièrement compatible avec XML.

Dans la structure de l'EAD, et dans une perspective d'intégration éventuelle pour le corpus de Châtillon, nous retenons particulièrement le deuxième segment appelé *description archivistique* (<archdesc>) qui permet la description de tous les niveaux hiérarchiques d'un fonds d'archives sans redondance d'informations ; chaque niveau subordonné hérite des informations présentes au niveau supérieur. Autre élément qui permet d'envisager son utilisation : l'EAD permet un encodage de profondeur variable, progressif et donc dynamique aussi bien sur la structure du fonds d'archives que sur son contenu⁴⁸.

En revanche, l'adaptation de cette DTD au corpus des archives de Châtillon nécessiterait un investissement particulier qui pourrait faire l'objet d'une mission à part entière. Malgré les limites en temps de notre intervention, nous préconisons

⁴⁶ Ainsi, en Grande Bretagne, les services de la British National Bibliography établirent en 1968 leur propre format MARC qui se voulait compatible avec le format américain. En France, la bibliothèque universitaire de Grenoble définit le format MONOCLE en 1970. Dans les années 1970, on vit se multiplier les formats MARC nationaux, adaptés aux pratiques de catalogage des différents pays : à côté de USMARC et de UKMARC, les pionniers, citons AUSMARC en Australie, CANMARC au Canada, DANMARC au Danemark, INTERMARC qui se voulait au départ un format européen, etc

LERESCHE Françoise, Bibliothèque Nationale de France. UA-Bibliotheken [en ligne].

Disponible sur <http://lib.ua.ac.be/KB/fl/leresc0.html> (consulté le 30/09/2004)

⁴⁷ « Une métalangue normalisée, ou syntaxe, conçue pour la spécification d'un nombre illimité de langages de marquage. Un document en langage SGML comporte trois éléments : la déclaration (décrit le milieu de traitement requis); la définition du type de document (DTD) (un renvoi défini qui forme un gabarit pour la description de la structure et du contenu d'un type particulier de document); et la suite de documents elle-même ».

CHARTRON Ghislaine. Standards – Normes - Documents numériques, Introduction générale. [en ligne]. Disponible sur : <http://web.ccr.jussieu.fr/urfist/presse/standard/coursintro.htm> (consulté le 30/09/2004).

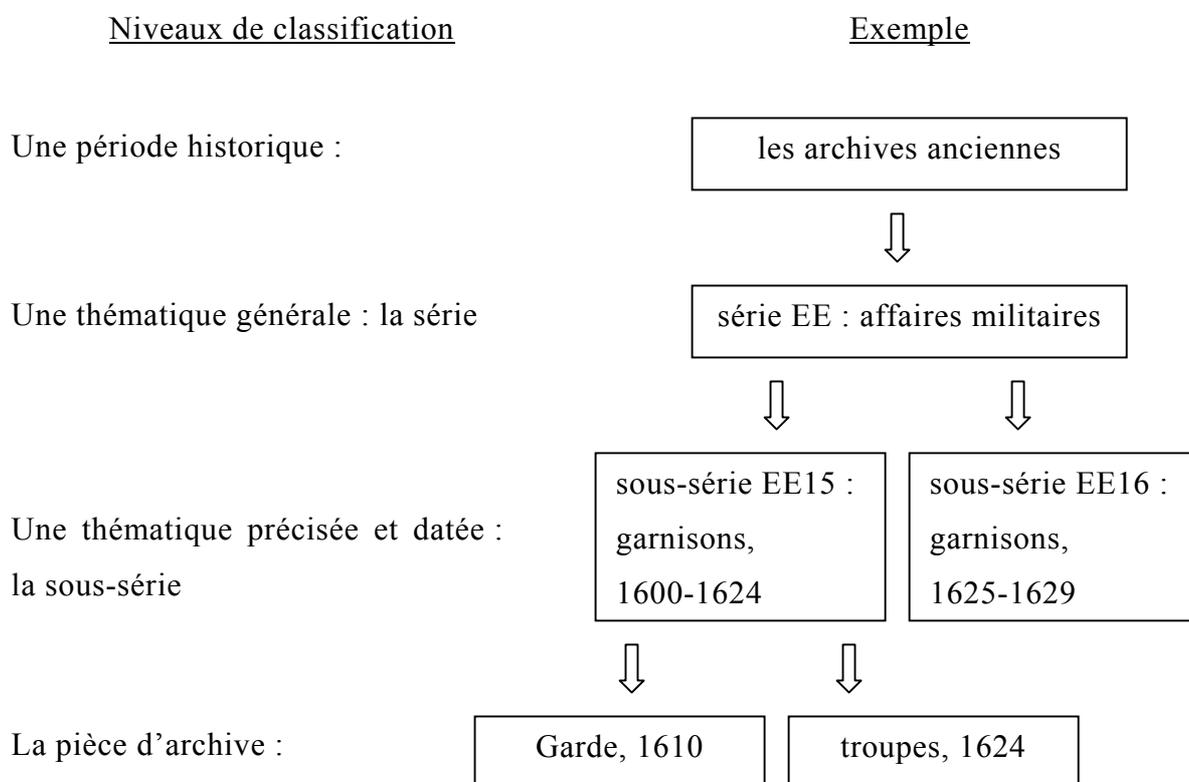
⁴⁸ Pour une présentation plus complète de l'EAD voir :

QUEYROUX Fabienne. *EAD, la description archivistique encodée*. In: WESTEEL I, AUBRY M. Coord. La numérisation des textes et des images: techniques et réalisations. Villeneuve d'Ascq : Ed. du conseil scientifique de l'Université Charles de Gaulle-Lille 3, 2003, p. 75-89.

néanmoins de suivre attentivement l'évolution de ce standard⁴⁹ tout en proposant d'ores et déjà des modalités de description et d'indexation des archives. Les données de cette indexation devront être organisées et stockées de façon à permettre aisément leur intégration éventuelle dans une structure EAD.

3. Organisation des archives et métadonnées

Comme nous l'avons vu précédemment les archives sont organisées selon une classification hiérarchique thématique, chronologique et selon le schéma suivant:



⁴⁹ A la date du 22 avril 2004, les Archives de France recensent 19 applications françaises en EAD, dont 10 concernent des archives départementales. Dans cet inventaire des applications en cours, nous pouvons noter la présence de la Bibliothèque municipale de Lyon pour l'encodage des instruments de recherche des fonds d'archives privées. Il pourrait être instructif, dans une phase ultérieure du programme de numérisation des archives de Châtillon, de prendre contact avec cet établissement et d'examiner en détail les choix de balises effectués.

Dans la mise en place du standard EAD, nous attendons le *Dictionnaire des éléments de l'EAD, version 2002*. Traduction française de la version anglaise réalisée par le groupe AFNOR CG46/CN357/GE3, avec une mise en ligne qui est prévue en octobre 2004.

ARCHIVES DE FRANCE. *Archives de France*. [en ligne]. Disponible sur : <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/referentielEAD.html> (consulté le 30/09/2004).

Le rôle de l'indexation par les métadonnées est de renseigner chaque niveau de cette organisation en vue de servir les différents usages envisagés.

3.1. Métadonnées et la notion de mot-clé métalinguistique

La notion de métadonnées est indissociable d'un traitement électronique de corpus textuels; elle est propre à la méthodologie d'exploitation automatique de bases de données et se prête à des usages diversifiés correspondant pour l'essentiel à la notion de *mot-clé métalinguistique* telle qu'elle a été mise en œuvre par T.R. Wooldridge et I.Turcan⁵⁰ : un mot-clé thématique ou un mot-clé métalinguistique identifiés comme tels dans un corpus textuel donné constituent, lors du traitement automatique du même corpus converti en base de données, un *mot outil*, une *métadonnée* permettant d'accéder automatiquement au contenu souhaité. Les métadonnées ne sauraient être déterminées de façon automatique et arbitraire dans des bases de données de textes anciens, comme c'est en particulier le cas pour les dictionnaires imprimés sous l'Ancien Régime ; elles doivent être définies pour chaque corpus dans le souci du respect de son identité⁵¹. A cet égard, elles peuvent fonctionner en complémentarité avec le thésaurus dont le principe de hiérarchisation arborescente est porteur de signification.

Le principe de classification hiérarchique, thématique et chronologique, nous servant de référence dans l'organisation des archives comme dans les modalités d'accès numériques envisagées, il est nécessaire d'en renseigner les différents niveaux.

3.2. D'une indexation de base au contenu

Compte tenu de l'ampleur du fonds d'archive et de ses caractéristiques linguistiques diversifiées, il est raisonnable d'envisager un accès évolutif aux archives. A partir de l'indexation de base réalisée par les archivistes, il est possible de fournir un premier accès aux images des archives en direction d'un public de spécialistes en sciences humaines et sociales (linguistique, histoire,...).⁵² En

⁵⁰ *Les mots-clés métalinguistiques comme outil d'interrogation structurante des dictionnaires anciens[...]*, in Actes des quatrième journées scientifiques (Lyon, septembre 1995), *Lexicomatique et dictionnaire*, publiés sous la direction d'A. Clas, P. Thoiron, H. Béjoint, Beyrouth, 1996, pp. 317-316.

⁵¹ Définition transmise par Isabelle Turcan (texte à paraître sur le site du laboratoire ATILF).

⁵² Cette possibilité d'accès peut aussi générer la mise en place d'un travail collaboratif de transcription, traduction et d'analyse de contenu.

revanche, cette indexation de base s'avère insuffisante si l'on souhaite, comme c'est le cas, ouvrir les archives à un public plus étendu et hétérogène dans ses compétences. Pour ce public, il est préférable d'aborder les archives par une approche thématique sur des éléments de la vie quotidienne (bâti, événements marquants, métiers) en donnant accès à des éléments de contexte (chronologie universelle, lexiques, dictionnaires). Si cette possibilité d'accès se mettra en place progressivement, au rythme du traitement des contenus, il est nécessaire d'en fixer le cadre et les champs d'indexation dès le départ.

3.3. La pièce d'archive comme unité de base d'indexation

En raison de son unité matérielle et de sa thématique unitaire, c'est la pièce d'archive qui constituera le niveau de base d'indexation dans l'arborescence. Les métadonnées retenues doivent décrire plusieurs aspects de la pièce d'archive :

- Un identifiant unique
- Son positionnement au sein du fonds d'archives (période historique, série, sous-série)
- Ses caractéristiques physiques et matérielles (dimension, support, conservation)
- Sa forme (feuille, cahier, plan)
- L'état d'avancement dans le traitement de l'archive (transcription, traduction, indexation du contenu)
- Sa date de création
- Les langues utilisées (latin, ancien français, moyen français, parlers franco-provençaux, français de référence,...)

Mais aussi des descripteurs qui identifient dans la pièce des thématiques plus spécifiques : par exemple,

- Les propriétés linguistiques (niveau de langue, termes techniques, noms propres)
- Les événements locaux
- Les éléments du bâti de Châtillon

- Les métiers⁵³

Nous avons vu que les informations correspondant aux premiers éléments ont été enregistrées au cours du travail de numérisation (Cf. *supra* 2.8), alors que les seconds plus spécifiques doivent l'être au moment de la transcription et traduction. Pour ce deuxième aspect nous verrons tout l'intérêt de l'utilisation d'un thésaurus qui peut permettre entre autres un usage partagé des descripteurs pour des transpositeurs/traducteurs multiples.

La notice présentée à la page suivante permet d'illustrer ce qui précède, sans correspondre forcément au contenu réel d'une pièce précise.

⁵³ L'examen des corporations et confréries au XVIII^e siècle nous permet de dresser une première liste des métiers en activité sur Châtillon (si l'on se range au principe de continuité historique depuis le XV^e, en étant conscient des risques de décalage): charpentiers, menuisiers, maçons, marchands tailleurs, chaussetiers et cabanaciers, cordiers, « tissiers en toile », cordonniers, tanneurs et corroyeurs, maréchaux et taillandiers, boulangers et cabaretiers, bouchers et chapeliers. Voir : FAVRE Adrien. *Histoire de Châtillon-sur-Chalaronne*. Ceyzériat : A. Favre, 1972, pp. 53.

Identifiant unique:	XX18_pièce002	
Période :	ancienne	
Série :	XX	
Sous-série :	XX18	
Dimension (hauteur x largeur, en cm) :	22,5 x 15	
Support :	papier	
Conservation :	Bon état général Folio 002_recto : taches Folio 005_verso: encre en transparence	
Forme :	Cahier cousu	
Traitement de l'archive :	Transcription, traduction, indexation de contenu	
Date de création :	04 mars 1634	
Langues :	Ancien français ; franco-provençal	
Propriétés linguistiques :	Niveau de langue	Langue parlée
	Termes techniques	Entrepiez, travon ⁵⁴
	Noms propres	Chastellion en dombes, évêque de Philippone, jehan de sayssel, marechal de savoye, Bourg
Événement local:	incendie de Châtillon	
Bâti concerné :	porte de l'Escarre, porte de Bourg, remparts	
Métiers :	maçon, charpentier	

Dans la perspective de constitution d'une base de données, le contenu de ces différents champs est interrogeable avec des possibilités de recherches croisées.

⁵⁴ Termes de charpente.

4. De l'intérêt d'un thésaurus

Le thésaurus⁵⁵ est un outil qui se présente sous la forme d'une liste de termes normalisés et organisée hiérarchiquement. Il est particulièrement adapté à un travail d'indexation quand plusieurs intervenants sont amenés à travailler sur un corpus commun. Dans le cas présent, il permet de répondre à plusieurs interrogations :

- Face à la multiplicité des mots qui peuvent désigner les mêmes objets, comment s'entendre sur un terme commun ?
- Comment éviter l'inflation des termes dans la description des documents d'archives ?

Dans un thésaurus, une notice donne la définition des termes retenus et leurs champs d'application. Les termes sémantiquement proches sont dits rejetés et renvoyés au terme de référence. Les termes sont organisés hiérarchiquement, des termes génériques aux plus spécifiques. La notion de *terme associé* permet d'établir une passerelle entre différentes thématiques du corpus.

Prenons un exemple précis sur le corpus des archives de Châtillon et le cas de nombreuses pièces qui sont indifféremment décrites sous les termes de « rempart », « fortification », « murailles » sans que l'on puisse en discerner précisément la différence. L'utilisation d'un thésaurus permettrait par exemple :

- De définir le terme de « fortification », de le retenir comme générique, de rejeter celui de « muraille » en le renvoyant sur celui de « fortification »
- De retenir le terme de « rempart » comme spécifique de « fortification » en se laissant la possibilité d'affiner dans la spécificité des usages du lexique (évolution inévitable depuis le XV^e).
- D'introduire des spécifiques au terme de « rempart » : « porte de Durlevant », « porte de Bourg », « porte de Villars ».
- De créer un lien avec des termes associés qui abordent le sujet mais à partir d'une approche différente: « gestion », « corvée ».

⁵⁵ Répertoire alphabétique de termes normalisés pour l'analyse de contenu et le classement des documents d'information. *Dictionnaire de la langue française*. Le Robert, 1993.

La création d'un thésaurus suppose une connaissance de l'éventail des thèmes abordés dans un corpus. Pour les archives de Châtillon, il est possible d'envisager que le classement des archives tel qu'il existe et les premiers travaux de transcription/traduction servent à un premier repérage des grandes thématiques. Si un thésaurus doit posséder une stabilité dans son arborescence de base, il peut évoluer dans l'introduction de spécificités au fur et à mesure de la connaissance du contenu des archives.

5. Des informations contextuelles complémentaires comme première phase de valorisation

Si l'opération d'indexation nous renseigne sur le contenu des documents, elle reste limitée dans la connaissance du contexte et pour l'approfondissement d'un sujet. Pour répondre à ces deux aspects et à partir des usages envisagés (par exemple, approches historique et linguistique), il est possible d'envisager l'accès à des éléments d'informations externes qui constituent une première phase envisageable de valorisation.

5.1. Une chronologie universelle et locale

C'est avec l'appui de repères chronologiques que l'examen d'un corpus d'archives, tel que celui de Châtillon étendu sur plus de trois siècles, peut livrer tout son intérêt. Si ces repères sont intégrés par des spécialistes historiens, ce n'est pas le cas des autres destinataires pressentis. Pour une localité comme Châtillon, située entre la Savoie et le Royaume de France, les décisions d'importance sont souvent tributaires du pouvoir de tutelle. Les archives de Châtillon rendent régulièrement compte des activités du conseil de syndic dont les délibérations et décisions sont liées à l'évolution du contexte extérieur⁵⁶ (par exemple : augmentation d'impôts, pénuries, réquisitions, corvées). Mais nous pouvons aussi constater une marge de négociation, de résistance dans l'application de directives. Or contrairement aux idées reçues d'une société livrée à elle-même, les archives marquent une volonté de régulation dans la gestion des conflits (réclamations, contrats, procès).

⁵⁶ Par exemple, l'entretien des remparts a été fluctuant et ceci en raison des menaces de guerres extérieures.

Pour appréhender ce contexte extérieur, il peut être utile de mettre à disposition du lecteur ou consultant d'archives, deux chronologies, l'une universelle⁵⁷ qui rendrait compte d'une histoire globale, l'autre plus locale qui aurait pour centre l'histoire ancienne de la Bresse et de la Dombes⁵⁸. A cet égard, l'idée d'informatiser comme document associé, l'ouvrage d'O. Morel qui concerne directement les comptes de syndics de Châtillon (1375-1500), constituerait un atout majeur du point de vue de la valorisation des manuscrits⁵⁹.

5.2. Des lexiques au *Dictionnaire Universel [...] de Trévoux* (1704-1771)

A l'image des événements historiques locaux qui trouvent leur sens dans un environnement plus global, la langue utilisée par une communauté peut être mise en relation avec l'état de la langue à un moment donné. Cet enregistrement de l'état de la langue a été le fait des grands dictionnaires à vocation nationale des XVI^e, XVII^e et XVIII^e siècles. Le *Dictionnaire de Trévoux*, né dans la souveraineté de Dombes (Trévoux 1704) fait partie de cet ensemble : ouvert sur le monde, il possède la particularité de ne pas puiser ses sources dans les seuls corpus littéraires et de s'être intéressé aux diverses formes de parlers régionaux aussi bien qu'aux termes techniques de l'artisanat⁶⁰.

5.2.1 Des manuscrits aux lexiques

Or le corpus des manuscrits de Châtillon, vierge jusqu'à présent de toute étude linguistique, présente des difficultés de compréhension, en particulier dans le lexique et la grammaire, et il nous importe de contribuer à clarifier ces textes pour les rendre accessibles au public des châillonnais, des habitants de l'Ain et plus

⁵⁷ Par chronologie universelle nous entendons des repères d'une histoire européenne, les contours d'un « territoire français » ayant évolué au gré des alliances entre puissances régionales et européennes.

Pour l'établissement de cette chronologie, l'ouvrage suivant pourra nous servir de référence : CENTRE NATIONAL DU LIVRE. *Journal de la France et des français*. Paris : Gallimard, 2001, 1060 p. (coll.Quarto)

⁵⁸ AUBRET Louis. *Mémoires pour servir à l'histoire des Dombes*. Trévoux : J.C. Damour, 1968, 4 vol.

GUICHENON Samuel. *Histoire généalogique de la roayle Maison de Savoye [...]*, Lyon, Barbier, 1960, 2 vol.in fol.

GUICHENON Samuel. *Histoire de Bresse et Bugey*. Lyon : Huguetau et Ravaud, 1650.

GUIGUE Marie-Claude. *Topographie historique du département de l'Ain*. Marseille : Lafitte, 1976, 518 p.

GUIGUE Marie-Claude. *Les voies antiques du Lyonnais, du Forez, du Beaujolais, de la Bresse, de la Dombes, du Bugey et de partie du Dauphiné déterminées par les hopitaux du Moyen-Age*, In Mémoires de la société littéraire, Historique et archéologique de Lyon, Lyon : A.Brun, 1877.

⁵⁹ Cependant l'ouvrage ayant été imprimé à Bourg (imprimerie du « courrier de l'Ain ») en 1921, une vérification s'impose concernant l'actuel propriétaire des droits.

largement au grand public en général de la France et de la francophonie. Le premier travail consistera à indexer les mots difficiles, par leur forme ou par leur sens, pour en dresser une liste qui sera présentée sous forme d'un menu déroulant dans l'interface destinée au public ; cette liste sera elle-même le point de départ d'un travail de plus longue haleine, l'élaboration d'un lexique assorti d'une grammaire pratique qui feront l'objet d'une base associée à celle des images des manuscrits. Ainsi, à chaque consultation d'un document d'archives, un consultant pourra à tout moment se référer aux pages de commentaires linguistiques.

Pour élaborer cette composante du travail de valorisation, nous disposons de deux catégories de sources, les études scientifiques contemporaines (souvent difficiles d'accès) et les dictionnaires anciens dont une partie importante a bénéficié d'éditions électroniques qui en facilitent la consultation.

5.2.2 Le corpus des dictionnaires anciens

Il nous offre une ressource documentaire d'informations très diversifiées, inscrite dans la diachronie de l'histoire de la langue et de l'histoire des mots, riche d'une nomenclature pré-encyclopédique, notamment depuis le fameux *Thresor de la langue françoise tant ancienne que moderne* de Jean Nicot (1606) en passant par le *Dictionnaire universel* de l'abbé Antoine Furetière (1690) et le *Dictionnaire étymologique ou Origines de la langue françoise* de Gilles Ménage, cela jusqu'aux différentes éditions du *Dictionnaire universel de Trévoux* (1704-1771), dont l'édition imprimée en 1743 bénéficie d'une version électronique consultable en plein texte⁶¹. Du fait de la proximité géographique Dombes-Bresse, c'est donc d'abord ce corpus lexicographique que nous privilégierons, d'autant qu'il est lui-même la somme des dictionnaires du XVIIe siècle. Ce n'est pas ici le lieu de développer cet aspect, mais on comprendra mieux notre démarche à la lumière d'exemples significatifs⁶².

⁶⁰ Cf. les Actes des journées internationales : *Connaissance et rayonnement du Dictionnaire de Trévoux (1704-1771)* (Trévoux, octobre 1999), en particulier, KIBBEE D. *Le Dictionnaire de Trévoux et la politique de la langue française*. Disponible sur <http://www.univ-lyon3.fr/siehdaweb/trevoux/Kibbee-T.ht> (consulté le 13 septembre 2004).

⁶¹ *Supra* note 8.

⁶² Nous remercions ici Isabelle TURCAN de nous avoir transmis le résultat de recherches qui n'ont pas encore été publiées.

5.2.3 Des manuscrits au *Dictionnaire de Trévoux*

5.2.3.1 *Un dictionnaire de la Dombes qui s'intéresse à la Bresse*

Prenons le mot *panage* présent dans les manuscrits de Châtillon : ce n'est pas un mot du quotidien des français du début du XXI^e siècle, mais il était vivant encore au XVIII^e siècle comme en témoigne le *Dictionnaire de Trévoux* où le mot *panage* apparaît comme propre à la Bresse :

« *Panage*. C'est en Bresse une quantité de blé ; par exemple, de 20 mesures, que le granger ou le métayer retient par le contrat de grangeage, à prendre sur les blés avant que de les partager, pour nourrir les moissonneurs ou batteurs. De Laurière.

Ce mot vient de *panis*, pain. Voyez Rével, sur les Statuts de Bresse. »
[1743]

Nous percevons par la définition qui en est donnée que l'usage du mot est propre aux métiers de la campagne, et certainement ancré de longue date sur ce territoire ; on sait par ailleurs que le droit coutumier des provinces de France est particulièrement conservateur d'usages anciens ; mais si nous consultons un dictionnaire postérieur, tel celui d'E. Littré au XIX^e s., nous nous trouvons confrontés à une définition différente qui ne comporte aucune précision relative aux anciennes coutumes de Bresse :

« PANAGE (pa-na-j'), s. m. Droit qui se paye au seigneur d'une forêt, pour avoir la liberté d'y faire paître les porcs. » [Litt.]

On observera le même cas de figure pour le droit féodal ce qui montre d'ailleurs contrairement aux idées reçues qu'il n'a pas fallu attendre le XIX^e siècle pour redécouvrir la civilisation médiévale.

5.2.3.2 *Un dictionnaire du XVIII^e siècle qui s'intéresse au droit féodal et au droit coutumier*

Prenons la forme *pontanage* présente dans les manuscrits de Châtillon et absente telle que dans les bases de données lexicographiques consultées ; la proximité formelle avec l'entrée *pontenage* ou *pontonage* dans le *Dictionnaire de Trévoux* nous permet, en vertu du contexte dans lequel la forme *pontanage* a été employée, de nous référer à la définition du dictionnaire :

« PONTENAGE, ou PONTONAGE. s. m. Est un droit que le Seigneur féodal prend sur les marchandises qui passent sur les rivières, sur les bacs & les ponts, qu'on a appelé en la basse Latinité, *pontaticum*, *pontagium*, & *pontonagium*. » [1743]

On trouve confirmation de cette définition grâce à une interrogation en plein texte du même dictionnaire qui nous donne accès à un alinéa de l'article BAC :

« Le droit de bac est un droit seigneurial qui s'affirme : ce qu'on appelle en quelques lieux *Pontenage*, ou *Pontonage*. Ce droit se lève sur ceux qui passent une rivière dans le bac, ou bateau du Seigneur, qui seul a le droit d'en avoir pour faire passer l'eau à ceux qui le souhaitent. Ce qui s'appelle passer le port, & se dit tant de celui qui conduit le bateau, que de ceux qu'il conduit à l'autre bord. Ce droit s'affirme à un Batelier, qui seul en jouit, & qui s'appelle *Pontonier*. » [1743]

Ces deux exemples nous paraissant suffisants pour montrer l'intérêt d'une consultation du *Dictionnaire de Trévoux* pour mieux cerner les usages lexicaux dans les manuscrits de Châtillon, il est désormais temps d'en dégager le modèle du processus de valorisation que nous envisageons de mettre en œuvre.

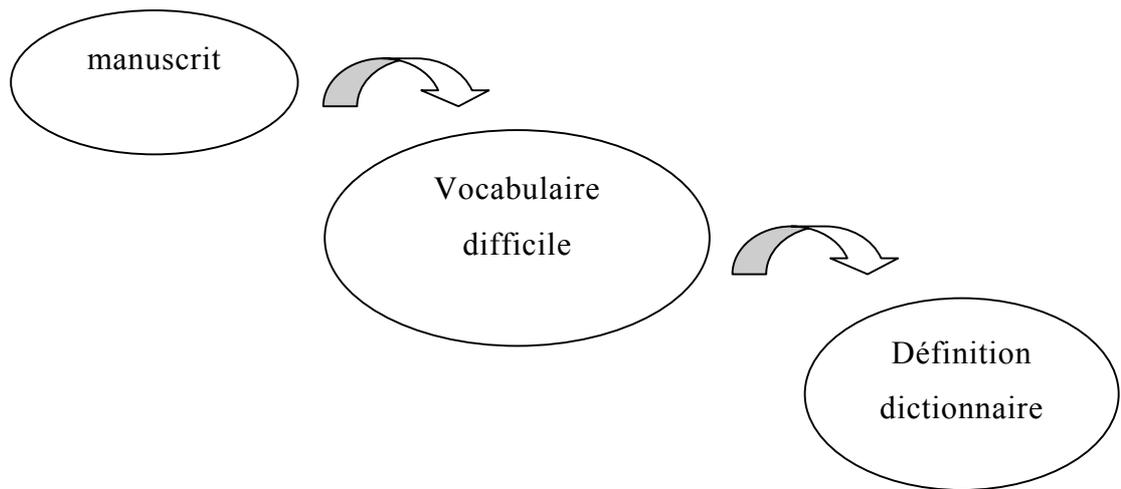
5.3. Représentation du modèle de valorisation des archives à l'aide du *Dictionnaire de Trévoux*

Dans les manuscrits de Châtillon nous rencontrons le mot *vinage* qui est inconnu à la plupart de nos contemporains. Si nous prenons la définition de mot dans le *Dictionnaire de Trévoux*, nous remarquons que trois usages de ce mot sont indiqués :

- Un usage technique juridique relevant du droit féodal
- Des usages divers dont une acception propre au droit coutumier
- Un usage plus généralisé

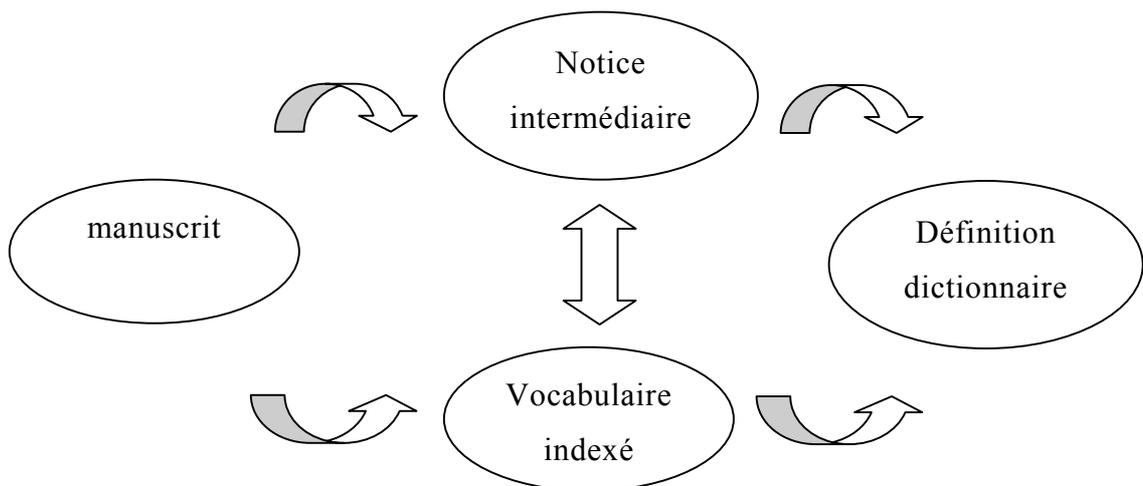
« VINAGE. s. m. Terme de Coutume. *Jus vinarium*. C'est un droit seigneurial qui est dû en plusieurs lieux sur les vignes, au lieu de censives, qui se doit payer à bord de cuve : c'est-à-dire, avant qu'on puisse tirer le vin de la cuve, comme le champart se paye avant que d'enlever les gerbes [...] Ce mot de *Vinage* a encore différens autres usages & différentes significations... Souvent *Vinage* se prend pour les droits qui se payent aux Seigneurs par des Communautés & territoires en bled, vin ou argent ; en conséquence de quoi les Seigneurs font réparer les ponts & passages... Enfin, quelquefois *Vinage* se prend pour réjouissance & bonne chère.»
[1743]

En revanche, le consultant des manuscrits de Châtillon pourra rencontrer des difficultés dans le choix entre ces trois définitions. Seule une connaissance du contexte des manuscrits peut éventuellement permettre de choisir le bon terme sans risque de contresens. Pour répondre à cette difficulté, une notice explicative intermédiaire doit être intercalée entre le manuscrit et le *Dictionnaire de Trévoux*. Les deux schémas qui suivent vont nous permettre de préciser l'intérêt de cette notice intermédiaire dans le processus de valorisation des archives de Châtillon visant à les rendre accessibles à tout public :



Modèle classique d'appel de définitions

Les mots difficiles des manuscrits sont indexés et un lien les relie aux articles du dictionnaire.



Modèle avec notice intermédiaire

A la différence du modèle précédent et quand cela est nécessaire, une notice intermédiaire permet d'expliquer et de retenir dans la définition du dictionnaire l'usage approprié à un contexte donné des manuscrits.

Nous voyons tout l'intérêt de la numérisation des archives de Châtillon et la mise en relation de ce corpus avec le *Dictionnaire de Trévoux*. Par cette mise en relation et l'explication du contexte, le contenu des archives de Châtillon se trouve valorisé et accessible à un grand nombre de consultants. Le *Dictionnaire de Trévoux*

trouve aussi dans le corpus des archives de Châtillon une base documentaire importante susceptible d'éclairer les usages diversifiés et quotidiens de certaines composantes de la langue française.

L'interface de consultation des archives

La description de la chaîne de numérisation avec la normalisation des procédures, le cadre d'indexation, la modélisation de la valorisation, trouve sa traduction concrète dans la mise en place d'une interface de consultation. Cette interface reprend les différents principes décrits au cours de ce rapport et insiste notamment sur l'exigence de double accès dans la consultation: en direction du grand public, en direction des chercheurs en sciences humaines et sociales.

1. La création d'un Cédérom de démonstration

A l'occasion du concours des *prix rhônalpins du patrimoine 2004*⁶³, la mairie de Châtillon-sur-Chalaronne présente un dossier qui comportera : l'ouvrage élaboré par l'association Saint-Guignefort⁶⁴, la présentation des travaux de Caroline Sarret⁶⁵, un cédérom de démonstration élaboré par nous-même sur les archives de Châtillon illustrant différentes possibilités de consultation des archives.

Dans ce cédérom nous proposerons :

- la présentation des archives, leur importance dans le patrimoine, le travail de numérisation, les intervenants du programme
- la consultation d'archives numérisées dont certaines sont transcrites, traduites, indexées
- la mise en relation des mots difficiles du manuscrit avec leur définition dans le *Dictionnaire de Trévoux*

Dans cette optique, nous avons réalisé une interface d'accueil qui s'appuie sur différentes modalités d'accès aux archives :

⁶³ Ce concours est organisée par *Patrimoine Rhônalpin* pour la dixième fois en 2004, pour récompenser des opérations de restauration et de valorisation du patrimoine. Il est doté d'une subvention par la région Rhône-Alpes et par EDF en Rhône-Alpes.

⁶⁴ *Supra* note 16.

⁶⁵ *Supra* note 17.



Notre proposition d'interface de la page d'accueil

Cette interface en permettant différentes modalités d'accès aux archives vise à répondre aux attentes du grand public comme celui des chercheurs.

Elle propose :

- un accès grand public qui permet d'atteindre directement les archives, à partir d'une sélection de thèmes liés au bâti de Châtillon, *plan de Châtillon*, ou par l'intermédiaire d'une nouvelle page contenant une sélection de thèmes, *les archives par thèmes*.
- un accès pour les chercheurs par menu déroulant, *Accès direct aux archives*, ou par l'intermédiaire de l'inventaire, *les archives par l'inventaire*.

2. Les développements envisagés

La réalisation de ce cédérom n'est qu'une première étape de valorisation concrète des archives de Châtillon. Ce cédérom doit permettre de porter à la connaissance d'un plus grand nombre la valeur patrimoniale inestimable de ce corpus et de

démontrer tout l'intérêt d'un développement du programme de numérisation à plus grande échelle. Dans une deuxième phase du programme, ce corpus doit devenir accessible sur le Web en utilisant les principes de navigation testés dans le cédérom. La technologie utilisée, le langage d'affichage HTML, permet d'ores et déjà d'envisager cette intégration dans une consultation en ligne. Pour favoriser une exploitation optimale des contenus d'indexation, nous suggérons l'intégration d'une base de données dont le modèle de champ a été décrit précédemment (Cf l'indexation des archives, chap 1.3.). Dans cette optique nous envisageons l'utilisation de deux applications libres et gratuites, MYSQL pour la base de données et PHP comme langage permettant son interrogation. Ces applications ont été conçues dès le départ pour le Web⁶⁶ et s'intègrent parfaitement au langage d'affichage HTML. Elles bénéficient de l'apport d'une large communauté d'utilisateurs/développeurs qui garantit l'amélioration constante de leur fonctionnalité et l'assurance de leur pérennité. De même ces outils permettent une mise en ligne progressive et évolutive du corpus des archives de Châtillon. Dans un premier temps, les images d'archives numérisées dans leur totalité peuvent être rendues accessibles en ligne. Puis au rythme progressif des opérations de transcription, traduction, indexation, renvois lexicographiques, – plus lourdes à mettre en œuvre – le corpus des archives de Châtillon deviendra accessible au plus grand nombre pour une multitude d'usages.

⁶⁶ En 1994, première version de PHP et du système de gestion de bases de données MYSQL.

Conclusion

Le corpus des archives historiques de Châtillon – Chastillion les Dombes – est unique à plus d'un titre. A notre connaissance, aucune localité ne possède un fonds documentaire d'une telle richesse et vierge de toute étude systématique récente. De plus, la proximité géographique de la localité de Trévoux, – ancienne principauté de la souveraineté de Dombes – permet de mettre en relation et d'élargir la compréhension réciproque de ces deux ensembles patrimoniaux d'une valeur culturelle et scientifique inestimable : *les archives historiques de Châtillon-sur-Chalaronne* et le *Dictionnaire universel de Trévoux*. Dans l'étude de cette ensemble, le programme de numérisation engagé offre des possibilités nouvelles. Il va pouvoir rendre accessible ce corpus à la connaissance des chercheurs et favoriser des études diversifiées en sciences humaines et sociales. Il va offrir au grand public et plus spécialement aux habitants de Châtillon, une connaissance approfondie de la vie quotidienne de leur cité à travers les siècles.

Par notre intervention et ce que nous en présentons dans ce rapport, nous avons tenté d'apporter notre contribution à ce programme de numérisation en nous appuyant sur la connaissance que nous pouvons avoir des documents et bibliothèques numériques, tout en tenant compte de la spécificité du corpus des archives de Châtillon. Une équipe motivée et pluridisciplinaire s'est désormais constituée autour de ce programme et c'est avec passion et professionnalisme que nous envisageons la poursuite de cette collaboration.

Bibliographie récapitulative

OUVRAGES

AUBRET Louis. *Mémoires pour servir à l'histoire des Dombes*. Trévoux : J.C. Damour, 1968, 4 vol.

CENTRE NATIONAL DU LIVRE. *Journal de la France et des français*. Paris : Gallimard, 2001, 1060 p. (Quarto)

FAVRE Adrien. *Histoire de Châtillon-sur-Chalaronne*. Ceyzériat : A. Favre, 1972, pp. 53.

GALLIMARD FLAVIGNY Bertrand. *Etre bibliophile, petit guide pratique*. Paris : Séguier, 2004, 235 p.

GUICHENON Samuel. *Histoire généalogique de la roayle Maison de Savoye [...]*, Lyon, Barbier, 1960, 2 vol.in fol.

GUICHENON Samuel. *Histoire généalogique de la roayle Maison de Savoye [...]*, Lyon, Barbier, 1960, 2 vol.in fol.

GUIGUE Marie-Claude. *Topographie historique du département de l'Ain*. Marseille : Lafitte, 1976, 518 p.

GUIGUE Marie-Claude. *Les voies antiques du Lyonnais, du Forez, du Beaujolais, de la Bresse, de la Dombes, du Bugey et de partie du Dauphiné déterminées par les hopitaux du Moyen-Age*, In Mémoires de la société littéraire, Historique et archéologique de Lyon, Lyon : A.Brun, 1877.

MOREL M. Octave. *Une petite ville forte en Bresse sous la première domination savoyarde; la vie à Châtillon-en-Dombes d'après les comptes de syndics 1375-1500*. Bourg en Bresse : Imprimerie « Courrier de l'Ain », 1925, pp. 50-64.

QUEYROUX Fabienne. *EAD, la description archivistique encodée*. **In**: WESTEEL I, AUBRY M. Coord. *La numérisation des textes et des images: techniques et réalisations*. Villeneuve d'Ascq : Ed. du conseil scientifique de l'Université Charles de Gaulle-Lille 3, 2003, p. 75-89.

CEDEROMS

TURCAN Isabelle. *Introduction au Grand atelier Historique de la langue française*, Cdrom publié chez Redon en 2001, diffusion Le Robert.

CONFERENCES ET COMMUNICATIONS

Les mots-clés métalinguistiques comme outil d'interrogation structurante des dictionnaires anciens[...], in Actes des quatrièmes journées scientifiques (Lyon, septembre 1995), *Lexicomatique et dictionnaire*, publiés sous la direction d'A. Clas, P. Thoiron, H. Béjoint, Beyrouth, 1996, pp. 317-316

TURCAN Isabelle, LE BOURGEOIS Frank. *La numérisation des archives historiques de la municipalité de Châtillon-sur-Chalaronne dans l'Ain*. 2004. La Rochelle.

TURCAN Isabelle. *Numérisation et édition électronique de dictionnaires anciens : éthique et pratique pour le respect des livres et de leurs consultants*, in Actes du colloque *Ethique numérique. Défis ou opportunités pour la compréhension des situations et des contextes de communication*. (Saint Cyr-sur-Mer, mai 2003), pp. 119-125.

CONFERENCES ET COMMUNICATIONS ELECTRONIQUES

KIBBEE D. *Le Dictionnaire de Trévoux et la politique de la langue française* In Actes des journées internationales : connaissance et rayonnement du Dictionnaire de Trévoux (1704-1771) (Trévoux, octobre 1999). Disponible sur <<http://www.univ-lyon3.fr/siehdaweb/trevoux/Kibbee-T.ht>> (consulté le 13 septembre 2004).

Numérisation et valorisation des collections d'ouvrages anciens du laboratoire ATILF, CNRS, UMR 7118, Nancy 2 in *Les institutions culturelles et la numérique/ Cultural institutions and digital technology*, Ecole du Louvre ICHIM Paris 2003 (actes sur Cdrom et sur <www.ichim.org>).

PERIODIQUES

LE BOURGEOIS F., EMPTOZ H., TRINH E. *Compression et accessibilité aux images de documents numérisés*. Document numérique, 2003, vol. 7, n°3-4, pp. 103-125.

ROUMIEUX olivier. *Scanners patrimoniaux, des matériels dédiés*. Guide pratique Archimag, 2004, Hors-série les scanners professionnels, pp. 74-90.

RAPPORTS

DEBORA : *projet européen no. LB 5608 A*. Coordinateur R. Bouché, juin 2000, 179 pages.

SITES INTERNET

CHARTRON Ghislaine. *Standards-Normes-Documents numériques, Introduction générale* [en ligne]. Disponible sur :

<<http://web.ccr.jussieu.fr/urfist/presse/standard/coursintro.htm>> (consulté le 30/09/2004).

CINES. *Liber Floridus* [en ligne]. Disponible sur : <<http://liberfloridus.cines.fr/>> (consulté le 30/09/2004)

LERESCHE Françoise, *UA-Bibliotheken* [en ligne]. Disponible sur <<http://lib.ua.ac.be/KB/fl/leresc0.html>> (consulté le 30/09/2004)

CENTRE FOR TECHNOLOGY AND THE ARTS, DE MONTFORT UNIVERSITY. *MASTER* [en ligne].

Disponiblesur :<<http://www.cta.dmu.ac.uk/projects/master/>> (consulté le 30/09/2004)

MINISTERE DE LA CULTURE *Enluminures* [en ligne] Disponible sur : <<http://www.enluminures.culture.fr/>> (consulté le 30/09/2004)