

Diplôme national de master

Domaine - sciences humaines et sociales

Mention - sciences de l'information et des bibliothèques

Spécialité - gestion et traitement de l'information spécialisée

Sélection de documents du Département Sciences et techniques de la BnF en vue de leur intégration dans Gallica 2

Alina CANTAU

Sous la direction de Madame Isabelle FORMONT

Chef du service Conservation, Département Sciences et techniques
Bibliothèque nationale de France

Et sous la direction de Monsieur Jean-Paul METZGER

Directeur des études
Professeur des Universités à l'ENSSIB

Remerciement

Je remercie tout particulièrement Madame Isabelle FORMONT, chef du Service Conservation du Département Sciences et techniques de la Bibliothèque nationale de France, pour ses conseils précieux tout au long du stage et pour le temps qu'elle m'a consacré.

Mes remerciements à Monsieur Jean-Paul Metzger, mon directeur de mémoire, qui a accepté de me guider dans ce travail.

Je remercie enfin toutes les personnes du Département Sciences et techniques de la BnF pour leur accueil chaleureux et de leur disponibilité.

Résumé :

Engagée dans la numérisation de masse, afin d'accomplir sa mission pour la future Bibliothèque numérique Européenne, la BnF mène une politique rigoureuse de sélection d'ouvrages selon des contraintes documentaires, juridiques et physiques. Ces contraintes auront-elles un impact sur l'organisation cohérente des connaissances au sein de corpus représentatifs ? Quel est le rapport entre ce que l'on s'attend à trouver en ligne et ce que l'on trouvera réellement ?

Descripteurs : Gallica, Bibliothèque numérique Européenne, Europeana, sélection, corpus, numérisation, sciences médicales, collections patrimoniales.

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

Abstract :

Engaged in a project of mass digital transformation, in order to create the future European Digital Library, the National Library of France (BnF) is employing a rigorous book selection process based on documentary, legal and physical standards. However, will these factors have an impact on the coherent organization of the resources included in the digital library? What is the relationship between what we would expect to find on the future website and what we really find?

Keywords : Gallica, The European Digital Library, Europeana, selection, corpus, digitalisation, medical sciences, patrimonial collection.

Sommaire

INTRODUCTION.....	8
PARTIE 1 : PATRIMOINE NATIONAL EN LIGNE : OFFRE, EVOLUTION ET ENJEUX	9
1. GALLICA – BIBLIOTHEQUE NUMERIQUE DE LA BNF.....	10
1.1. <i>Politique documentaire.....</i>	<i>10</i>
1.1.1. Critères juridiques	10
1.2. <i>Le fonds de Gallica.....</i>	<i>11</i>
1.2.1. Les imprimés (monographies et périodiques)	11
1.2.2. Images, sons et presse	11
1.2.3. Politique de numérisation partagée	12
1.3. <i>L'existant en ligne pour le Département Sciences et techniques.....</i>	<i>13</i>
2. EUROPEANA – CONTRIBUTION FRANÇAISE A LA BNUE	14
2.1. <i>Historique</i>	<i>14</i>
2.2. <i>Autour du prototype Europeana</i>	<i>16</i>
2.2.1. Politique documentaire d'Europeana.....	17
2.2.2. Informations techniques et documentaires sur Europeana	18
2.3. <i>Enjeux d'une bibliothèque numérique européenne.....</i>	<i>19</i>
3. GALLICA 2 – VECTEUR D'EVOLUTION DE GALLICA.....	19
3.1. <i>Politique documentaire.....</i>	<i>20</i>
3.1.1. Le patrimoine national.....	20
3.1.2. Un corpus à vocation internationale	20
3.1.3. Un corpus français sur l'Europe.....	21
3.1.4. Les plus-values d'une offre d'ouvrages sous droits.....	21
3.2. <i>Politique logistique et budgétaire de la numérisation</i>	<i>22</i>
3.2.1. Marchés de dématérialisation.....	22
3.2.2. Filières de production	23
3.2.3. Choix techniques de numérisation.....	24
3.2.4. Chaîne de traitement et de numérisation.....	24
3.2.5. Aspects économiques des programmes de numérisation	26
3.3. <i>Calendrier prévisionnel.....</i>	<i>27</i>

4. LES MISSIONS DU STAGE.....	29
-------------------------------	----

PARTIE 2 : LA CHAÎNE DE TRAITEMENT DES COLLECTIONS

SCIENTIFIQUES ET TECHNIQUES POUR LA BNUE.....	30
--	-----------

1. PRESENTATION DES COLLECTIONS PATRIMONIALES SCIENTIFIQUES ET TECHNIQUES	30
1.1. <i>Séries : R, S, T, V, Thèses</i>	31
1.1.1. Séries S « Sciences naturelles »	31
1.1.2. Série T « Sciences médicales »	32
1.1.3. Série R « Sciences philosophiques, morales et physiques »	32
1.1.4. Série V « Sciences et arts ».....	32
1.1.5. Thèses étrangères et françaises	33
1.1.6. La série Z « polygraphie, mélanges »	34
1.1.7. Autres collections scientifiques conservées à la BnF	34
1.2. <i>Les sciences médicales: cotation et catalogage</i>	34
1.2.1. Le catalogage des collections de la section T : quelques repères historiques	34
2. ETAT DE LIEUX DE LA BNUE DANS LE DEPARTEMENT SCIENCES ET TECHNIQUES..	39
2.1. <i>Sélection de printemps</i>	39
2.1.1. Les expositions universelles	40
2.1.2. Les Sciences naturelles.....	40
2.1.3. Les périodiques médicaux.....	41
2.1.4. Autres propositions	41
2.2. <i>Listes « Europ »</i>	41
2.3. <i>Justification et choix d'un corpus</i>	43
2.3.1. Philosophie médicale.....	44
2.3.2. L'anesthésie	46
2.3.3. La dimension historique du corpus.....	48
3. PROCESSUS DE SELECTION.....	48
3.1. <i>Critères juridiques</i>	50
3.1.1. Droit d'auteur pour les monographies	50
3.1.2. Droit d'auteur pour les périodiques	51
3.2. <i>Critères documentaires</i>	52
3.3. <i>Critères liés à l'état physique du document</i>	52

3.3.1.	Critères physiques bloquants	52
3.3.2.	Critères physiques non bloquants.....	53
PARTIE 3 : ANALYSE, EVALUATION ET BILAN DU CORPUS ETUDIE		54
1.	BILAN DES SELECTIONS.....	54
1.1.	<i>Présentation du corpus</i>	54
1.2.	<i>Corpus après l'examen documentaire</i>	57
1.2.1.	Répartition par critère documentaire	57
1.2.2.	Répartition par cote	58
1.2.3.	Répartition par date de publication	59
1.3.	<i>Corpus après l'examen physique</i>	61
1.3.1.	Résultats généraux.....	61
1.3.2.	Analyse des ouvrages rejetés	62
1.3.3.	Analyse du corpus sélectionné	63
1.4.	<i>Validations du corpus par Gallica et DCO</i>	65
2.	BILAN DES SELECTIONS DES DOCUMENTS AU SEIN DE LA BNF.....	66
2.1.1.	Département Sciences et techniques.....	66
2.1.2.	Direction des collections	67
2.2.	<i>Analyse d'une méthode de sélection expérimentale</i>	68
3.	ELARGISSEMENT ET ENRICHISSEMENT DU CORPUS	69
3.1.	<i>Catalogue méthodique des sciences médicales</i>	69
3.2.	<i>Autre pistes d'élargissement</i>	71
CONCLUSION		72
LISTE DES SIGLES ET DES ACRONYMES		74
BIBLIOGRAPHIE.....		77
DIAGRAMME DE GANTT.....		82
TABLE DES ANNEXES.....		83

Introduction

L'existence même de L'Europe comme espace de libre circulation des personnes et des valeurs, sans frontières, le droit et le besoin des citoyens de s'informer et d'accéder au savoir selon les traditions démocratiques, imposent une attention particulière à l'égard de la conservation et de la disponibilité de son patrimoine. La France a brillamment assumé son rôle dans l'affirmation de l'identité culturelle européenne. La pierre qu'apporte la BnF à la future Bibliothèque numérique Européenne (BNuE) en est une illustration manifeste. Tout d'abord par la réalisation du prototype Europeana, disponible en ligne sur le site de la BnF et présenté publiquement lors du Salon du Livre à Paris, en mars 2007. Ensuite par le rythme d'enrichissement annuel de sa propre bibliothèque numérique Gallica, qui passera de 6 000 documents à 100 000 documents par an, soit environ 300 000 documents d'ici à trois ans, la période prévue pour le projet. Tandis que la Direction des collections, dont le Département Sciences et techniques, qui m'a confié la mission de stage, assure la sélection documentaire, les services de numérisation et d'informatisation de la BnF ont la charge de la logistique du projet de numérisation de masse, dont les objectifs de qualité restent très exigeants.

L'objectif du projet de la BNuE se construit autour de l'accès aux sources de la pensée à travers une bibliothèque organisée en corpus documentaires cohérents et représentatifs de tous les champs de la connaissance, avec une indexation professionnelle rendue possible grâce au mode texte associé au mode image.

Notre travail a pour but la constitution et l'analyse d'un corpus en sciences médicales à partir des collections patrimoniales du Département Sciences et techniques de la BnF. Nous proposons d'abord une présentation évolutive de la bibliothèque numérique de la BnF, ce qu'elle est à présent, grâce à Gallica, ce qu'elle pourrait être avec Europeana, et ce qu'elle sera réellement, à partir de mi 2008, avec Gallica 2. Ensuite, tout en respectant les exigences documentaires, juridiques et physiques que la numérisation, la mise en ligne et la politique de sélection de la BnF imposent, nous justifions le choix de notre corpus. Enfin, nous dressons une analyse et une évaluation des résultats obtenus et indiquons des pistes pour l'élargissement du corpus étudié.

Partie 1 : Patrimoine national en ligne : offre, évolution et enjeux

Dès 1992, la BnF commence à constituer ses collections numériques grâce au développement de ses programmes de numérisation, ainsi qu'à l'évolution de la législation sur le dépôt légal, qui a introduit la même année, l'obligation de dépôt pour les ressources électroniques sur support. Ce dépôt est devenu effectif début 1994 après la promulgation du décret d'application.

Aujourd'hui, l'évolution des techniques de numérisation, les exigences des lecteurs ainsi que l'enjeu majeur de la démocratisation du savoir conduisent la BnF à affronter le défi d'un projet de numérisation sans précédent. Le passage de 6 000 ouvrages à numériser par an à 100 000 demande des efforts concertés à tous les niveaux. Cette démarche ne s'inscrit pas dans une logique de concurrence ou « d'orgueil européen » avec d'autres projets en cours, dont Google Recherche de Livres, mais dans une direction complémentaire, nécessaire pour l'identité culturelle européenne. A travers la Bibliothèque numérique Européenne on s'efforce à mettre à la disposition de tous, d'une manière cohérente et organisée, la culture européenne, dont on est fier, et qui constitue notre patrimoine écrit à transmettre et à faire connaître.

Notre premier chapitre présentera les trois étapes constitutives du projet de numérisation menée par la BnF, en commençant avec l'actuelle bibliothèque numérique, Gallica, avec ses points forts et ses faiblesses. Une courte revue des collections scientifiques et techniques contenues dans Gallica nous a semblé intéressante, ainsi que la présentation de grandes lignes de sa politique documentaire. Par suite, nous proposons un historique du prototype de la Bibliothèque numérique Européenne (BNuE), Europeana, ainsi que ses fonctionnalités qui émergent du Web 2.0 et ses axes documentaires, définis par la BnF. Mais Europeana reste un prototype qui n'évoluera plus et il est le reflet de la contribution française à la BNuE. Nous nous intéresserons ainsi à ce qui pourra être la Bibliothèque numérique Européenne dans le futur. Enfin nous projetterons, grâce à Gallica et à Europeana, la future bibliothèque numérique française, un Gallica renouvelé, un Gallica 2.

1. Gallica – Bibliothèque numérique de la BnF

1.1. Politique documentaire

Mise en chantier en 1992 et ouverte à titre expérimental en octobre 1997, Gallica est conçue comme une bibliothèque patrimoniale et encyclopédique. Toutefois, en dépit de cette vocation encyclopédique, on constate des disparités dans sa représentation du savoir : prépondérance conséquente des sciences humaines et sociales par rapport aux sciences dures. Toutefois la religion constitue une exception, car largement représentée dans les collections de la BnF, elle reste minoritaire dans les corpus de Gallica.

Plus de 40 % du fond de Gallica couvre le XIX^e siècle, période d'un grand essor intellectuel, qui constitue aujourd'hui un important terrain de recherche notamment historique. Une autre raison de la prédominance du XIX^e siècle dans Gallica est celle de l'origine des ouvrages pour la numérisation : si les livres du XX^e siècle sont couverts par les droits d'auteur, les livres du XVI^e au XVIII^e sont trop chers à l'achat. Toutefois, l'usage de microformes permet de nuancer cette tendance, ainsi que la riche production éditoriale dans le XIX^e siècle des ouvrages de l'Antiquité et de Moyen Age.

La couverture chronologique des autres siècles se répartit comme suit : 21 % pour le XVIII^e siècle, 6% pour le XVII^e, 7 % pour le XVI^e siècle et 3 % pour les documents de l'Antiquité, principalement d'ordre numismatique.

1.1.1. Critères juridiques

Gallica accorde une priorité absolue aux ouvrages libres de droit, bien que des négociations ponctuelles aient permis de présenter des éditions sous droits : ce qui explique pour partie les 18% de documents du XX^e siècle. Cependant, les ouvrages numérisés de Gallica s'arrêtent en 1937, pour respecter d'une manière stricte la législation liée aux droits d'auteur. Toutefois, les ouvrages sous droits peuvent être numérisés et sont consultables uniquement dans les salles de lecture de la BnF, lorsqu'un accord préalable avec l'éditeur a été défini.

1.2. Le fonds de Gallica

Gallica pouvait se prévaloir au 1^{er} novembre 2006 de 185 000 imprimés en mode image¹, dont 60 000 fascicules de presse (mode image + mode texte), de 105 000 images fixes, issus des fonds patrimoniaux de la BnF et des collections d'autres établissements. A cela s'ajoutent environ 30 heures de sons, extraites principalement des *Archives de la parole*. Il faut préciser que le fonds de Gallica est en constante évolution suite aux marchés de numérisation engagés par la BnF.

1.2.1. Les imprimés (monographies et périodiques)

Gallica se compose à 66 % de monographies imprimées, publiées en France, en langue française. On y retrouve un millier de dictionnaires et encyclopédies, spécialisés comme le *Dictionnaire de droit et de pratique* de Ferrière ou *Dictionnaire encyclopédique des sciences médicales* de Dechambre, ou des dictionnaires généralistes, comme le *Dictionnaire* de Furetière. Il faut noter la forte demande de la part des chercheurs de numérisation d'instruments tels que les dictionnaires ou les catalogues spécialisés, les outils bibliographiques, notamment les titres épuisés ou difficilement accessibles. A ces dictionnaires et catalogues s'ajoutent des textes de référence : corpus complets des grands auteurs comme Balzac ou Chateaubriand et des textes fondateurs de différents domaines de la connaissance. Des améliorations sont attendues dans la nouvelle version de Gallica concernant surtout les modalités de recherche en texte intégral, une fois la retroconversion des caractères conclue.

Les 33 % d'imprimés restants sont des périodiques (plus de 800 titres) pour lesquels la BnF s'attache à combler systématiquement les lacunes (numéros manquants) et à ajouter les titres les plus importants dans la chaîne de production. Gallica détient essentiellement des revues de création, de critique et d'érudition, organisés d'une manière thématique. Des progrès techniques doivent être apportés dans Gallica 2 facilitant l'accès à ces revues par une navigation dans les tables de matière.

1.2.2. Images, sons et presse

Les images de Gallica sont principalement issues de la numérisation exhaustive de fonds constitués à la BnF et retenus en fonction de leur intérêt pour la recherche. On peut citer les cartes de la collection du géographe Jean-Baptiste Bourguignon d'Anville, du

¹ Dont plus de 1240 en mode texte, à titre expérimental actuellement.

Département Cartes et Plans, les plans d'architectures d'Etienne-Louis Boullée ou de Jean-Jacques Lequeu du Département des Estampes et de la Photographie ou le fonds musical Philidor. Ces images peuvent être consultées via le catalogue ou feuilletées par fonds, comme leur version d'origine.

Les ressources sonores relevant du fonds général, sont principalement issues d'un corpus constitué, les *Archives de la parole*, des enregistrements réalisés au début du siècle, par le linguiste Ferdinand Brunot. Ce fonds présente un intérêt intrinsèque mais il peut également être considéré comme étant le noyau de ce qui deviendra un siècle plus tard le Département de l'audiovisuel de la BnF.

La presse quotidienne constitue un véritable chantier scientifique et technique de Gallica, qui a entrepris la numérisation de 26 titres significatifs de la presse française du XIX^e siècle et du début du XX^e siècle. Les estimations à atteindre s'élèvent à plus de 3 millions de pages en mode image et en mode texte, pour une durée prévue de 5 à 7 ans. Ce corpus constituera l'ensemble de presse quotidienne le plus important au monde mis en ligne, dans sa partie rétrospective. Des accords entre les éditeurs détenteurs de droits et Gallica vont s'opérer pour la mise en ligne des titres des années récentes. La pratique de la « barrière mobile » préserve les intérêts des partenaires : celui de la BnF qui a vocation à mettre gratuitement le patrimoine écrit français dans ses salles de lecture et en ligne pour la partie libre de droits mais également celui des organismes de presse qui peuvent exploiter commercialement la partie récente.

1.2.3. Politique de numérisation partagée

L'offre documentaire de Gallica s'est constituée grâce aux différents partenariats et collaborations conclus avec d'autres établissements français, par la mise en place de nombreux programmes de numérisation partagée².

Les principes décrits dans la charte documentaire de Gallica ne pourront pas être accomplis sur les seules collections de la BnF. Les programmes de numérisation partagée se définissent dans deux perspectives :

- le regroupement virtuel de corpus complémentaires (fonds Philidor de musique baroque réparti entre Paris et Versailles, par exemple) et l'articulation des différents portails ;

² Medic@ (BIUM) ; Cnum (CNAM) ; bibliothèque numérique du Muséum, de l'Ecole vétérinaire, Lyon 1, CNRS Paris 1, Médiathèque de la Cité des sciences de la Villette, Ecole Centrale de Lyon etc.

- coopération documentaire entre la BnF et des organismes spécialisés, vue également comme une complémentarité en termes de compétences. L'apport de la cellule Mathdoc (Unité mixte Université Joseph Fourier-Grenoble, CNRS) constitue une réelle valeur ajoutée en termes d'indexation et de navigation au document.

La BnF mène une politique de coopération en régions notamment dans les programmes des revues des Sociétés savantes du XIX^e afin de valoriser les fonds patrimoniaux réparties en France entre la BnF et les bibliothèques municipales.

Le protocole OAI, protocole d'échange de métadonnées, se situe au cœur de la politique de numérisation partagée, car il permet de réunir des ressources hébergées par les serveurs de différentes institutions. Un exemple de réussite est constitué par le dossier France-Amérique, mis en ligne en 2005 qui, permet à l'utilisateur, à partir du site Gallica ou du site de la Bibliothèque du Congrès de mener une recherche sur les ressources des deux établissements.

1.3. L'existant en ligne pour le Département Sciences et techniques

Les collections scientifiques et techniques en ligne comprennent plus de 11 594 volumes, avec une prépondérance significative pour le XIX^e siècle. La volumétrie du domaine de la médecine, qui nous intéresse particulièrement, s'élève à 3000 volumes environ. Parmi les 400 volumes d'instruments intellectuels, on peut compter le célèbre *Dictionnaire encyclopédique des sciences médicales* de Dechambre ou la *Bibliothecae Medicinae* de Haller. Gallica permet l'interrogation, pour l'instant en mode image, à terme en texte intégral, de plus de 8000 études et traités de production scientifique pure, répartis dans des corpus thématiques. A cela s'ajoutent environ 500 ouvrages de vulgarisation (presque 80% du XIX^e siècle) et d'ouvrages d'enseignement élémentaire ou professionnel.

L'offre de périodiques de vulgarisation ou spécialisés, disponible en ligne, dépasse 55 titres, soit 2592 fascicules couvrant toutes les disciplines, sauf les sciences de la terre. Les 489 images proviennent notamment des organismes partenaires, comme le Musée National d'Histoire Naturelle ou Musée de l'Homme.

Ayant comme but d'accroître l'offre documentaire et sa visibilité, Gallica prévoit des enrichissements de son fonds pour compléter des lacunes dans diverses disciplines (astronomie, sciences de la terre, psychologie), pour constituer des ensembles

thématiques, des corpus de revues scientifiques, des corpus d'images ou bien des œuvres complètes des savants ou des vulgarisateurs du XIX^e siècle.

La version actuelle de Gallica reste disponible en ligne et sera modernisée grâce à la conversion OCR en mode texte des ouvrages numérisés au sein du projet de la Bibliothèque numérique Européenne. Gallica s'enrichira également grâce à la numérisation de la presse, en mode image et texte. Néanmoins, son architecture et ses fonctionnalités n'évolueront plus.

La fermeture de Gallica est prévue pour l'été 2008 et elle va être remplacée par Gallica 2 qui va intégrer les fonctionnalités actuelles de Gallica. A cela s'ajouteront les fonctionnalités présentes dans Europeana et celles prévues mais pas encore développées dans la maquette Europeana. Il faut préciser que Gallica 2 existe dès maintenant mais qu'elle est à un stade de préparation.

2. Europeana – contribution française a la BNUE

2.1. Historique

Face aux défis lancés par le projet de mise en accès de 15 millions de livres numérisés lancé fin 2004 par la firme Google et suite au débat public ouvert par Jean-Noël Jeanneney³, Jacques Chirac, ancien président de la République, lance l'idée d'une bibliothèque numérique européenne. Il a le soutien d'autres chefs d'Etat européens (Espagne, Allemagne, Italie, Hongrie, Pologne), des institutions de l'Europe, de Jean-Claude Juncker, alors président du Conseil européen et de Jose Manuel Barroso, président de la Commission européenne. Vingt-trois bibliothèques nationales des Etats de l'Union européenne sont signataires d'une motion pour soutenir cet immense projet de numérisation.

Renaud Donnedieu de Vabres, alors ministre de la Culture et de la Communication, a demandé à la BnF le 30 août 2005 d'établir un état de lieux précis sur une contribution française au projet de la BNUE. Cinq groupes de travail réunissaient des acteurs du privé (représentants du monde de l'édition, des secteurs industriels de la numérisation et des moteurs de recherche) et du secteur public (Culture, Education nationale, Recherche, Industrie, Finances). Renaud Donnedieu de Vabres présidait le comité de pilotage, créé en juillet 2005 et l'ancien président de la BnF, Jean-Noël Jeanneney, en assurait la vice-

présidence. Le comité a remis le 11 janvier 2006 un « livre blanc », présenté en Conseil des ministres le 8 février 2006. Ce document propose une analyse stratégique du positionnement de la BNuE, les plus-values attendues, les acteurs concernés en priorité et les scénarios envisagés. « Le livre blanc » énonce quelques propositions consensuelles :

- les textes de référence de l'identité européenne, organisés en corpus, choisis grâce à une charte documentaire de la BNuE, seront mis à la disposition du public ;
- des livres, des revues et des journaux seront proposés dans un premier temps mais un élargissement à d'autres média est envisagé ;
- les ouvrages libres de droits sont prioritaires mais, sous réserve d'accords avec les éditeurs européens, des ouvrages sous droits pourront également être mis en ligne ;
- des financements publics, nationaux et communautaires sont prévus mais également d'autres types de ressources.

La Bibliothèque nationale de France a reçu le 2 mai 2006 du gouvernement français la responsabilité officielle du pilotage opérationnel du projet en France.

Parallèlement, pour informer de l'état de réflexion en France, des contacts sont pris tant avec la Commission européenne, qu'avec la présidence de l'Union européenne. La Commission européenne organise la consultation du programme *i2010 Bibliothèque numériques* sur ce sujet, dont les résultats sont annoncés le 2 mars 2006 dans un communiqué de presse sur les démarches à accomplir pour mettre à la disposition de tous la « mémoire de l'Europe ». La future bibliothèque européenne reposera sur le programme TEL⁴, soutenu par CENL⁵ dont l'objectif est de permettre dans les années à venir l'accès aux collections des 47 bibliothèques nationales, membres de la CENL.

TEL reste encore assez méconnu par les internautes mais il offre l'avantage d'une expérience de projet commun à plusieurs bibliothèques nationales européennes et pourrait jouer aussi un rôle dans la réalisation de la BNuE.

La Commission a publié le 24 août 2006 une recommandation concernant la numérisation et l'accès sur la Toile au « matériel culturel et à la conservation

³ JEANNENEY, Jean-Noël. *Quand Google défie l'Europe* : plaidoyer pour un sursaut, Paris, Mille et une nuits, 2005.

⁴ The European Library permet l'accès en ligne aux catalogues de 23 bibliothèques nationales européennes.

numérique », approuvée par le Conseil des ministres de la Culture et de l'Audiovisuel, le 13 novembre 2006.

2.2. Autour du prototype Europeana

Afin de mener à bien sa mission, la BnF ouvre plusieurs chantiers de travail et met en place une stratégie culturelle complexe en mobilisant de nombreux acteurs pour la sélection des ouvrages à numériser, les manipulations physiques, le contrôle de qualité, la mise en ligne des ouvrages, la conservation des données produites. La coopération avec les partenaires européens se situe également au cœur de la réflexion, notamment en ce qui concerne l'élaboration de la charte documentaire de la BNUe (les documents à numériser, à quel rythme, selon quelles modalités) et la prise de décision sur les aspects juridiques, techniques et financiers du projet.

Un premier chantier décisif, interne à la BnF, suppose une intensification du programme de numérisation des collections patrimoniales, indispensable pour une contribution pertinente à la BNUe. Le rythme de numérisation passe, à partir de cette année, de 6 000 documents à 100 000 par an. De février à septembre 2007, il est prévu la numérisation des 30 000 documents supplémentaires. A partir de l'automne 2007 jusqu'à août 2010, une dématérialisation de 100 000 documents par an alimentera la future bibliothèque numérique de la BnF.

Une concrétisation des opérations de modernisation technique de Gallica a été réalisée par le biais du prototype Europeana, dont la version « beta » a été ouverte au public lors de l'inauguration du Salon du livre à Paris le 22 mars 2007. Europeana est disponible en ligne sur le site de la BnF à l'adresse : <http://www.europeana.eu>. Ce prototype, qui n'évolue plus, a pour objectif de montrer de manière concrète et pragmatique les résultats auxquels il est possible d'aboutir, avec comme point de départ la maquette présentée à l'automne 2006 aux instances politiques et aux professionnels de bibliothèques. Le prototype met en oeuvre sur un volume conséquent de documents numérisés, des fonctionnalités nouvelles, comme la recherche plein texte au sein d'un ouvrage ou l'étiquetage des textes dans les espaces personnels. Nous les présenterons d'une manière détaillée dans l'Annexe 1.

⁵ Conference of European National Libraries, réunit les directeurs des bibliothèques nationales en vue d'une meilleure coopération documentaire.

2.2.1. Politique documentaire d'Europeana

Le prototype comprend un ensemble de 7 000 monographies des collections patrimoniales françaises en mode image et en mode texte. Les collections qui y figurent répondent à l'objectif scientifique de constituer une bibliothèque à vocation encyclopédique, représentative du patrimoine francophone et européen.

Le choix de textes d'Europeana s'articule autour de trois axes documentaires complémentaires :

- un corpus construit par des ouvrages de référence, fondateurs de la culture francophone et européenne dans le domaine littéraire et dans le domaine scientifique : œuvres de Corneille, Diderot, Rousseau, Shakespeare, Dickens mais aussi Kepler, Galilée, Newton.
- un corpus autour des œuvres classiques : critique littéraire, textes de vulgarisation scientifique, essais, correspondances, sélection de mémoires et de biographies : correspondances d'Auguste Comte ou Berlioz et aussi les mémoires de Saint-Simon, Louise Michel, la vie de Bonaparte, George Sand ou Victor Cousin ;
- un ensemble d'outils intellectuels, comme des encyclopédies (*Encyclopédie du XIX^e siècle*), des dictionnaires (*Dictionnaire des finances* ou *Dictionnaire de la Bible*), des manuels biographiques ou bibliographiques. La presse suscite une forte demande au sein des chercheurs, au-delà de la littérature et des sciences, domaines pour lesquels il y a un intérêt avéré.

Le prototype permet également l'accès à 4 000 documents de la Bibliothèque Nationale de Hongrie, qui comprend notamment des études sur l'histoire, la tradition populaire et la littérature hongroises, ainsi que des œuvres classiques d'auteurs hongrois et des traductions en hongrois de classiques européens. Ces documents proviennent de la bibliothèque électronique hongroise Magyar Elektronikus Könyvtár.

En plus des ouvrages hongrois, on compte également une collection portugaise de 1000 ouvrages, composée principalement de documents sur l'histoire du Portugal, de récits de voyages, ainsi que de divers ouvrages sur le droit, les arts et l'éducation. Ils proviennent de la bibliothèque numérique du Portugal, Biblioteca Nacional Digital. Les documents de ces deux bibliothèques nationales sont consultables sur leur site d'origine.

2.2.2. Informations techniques et documentaires sur Europeana

Le prototype Europeana est développé sur une plateforme indépendante de celle de Gallica et la volumétrie totale des documents est de 700 Go. Pour sa mise en œuvre, 6 000 documents de Gallica ont été repris en leur attachant le mode texte au mode image déjà existant. Le texte est issu de la conversion automatique par un logiciel de reconnaissance de caractères. La conversion est non corrigée et peut comporter un certain nombre d'erreurs. Le taux de reconnaissance plancher est de 96% mais 10% environ des documents numérisés par la BnF a un taux de 99,9 %.

Le mode texte fournit une série de facilités à la recherche et est intimement lié à une indexation pertinente d'œuvres et de corpus. Il est également compatible avec les outils de lecture d'écrans, utilisés dans le cadre de la déficience visuelle.

Le moteur de recherche d'Europeana est un logiciel libre intitulé Lucene, de la fondation Apache, qui propose un certain nombre de fonctionnalités : la recherche plein texte, la recherche dans les notices, le classement des résultats. Une plus value apportée par la BNUe sera la capacité de son moteur de recherche à présenter de façon organisée, par une navigation intuitive et utilisant les ressources du langage naturel, des contenus qui ont de fortes connexions entre eux. Par exemple, par une même interrogation on pourra avoir les titres de revues politiques publiées entre 1920 et 1950 en France, Allemagne et Grande-Bretagne.

Une série de fonctionnalités supplémentaires, non développées dans le prototype, est envisagée pour Gallica 2, comme par exemple la recherche avancée, les opérateurs booléens et les outils sémantiques. Ils étaient d'ailleurs envisagés dans la maquette Europeana.

L'algorithme de pertinence du moteur Lucene propose un classement des résultats de recherche en fonction d'une analyse des mots clés, du nombre d'occurrences retrouvées, de leur fréquence et de leur emplacement dans le contenu. Par exemple, un mot trouvé dans le titre d'un document a un taux de pertinence plus élevé qu'un mot trouvé dans le corps du texte.

L'application informatique du prototype est déployée sur 26 serveurs de dernière génération (bi-processeurs) hébergés par la BnF. En cas de forte sollicitation la société Atos Wordline a prévu des dispositifs complémentaires.

2.3. Enjeux d'une bibliothèque numérique européenne

La Bibliothèque numérique Européenne apporte une réponse européenne aux enjeux culturels et commerciaux de la diffusion du savoir sur la Toile. Ainsi, la mise en ligne de la culture est devenue à travers ce projet un problème autant industriel que politique.

Nous proposons de distinguer deux conceptions différentes du rapport aux documents et à la bibliothèque, que l'on peut dresser à travers les projets de numérisation en cours.

On peut mettre en évidence une approche latine d'une part, qui privilégie l'unité, la construction des contenus et la transmission du savoir. C'est le cas d'une Bibliothèque Numérique Européenne pour laquelle l'universel trouve sa source en amont de ce qui est produit et qui accorde une priorité à la préservation du patrimoine commun, symbolisé par le dépôt légal de tout document. Le projet européen est construit selon une planification prudente dans un esprit très organisé et rigoureux. Les acteurs qui assument les responsabilités sont les Bibliothèques nationales et les institutions gouvernementales.

L'autre approche, plus anglo-saxonne, insiste au contraire sur la diversité et sur l'accès à l'information. L'universel est en aval dans l'accès pour tous au patrimoine commun. Le côté pragmatique du projet (rapidité d'accès) est mis en avant au profit d'une organisation des savoirs.

Ces deux conceptions sont complémentaires, comme le souligne Jean-Michel Salaun⁶ et le défi d'une bibliothèque numérique européenne est de construire un équilibre satisfaisant entre l'une et l'autre et de faire ressortir ainsi la richesse d'une telle dialectique. L'accès aux sources de la pensée à travers une bibliothèque organisée en ligne constitue donc le défi majeur de la future Bibliothèque numérique Européenne.

3. Gallica 2 – vecteur d'évolution de Gallica

Gallica a apporté à la BnF une expérience importante des choix documentaires, des techniques de numérisation et des usages des collections. Projection de la future bibliothèque numérique de la BnF, Gallica 2 sera la version modernisée de Gallica, qui intégrera des fonctionnalités du Web 2.0 et qui se construira probablement selon le modèle d'Europeana.

⁶ SALAUN, Jean-Michel, Bibliothèque numériques et Google-Print. *Regard sur l'actualité*. La Documentation française, 2005, p.11.

3.1. Politique documentaire

La politique documentaire de Gallica 2 suit les grands axes définis pour Gallica avec quelques remaniements, mais à une échelle beaucoup plus grande. Dans Gallica 2, l'accent sera mis sur les textes plutôt que sur les images, sur les documents primaires plutôt que sur les outils critiques. Le domaine de l'art sera très peu représenté car la numérisation de l'image est plus coûteuse et l'INHA est également impliqué dans des projets de numérisation.

Quelques catégories des documents seront refusées dans un premier temps, comme c'est le cas pour les livres riches en images (albums, guides de voyages, « beaux-livres », bandes dessinées), les manuels scolaires et parascolaires.

3.1.1. Le patrimoine national

La constitution d'un corpus de l'offre patrimoniale française constitue un des premiers objectifs de Gallica et *Gallica Classiques* en est un excellent exemple. La numérisation des ouvrages en vue de leur intégration dans Gallica 2 se poursuivra en complémentarité avec les collections déjà existantes : des ouvrages de critique littéraire et des revues littéraires s'ajouteront au domaine littéraire, les sources du droit au domaine juridique etc. Un accent particulier sera accordé à l'histoire de France du XIX^e siècle, en ciblant quelques moments importants pour l'essor de la démocratie : la révolution de 1848, l'affaire Dreyfus, le syndicalisme, l'immigration, le colonialisme. Les régions françaises jouiront de plus de visibilité par la continuation de mise en ligne des publications de Sociétés savantes et par des textes sur leurs traditions spécifiques.

Le corpus sur les sciences s'agrandira surtout autour des sciences naturelles et les grands naturalistes français ou sur la médecine dans ses aspects les plus sociaux.

Un chantier conséquent est envisagé pour la numérisation des outils bibliographiques : dictionnaires et bibliographies spécialisés, annuaires de la presse, catalogues d'éditeurs, revues françaises scientifiques et d'érudition.

3.1.2. Un corpus à vocation internationale

Afin de renforcer la coopération avec d'autres bibliothèques nationales européennes, quelques domaines de connaissance ont été choisis pour être numérisés par différents programmes nationaux de numérisation. Il s'agit de la presse quotidienne nationale pour laquelle La Belgique, l'Autriche, la Grande Bretagne ont lancé des programmes similaires de numérisation. D'autres axes sont constitués par la généalogie (recherche

identitaire des personnes ou des territoires), les langues (constitution d'un fonds sur l'apprentissage de la langue française par les étrangers, l'histoire de la francophonie illustré par exemple par le statut de l'Alliance française) ou la documentation juridique (archives parlementaires, recueils de lois, jurisprudence). A cela s'ajoute les traductions dans les langues européennes, des chefs-d'œuvre des différentes littératures nationales.

3.1.3. Un corpus français sur l'Europe

Gallica 2 doit faire connaître à ses internautes les spécificités du continent européen, emblème de culture et civilisation dans le monde. L'intérêt est de présenter l'Europe dans toutes les dimensions qui ont contribué à son essor : les identités nationales, l'économie, le commerce, le développement de grandes institutions financières, le travail, les métiers, les sciences et les techniques (développement du chemin de fer, des méthodes agricoles, de la médecine), les problématiques sociales, l'éducation, l'enseignement, les voyages, les expositions universelles. Le corpus sur l'Europe reflétera les relations et les échanges extra-européens avec les Etats-Unis, l'Afrique ou l'Asie. La presse européenne occupe une place prioritaire dans la constitution du corpus : des gazettes du XVIII^e siècle, *L'économiste européen* ou *Gazette politique de l'Europe* en sont juste quelques exemples.

La BnF a déjà une expérience dans la constitution du corpus sur l'Europe avec la sélection des ouvrages pour les « Listes Europ* » au sein du « marché des 30 000 ». L'axe documentaire du « marché des 100 000 » suit donc ce qui a été déjà entrepris. Le Département Sciences et techniques poursuit notamment la numérisation des documents publiés à l'occasion des expositions universelles, projet conduit avec le CNAM, mais également la numérisation des ouvrages consacrés aux inventions et aux travaux d'ingénieurs.

Une des difficultés de la constitution de ce corpus consiste dans la dissémination des ouvrages concernés dans l'ensemble des collections de la BnF.

3.1.4. Les plus-values d'une offre d'ouvrages sous droits

L'offre des ouvrages sous droits dans Gallica 2 s'inscrit dans la politique documentaire de la BNUe. Il s'agit d'une complémentarité éditoriale lorsqu'on veut retrouver des ouvrages de critique récente sur les auteurs ou scientifiques classiques, qui présentent un véritable intérêt pour les chercheurs d'aujourd'hui. De même, les dictionnaires et les

encyclopédies du début du siècle ont un intérêt historique, mais sont évidemment dépassés par l'évolution des sciences et des disciplines. Nous pensons notamment à l'économie, la psychologie, la psychiatrie, les sciences politiques, les sciences de l'éducation, des média ou de la mondialisation. La présence d'ouvrages du XX^e siècle dans ces domaines entraînera dans un dialogue intellectuel le savoir du passé et le savoir contemporain. Cela enrichira également les thématiques des siècles précédents (droit, histoire de France, la langue française, travaux critiques autour de la presse) qui bénéficieraient d'un regard contemporain renouvelé. En plus, une offre sous droits des éditions à diffusion restreinte et très fragmentée (éditions régionales ou publiques, éditions érudites) les rendra plus visibles en regard de l'offre patrimoniale.

Toutefois, la mise en ligne des ouvrages sous droits doit se prévoir dans le cadre d'un dispositif concerté assurant une juste rémunération des ayants droit. Des discussions entre le Syndicat national de l'édition et la Bibliothèque nationale de France ont été engagées afin de proposer un modèle économique permettant cette rémunération. La société Numilog a été désignée pour conduire une étude d'expert sur ce modèle et les résultats seront bientôt disponibles.

3.2. Politique logistique et budgétaire de la numérisation

La chaîne de production de la numérisation de la BnF doit s'adapter aux nouveaux défis imposés par la Bibliothèque numérique Européenne. Comme il s'agit d'une numérisation de masse, de type industriel, la réalisation en a été déléguée dans une certaine mesure à des prestataires extérieurs spécialisés.

3.2.1. Marchés de dématérialisation

Plusieurs marchés de dématérialisation se juxtaposent. Nous présenterons les marchés en cours et en prévision à la BnF pour les trois années à venir :

- le « marché numérisation annuelle Gallica » de 6 000 documents (1 800 000 pages) prévue entre janvier et septembre 2007, sur le modèle interne : sélection => sortie des documents => catalogage => pistage => numérisation => contrôle => chargement. Seule la numérisation est sous-traitée ;
- la conversion de 6 000 documents de Gallica en vue de la réalisation du prototype Europeana ;

- la reprise de 60 000 documents de Gallica pour leur associer le mode texte par traitement OCR (environ 16 000 000 de pages dont 600 000 en haute résolution). Le marché est prévu de janvier 2007 à janvier 2008 ;
- le « marché des 30 000 » documents (9 000 000 pages) de février à septembre 2007, pour lequel la numérisation sera effectuée à la fois en mode image et en mode texte ; les prestataires extérieurs (Jouve, Diadeis et Infotechnique) sont chargés de la numérisation, de l'OCR et du contrôle des exemplaires numérisés. Le contrôle des 30 000 documents est à effectuer de mai 2007 à mars 2008 et l'exemplarisation/dépouillement entre mars et juillet 2007 ;
- le « marché des 100 000 » prévoit la dématérialisation de 100 000 documents⁷ par an (14 400 000 pages) de septembre 2007 à août 2010. Ce chantier suppose un changement de chaîne de production et de système d'information mais également l'externalisation massive des charges (numérisation, catalogage, OCR, contrôle qualité, correction).

3.2.2. Filières de production

Trois filières de production de numérisation ont été définies :

- la filière « papier massicoté », filière classique de Gallica : la numérisation s'exécute à partir des documents achetés dont les pages sont massicotées pour une numérisation plus rapide et à moindre coût. Il s'agit de 10 000 documents numériques par an (environ 2 400 000 pages) au sein du « marché des 100 000 » ;
- la filière « papier relié », instaurée en 2006 porte sur les collections conservées dans les magasins de la BnF, constituées grâce au dépôt légal, pour la plupart. Cette filière concerne 50 000 documents numériques par an soit 7 500 000 pages sur le « marché des 100 000 » ;
- la filière « microforme » comprend les microfiches et les microfilms prélevés dans la collection de sécurité du Centre de conservation de la BnF, à Bussy-Saint-Georges. Comme ils sont exclus de la communication au public, ils se trouvent dans un bon état et se prêtent à la numérisation. Il s'agit d'environ 45 000 documents par an, soit 4 500 000 pages dans le « marché des 100 000 ».

⁷ Voir Annexe 2 pour les variations minima/maxima dans les départements.

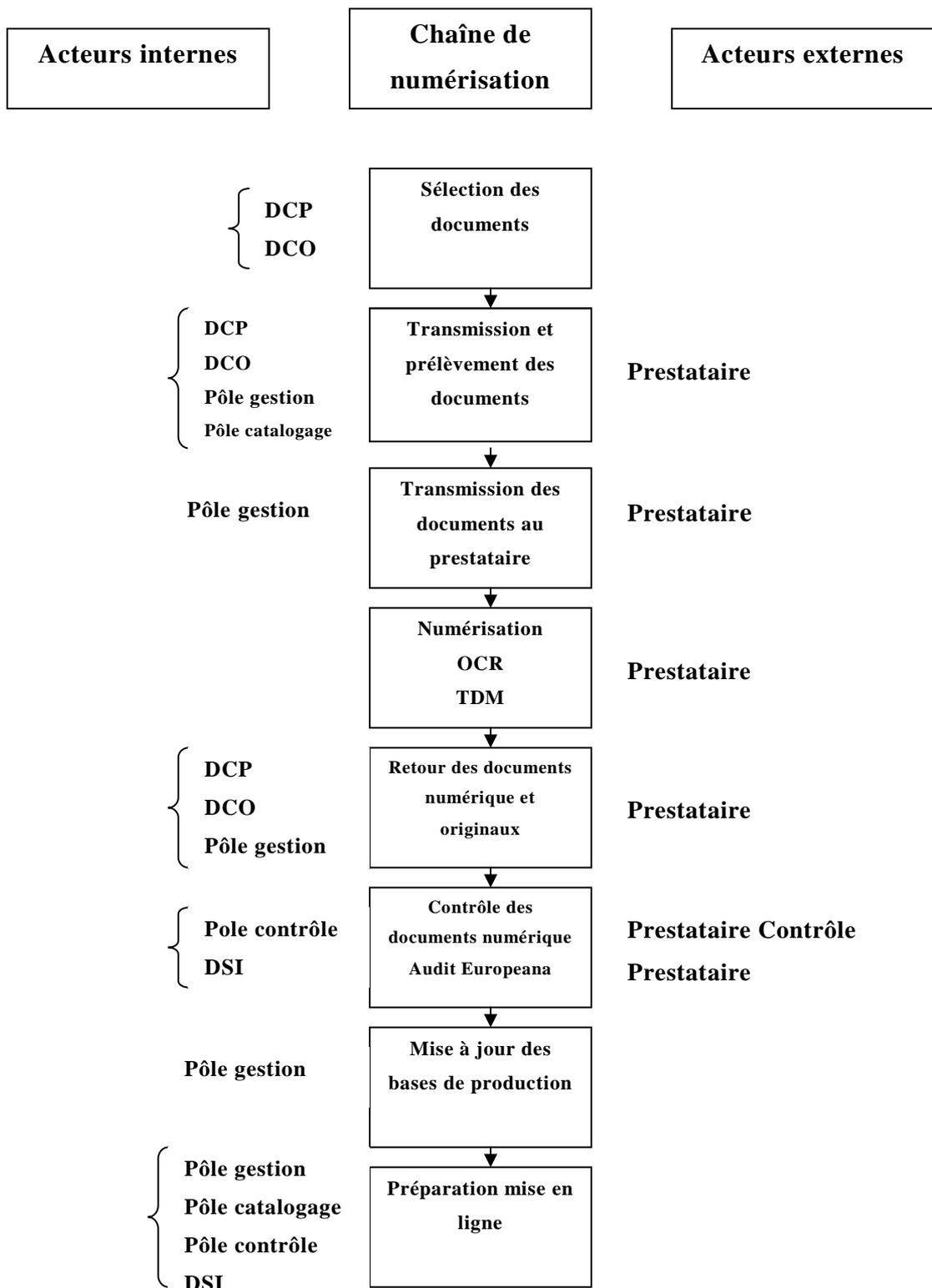
3.2.3. Choix techniques de numérisation

La BnF a défini ses propres critères en ce qui concerne la numérisation des ouvrages :

- choix de la profondeur d'acquisition : 1 bit = 2 tons, recommandé pour les textes ; 8 bits = 256 tons, pour les images et les manuscrites ; 24 bits = 16,7 millions de tons, pour les images en couleur ;
- choix de la résolution : la BnF a choisi 300 dpi, comme règle de base pour la numérisation d'un original et 600 dpi pour saisir les petits détails, les traits pour la presse par exemple ;
- pour la compression des images, la BnF a décidé que les formats engendrant des pertes de données irréversibles ne doivent pas être utilisés pour la sauvegarde des images. L'archivage des ouvrages numérisés s'effectue en TIFF non compressé et la compression en UIT-T GIF est admise seulement pour les documents en noir et blanc. Les fichiers numériques obtenus doivent être renseignés afin d'une alimentation en métadonnées de préservation : informations techniques sur l'image (taille, résolution, compression, données colorimétriques), informations de production (matériel utilisé, producteur et date de production), informations administratives (auteur, propriété, localisation).

3.2.4. Chaîne de traitement et de numérisation

Nous présentons dans le schéma suivant la chaîne de numérisation de masse ainsi que les acteurs qui y participent. Il s'agit d'une forte mobilisation de toutes les équipes de la BnF et d'un intense travail collaboratif. Par ailleurs, si pour les marchés antérieurs, ceux de Gallica, Europeana ou le « marché des 30 000 », la BnF a externalisé seulement une partie des tâches auprès des prestataires, la situation change avec la numérisation de masse pour laquelle les prestataires extérieurs prennent en charge plusieurs étapes de la chaîne de traitement.



Le pôle de gestion du service de numérisation joue un rôle actif dans la plupart des étapes du projet : en amont et en aval des tâches confiées aux prestataires, mais également dans tous les travaux de numérisation de sauvegarde, qui sont menés en interne à la BnF.

La DCO veille à l'état des collections lors des prélèvements dans le magasin, puis à leur retour, et elle participe pleinement au processus intellectuel de sélection des ouvrages.

Le pôle catalogage intervient à deux moments de la chaîne de numérisation : dans un premier temps, les catalogueurs doivent récupérer dans une base interne les fiches descriptives d'ouvrages, préalablement validées par la DCO. Cela est nécessaire pour la création de l'exemplaire numérique, rattaché à la fiche descriptive (exemplarisation). Tout document qui sera numérisé aura d'abord un exemplaire numérique signalé dans le catalogue BN Opale Plus, en correspondance avec la notice bibliographique de l'ouvrage. A partir des fiches descriptives un bordereau électronique sera édité, comprenant les métadonnées du document, comme par exemple les descriptions bibliographiques et les informations sur le droit d'auteur. Ces informations seront envoyées en ligne au prestataire, qui les complétera avec les métadonnées techniques de numérisation et de la structure interne du document.

La transmission des documents vers le prestataire s'effectue dans des caisses et les livres sont protégés sous pochettes de plastique munies de code-barres. Lors de la réception des fichiers numérisés, c'est une application informatique interne qui assure le contrôle qualité automatique des informations techniques et bibliographiques. Un autre contrôle humain complémentaire s'effectue sur la qualité et la cohérence des images. Si les critères de qualité de la BnF ne sont pas satisfaits, les fichiers sont renvoyés au prestataire.

Les documents originels reviennent dans les magasins après le pistage obligatoire et la base de production est ainsi mise à jour. Une fois ces opérations accomplies, les documents peuvent migrer sur Gallica, par l'exercice conforté des pôles gestion, catalogage, contrôle et le DSI.

3.2.5. Aspects économiques des programmes de numérisation

Avant de se lancer dans la numérisation de masse, la BnF a commandé un test de numérisation au printemps 2006 auprès d'IBM, arrivant à un résultat moyen de 0,09 €/brut par page (pour un taux de reconnaissance des caractères de 96 %) pour la

numérisation et la conversion en OCR. Mais une meilleure restitution en ligne de certaines types d'ouvrages (dictionnaires, encyclopédies) ou de parties d'ouvrages (table des matières, index) impose le recours à une conversion par reconnaissance optique de caractères de haute qualité (taux de reconnaissance de 99,8 %) pour un prix nettement supérieur. Donc le coût de la numérisation, de l'OCR et du contrôle qualité varie de 0,065 € à 0,19 € TTC par page selon le type de document (massicoté, relié, microformes).

En 2006 la BnF a obtenu 3 millions d'euros pour les marchés de numérisation et 375 000 € provenant du Ministère de la Culture et de la Communication, nécessaires pour le test de numérisation de masse et pour l'acquisition des équipements informatiques.

En 2007 la BnF bénéficie des 10 millions euros versés par le Centre National du Livre pour financer la contribution française à la Bibliothèque numérique Européenne. L'achat de serveurs, des travaux d'aménagement des locaux et des études préalables nécessaires ont été déjà engagés, dont la valeur estimée s'élève à 200 000 €.

La Commission européenne a accueilli favorablement le projet de la Bibliothèque numérique Européenne et elle a tracé les grandes lignes qui contribueront à sa réalisation, par la mise en œuvre de programmes coordonnant et finançant les différentes initiatives nationales. Le programme « E-content Plus » en est un très bon exemple. Depuis 2005 et jusqu'en 2008, la Commission finance des programmes, dont le but est l'accessibilité en ligne et l'interopérabilité des contenus numériques multilingues. Sa participation totale s'élève à 149 M€.

3.3. Calendrier prévisionnel

Nous présentons en cette fin de chapitre un calendrier prévisionnel des activités attendues dans le cadre du projet de la Bibliothèque numérique Européenne à la BnF jusqu'à fin 2008 :

Ete 2007	<ul style="list-style-type: none"> -Enrichissement des collections de la nouvelle version de Gallica (en texte et image), parallèlement à l'ancienne -Multilinguisme de l'interface (anglais, espagnol) -Corrections éventuelles -L'ajout de zoom pour les monographies
Décembre 2007	<ul style="list-style-type: none"> -Visualisation des images -Visualisation des textes -Visualisation des périodiques -Recherche avancée
Mi 2008	<ul style="list-style-type: none"> -Visualisation des cartes (fonction zoom) -Visualisation de la presse (grand format) -Diffusion des documents audiovisuels -Accès collections partenaires -Création de corpus -Gestion de contenu -Fermeture du site Gallica actuel
Fin 2008	<ul style="list-style-type: none"> -Moteur sémantique -Travail collaboratif

En ce qui concerne les fonctionnalités présentes et celles attendues sur le site de Gallica 2 d'ici mi-2008, nous dressons le bilan suivant :

Fonctionnalités actuelles de Gallica	Fonctionnalités présentes dans le prototype Europeana	Fonctionnalités attendues pour Gallica 2
Périodiques Presse Images Cartes Texte structuré Zoom Valorisation éditoriale et scientifique Recherche avancée Accès collections partenaires	Imprimés livres Téléchargement PDF Recherche simple Recherche plein texte Espace personnes Accès collections partenaires	Corrections Gestion de contenu Multilinguisme Espace collaboratif Vidéos Sons Moteur sémantique

4. Les missions du stage

Les missions du stage ont consisté à réfléchir sur la pertinence documentaire d'un corpus d'ouvrages en sciences médicales, à contribuer à sa constitution effective en respectant la politique de numérisation de la BnF et à mener des analyses afin de s'assurer de la cohérence du corpus tout au long du processus de sélection.

L'objectif final est de proposer un corpus cohérent qui sera, après une validation finale de la DCO et de Gallica, intégré dans la future Gallica 2. Le travail de sélection d'ouvrages a été réalisé dans le cadre du « marché des 100 000 » pour la BNuE.

Mon travail peut se décliner en cinq parties⁸ :

- Un premier travail en amont : état de lieu des collections disponibles en ligne et des collections patrimoniales, suivi par une analyse de leur classification en vue d'une préparation du processus de sélection des ouvrages ;
- La proposition d'une thématique médicale afin d'enrichir les futures collections en ligne et l'identification des cotes ou des tranches de cotes relatives à cette thématique ;
- La sélection concrète des ouvrages identifiés lors de l'étape précédente en vue d'une potentielle numérisation, en conformité avec des critères documentaires, juridiques et physiques ;
- Une analyse de la cohérence du corpus ainsi constitué et, dans un cadre plus général, de l'avancement du processus de sélection dans le Département Sciences et techniques ;
- Enfin, le stage m'a permis de travailler en binôme afin de poursuivre les sélections déjà commencées dans le département et de participer à l'expérimentation et à l'évaluation d'une méthode nouvelle de sélection des ouvrages.

J'ai également participé à trois formations internes. La première a été consacrée au projet de la Bibliothèque numérique Européenne au sein de la BnF, présenté par les directeurs des départements concernés. Les deux autres formations intitulées « Recherche sur Internet : veille documentaire » et « Etat matériel des collections imprimés », cette dernière ayant eu lieu sur le site de Bussy Saint George, m'ont été très utiles pour la mission qui m'a été confiée.

⁸ Une diagramme de Gantt de la fin du mémoire est dédié à la gestion du projet.

Partie 2 : La chaîne de traitement des collections scientifiques et techniques pour la BNUE

Après la présentation du grand projet de numérisation engagé par la Bibliothèque nationale de France, nous nous proposons de délimiter la place et l'apport de la Direction des Collections, Département Sciences et techniques, au sein de ce projet et de mettre en avant le travail intellectuel de constitution des corpus d'ouvrages en vue de leur intégration dans Gallica 2/Europeana.

Pour cela, une brève description des collections scientifiques nous a paru importante, avec un accent particulier porté sur les collections médicales ; nous décrirons également leur cotation, ce qui permet de les identifier aisément sur le rayonnage. Nous poursuivrons notre mission par une présentation de l'état de lieux d'Europeana dans le cadre du « marché des 30 000 » (en cours) et du « marché des 100 000 » (en phase initiale). Un court bilan des choix documentaires du Département Sciences et techniques sera dressé, ainsi qu'une justification du corpus à constituer. A cela s'ajoute la description du processus de sélection des ouvrages avec les contraintes documentaires, juridiques et physiques spécifiques.

1. Présentation des collections patrimoniales scientifiques et techniques

Les collections scientifiques et techniques sont moins nombreuses que celles concernant d'autres domaines, notamment les sciences humaines. Exhaustives ou presque en ce qui concerne les publications françaises, ces collections avaient été complétées par d'importantes acquisitions, dons ou échanges étrangers jusque vers 1920, ou à des dates variables suivant les disciplines. Depuis lors, les administrateurs généraux, parmi lesquels Julien Cain, mènent une politique de renoncement à des acquisitions étrangères dans les disciplines scientifiques et techniques, exprimant ainsi la conscience d'une rupture avec la politique documentaire du siècle précédent, essentiellement encyclopédique. L'interruption de ces acquisitions n'a pas permis à la Bibliothèque Nationale d'accompagner le développement des publications scientifiques, comme ont

pu le faire les grandes bibliothèques anglo-saxonnes. La politique documentaire de la bibliothèque a repris toutefois les acquisitions étrangères depuis 1990, d'une manière soutenue, qu'elles soient de monographies, ou d'abonnements à des périodiques, sous forme papier ou électronique, en sciences et techniques. Elle vise la constitution d'un fonds de référence comprenant les textes fondamentaux de la recherche contemporaine, tandis que certains domaines très spécialisés sont confiés aux pôles associés.

La création du Département Sciences et techniques, en 1994, est une preuve de la volonté de l'établissement de renouer avec la tradition encyclopédique de la Bibliothèque au XIX^e siècle.

1.1. Séries : R, S, T, V, Thèses

Tous les ouvrages qui sont entrés dans la Bibliothèque Nationale de France jusqu'à la fin de 1996 ont été classés selon la cotation Clément. Elle a été conçue par Nicolas Clément à la fin du XVII^e siècle. Les livres sont classés à l'intérieur de 23 divisions systématiques (de la lettre A – Ecriture sainte – à la lettre Z – Polygraphie et mélanges)⁹. Les fonds scientifiques et techniques sont répartis entre ceux qui sont entièrement consacrés aux sciences, comme S et T d'une part, et ceux qui sont en grande partie consacrés aux sciences, comme R et V d'autre part.

1.1.1. Séries S « Sciences naturelles »

La lettre S comprend le fonds inventaire (ouvrages entrés jusqu'en 1875) dont les premiers documents remontent au XVI^e siècle, et le nouveau fonds (ouvrages entrés de 1876 à 1996), soit environ 140 000 volumes. Divers domaines thématiques sont couverts parmi lesquels l'agriculture, les sciences biologiques y compris la botanique et la zoologie, les sciences de la terre et la paléontologie. Les collections existantes retracent l'histoire de ces disciplines. Avec l'émergence de nouvelles disciplines dans le domaine de la science de la nature, le fonds s'est enrichi, au XX^e siècle, d'ouvrages, essentiellement français, de génétique, biologie moléculaire et microbiologie.

La collection de sciences naturelles présente un grand intérêt historique et scientifique, et fait objet de recherche dans les études actuelles en stratigraphie, en biodiversité ou en biologie de l'évolution.

⁹ Nous ajouterons quelques éléments historiques complémentaires concernant ce plan de classement dans la section 1.2 de la deuxième partie.

1.1.2. Série T « Sciences médicales »

La Bibliothèque nationale de France possède un fonds d'une valeur inestimable de livres de médecine qui constitue le terrain d'investigation et d'analyse en vue d'une numérisation ultérieure à grande échelle. Le nombre d'ouvrages médicaux s'élève à 250 000 volumes dont 100 000 thèses. Une présentation détaillée des collections médicales sera réalisée dans la partie 1.2 de notre travail.

1.1.3. Série R « Sciences philosophiques, morales et physiques »

Les ouvrages regroupés sous la lettre R « Sciences philosophiques, morales et physiques » du classement Clément sont conservés dans le département Philosophie, Histoire, Sciences de l'Homme. La partie ancienne du fonds, qui comprend les livres entrés avant 1875, compte 55 127 volumes, tandis que le nouveau fonds ouvert par Léopold Delisle après 1875 compte plus de 21 500 volumes.

La série R rassemble des ouvrages de philosophie, de physique (y compris astronomie et chimie) et de pédagogie. Les mémoires de sociétés savantes occupent une place importante dans ce fonds. Les textes de physique ont été classés dans la lettre R jusqu'au 14 janvier 1997, date de lancement de la nouvelle cotation pour les monographies.

La physique constitue un cas particulier dans la répartition des collections entre les départements mais tout à fait explicable par son histoire. Elle était appelée « philosophie naturelle » au XVII^e siècle lorsque Nicolas Clément instaura son plan de classement. Elle était alors considérée comme faisant partie de la philosophie, qui était divisée dans l'enseignement des collèges jésuites en quatre parties : logique, physique, métaphysique et morale.

La « rupture » entre science et philosophie ne remontant qu'à deux siècles, leur histoire commune explique ce qui nous paraît comme une anomalie aujourd'hui, à savoir le regroupement au sein du même lettrage de disciplines comme la philosophie et la physique.

1.1.4. Série V « Sciences et arts »

L'intitulé « Sciences et arts » est conforme à l'idéal classique, qui ne sépare pas les beaux-arts des techniques, voire des sciences. C'est la raison pour laquelle l'on trouve, en V, des traités de mathématiques, des ouvrages sur l'architecture, la peinture, les techniques et les matériaux de construction. La série V fait l'objet d'un partage entre

deux départements : le fonds inventaire (ouvrages entrés jusqu'en 1875), qui compte environ 55 000 volumes sans les recueils, est conservé dans le département Littérature et Art ; le nouveau fonds (environ 250 000 ouvrages entrés après 1875), a été attribué au Département Sciences et Techniques. Il offre notamment des collections de manuels et de livres techniques et pratiques souvent introuvables par ailleurs.

A la lettre V se rattachent les recueils de pièces Vp et Vz, soit environ 31 300 pièces. En novembre 1994 a été créée la cote V2 pour anticiper la départementalisation des collections. Elle regroupe des publications concernant les arts majeurs et mineurs, architecture, peinture, sculpture, arts décoratifs, cinéma, danse.

Le fonds de la série V présente une forte diversification, due à l'essor de certaines disciplines comme les sciences de l'ingénieur et les applications des sciences fondamentales (chimie industrielle, microbiologie appliquée) sans oublier pour autant la cuisine, les jeux, le sport et le cinéma.

1.1.5. Thèses étrangères et françaises

Le fonds des thèses que la Bibliothèque nationale de France abrite est pluridisciplinaire et il comprend les thèses étrangères cataloguées Θ , Θg , THE, Th varia et les thèses françaises scientifiques et médicales.

La lettre Θ (thêta) couvre la période 1875-1975 et son classement est fait par format, ville de soutenance et discipline. A partir de 1975 et jusqu'à 1979, les thèses étrangères THE, environ 2267 volumes, sont classées par format et ville de soutenance. Depuis 1980 elles sont réparties dans les séries Clément correspondantes et proviennent, pour la plupart, des échanges internationaux.

La sous série Θg compte environ 14 795 thèses allemandes publiées pendant la première guerre mondiale. Le Th varia est un fonds clos d'environ 4529 thèses étrangères de médecine dont les premières remontent au XVII^e siècle, les dernières datant du début du XX^e siècle.

Les thèses scientifiques médicales françaises, couvrant la médecine, la pharmacie, la médecine vétérinaire et les sciences sont cotées de Th à Tk jusqu'en 1970 puis Thm, Thp, Thv et Ths.

Le Département Sciences et techniques reçoit sous forme de microfiche de l'Atelier National de Reproduction des Thèses (ANRT) un exemplaire de chaque thèse soutenue en France depuis 1986, à l'exception des thèses de médecine et à condition que la thèse ait été microfichée.

1.1.6. La série Z « polygraphie, mélanges »

Parmi le fonds dense et complexe de la série Z, on peut retrouver des collections scientifiques et techniques, car il regroupe souvent les ouvrages inclassables ailleurs, dépassant la seule représentation de la littérature.

1.1.7. Autres collections scientifiques conservées à la BnF

Les départements spécialisés de la BnF, des Estampes, des Cartes et Plans, et des Manuscrits conservent des documents spécialisés à caractère scientifique : la Division Occidentale des Manuscrits détient par exemple des manuscrits anciens en rapport avec la botanique, la zoologie et la médecine, ainsi que les carnets de Marie Curie et de Louis Pasteur.

Le Département des Cartes et Plans regroupe un fonds de cartes et globes célestes qui complète les fonds d'astronomie, ainsi qu'un fonds de cartes récentes topographiques, géologiques, de la végétation, d'images de télédétection qui complètent les fonds de géologie et de botanique.

Le Département de l' Arsenal conserve des fonds importants concernant la gastronomie, qui ont donné lieu en 2002 à l'exposition *Livres en bouche*.

Le département de l'Audiovisuel collecte et conserve des films documentaires scientifiques dans le domaine médical ou logiciel.

1.2. Les sciences médicales: cotation et catalogage

Une étude de la cotation et du catalogage des collections médicales complète notre présentation sur les collections scientifiques patrimoniales, car elle a une importance historique particulière. L'organisation et la classification des connaissances passe toujours par un système de cotation au sein d'une bibliothèque, et constitue un reflet fidèle de l'outillage mental de l'époque.

1.2.1. Le catalogage des collections de la section T : quelques repères historiques

Nicolas Clément, bibliothécaire de la Bibliothèque du Roi, a reçu en 1675 la mission d'organiser les collections royales suivant un cadre méthodique, inspiré de la classification en cinq catégories, définies en 1627 par Gabriel Naudé, à savoir : la Théologie, la Jurisprudence, l'Histoire, les Sciences et les Arts, les Belles Lettres.

Ainsi, 35 589 volumes ont été repartis en 23 divisions¹⁰, avant d'être inscrits sur les 7 registres in folio, ce qui constituera la première série du *Catalogus librorum impressorum bibliothecae regiae*. La lettre T est affectée aux sciences médicales. Dans chacune des 23 divisions désignées par des lettres, les livres étaient classés par format, puis par sujet et chaque volume était affecté d'un numéro (en vue d'accroissements futurs, des numéros avaient été laissés vacants).

Un récolement des collections réalisé en 1684 par l'abbé de Varès, à partir du premier catalogue de Clément, met en évidence la nécessité d'une révision de la rédaction des notices et du cadre de classement. Un deuxième catalogue, auquel la contribution de Jean Buvat a été importante, a été achevé en 1697 mais la lettre T n'a subi que de légères remaniements. Ce second catalogue donnera lieu à un nouveau récolement d'octobre 1719 à décembre 1720, à l'issue duquel la lettre T comptait 3 787 volumes¹¹.

À partir de 1730, le nombre des nouveaux ouvrages à intercaler ne cessa pas de croître, ce qui crée la nécessité d'une révision du cadre de classement de Nicolas Clément. Des nouvelles divisions sont introduites¹² et d'autres reclassés¹³. Cette entreprise aboutira avec l'impression du *Catalogue des livres imprimés de la Bibliothèque du Roi*. 1739-1753, 7 volumes in-fol., publication interrompue au début de la Révolution. Jusqu'au milieu du XIX^e siècle, les ouvrages entrés dans la Bibliothèque Nationale suivent les deux systèmes de cotation différents : celui de Clément pour les lettres F-V et le Z non classé, et celui du catalogue imprimé pour les lettres A-E* et X-Z.

Puis, les 300 000 ouvrages confisqués pendant la Révolution ont été insérés dans le cadre de l'ancienne Bibliothèque royale, qui comportait un fond « porté » et « non porté ».¹⁴

En 1852, sous la direction de Jules Taschereau (administrateur général de 1858 à 1874) les deux fonds « portés » et « non portés » du T et de sept lettres historiques¹⁵ sont entièrement recotés, suivant un ordre méthodique.

¹⁰ Religion (lettres A à D) ; Droit (lettres E et F) ; Histoire (lettres G à Q) ; Sciences et arts (lettres R à V) ; Littérature (lettres X à Z), chaque lettre recouvrant un domaine propre

¹¹ LEDOS, E. *Histoire des catalogues des livres imprimés de la Bibliothèque Nationale*. Paris. 1936, p. 62.

¹² D2 – Théologie hétérodoxe ; E* - Droit de la nature et des gens ; Y2 – Romans)

¹³ Reclassement des divisions A-E*, X-Y et une grande partie du Z, sans prendre en compte le format et affectant des numéros apr par ouvrage et non par volume.

¹⁴ Le fonds porté au catalogue et reclassé fait l'objet de la constitution des 53 volumes des carnets d'inventaire dressés de 1839 à 1844. Le fond non porté destiné uniquement aux recherches en magasin n'était ni coté, ni inventorié mais une seconde tranche a été ouverte dans chaque lettre, classée par format, puis par nom d'auteur ou par titre anonyme.

¹⁵ L – Histoire de France ; N – Histoire d'Angleterre, O – Histoire d'Espagne et du Portugal ; O2 – Histoire d'Asie ; O3 – Histoire d'Afrique ; P – Histoire d'Amérique, P2 – Histoire d'Océanie)

La même année a été décidée l'édition du *Catalogue des sciences médicales*, en même temps que celui de l'Histoire de France. Dans la lettre de présentation du premier volume du *Catalogue*, Taschereau explique que le ministre de Napoléon III, Hippolyte Fortoul, avait sollicité auprès de l'Académie de médecine un plan de classement, qui a été confié trois ans plus tard à Frédéric Dubois d'Amiens, agrégé de médecine et secrétaire perpétuel de l'Académie de médecine à partir de 1847.

1.2.1.1. Les subdivisions : Ta, Tb, Tc, Td, Te, Tf, Tg

Le classement de Dubois d'Amiens a bouleversé totalement celui établi par Nicolas Clément, de plus en plus difficile à appliquer et s'avérant incomplet (pas de place pour les Eaux minérales et l'Art vétérinaire, rattachement de la Parfumerie et de la Cuisine à la Chimie, ou mélange des ouvrages d'anatomie à des traités de chirurgie). Dubois d'Amiens va entreprendre un remaniement radical du classement Clément et une critique sévère de celui-ci, lors d'un Rapport officiel, en 1855.

Avec l'aide du Dr. Briau, bibliothécaire de l'Académie impériale de Médecine, Dubois D'Amiens a rédigé la première partie de son catalogue consacré à l'anatomie, la physiologie et l'hygiène. La plus grande part de la rédaction du catalogue a été confiée à A. Pauly, auteur en 1874 d'une *Bibliographie des Sciences médicales*.

Le *Catalogue méthodique des Sciences médicales* paraît en trois volumes de 1857 à 1889. Y était recensés 43 000 articles environ, répartis en 8 chapitres et subdivisés en 615 sections.

T	Préliminaire et Généralités	54 sections
Ta	Anatomie	64 sections
Tb	Physiologie	78 sections
Tc	Hygiène	55 sections
Td	Pathologie	140 sections
Te	Thérapeutique médicale	165 sections
Tf	Médecine légale et jurisprudence médicale	21 sections
Tg	Art vétérinaire	21 sections

Le *Catalogue* commence avec un chapitre général « Préliminaires et Généralités » réservé à l'histoire des sciences médicales (bibliographies, biographies, traités de philosophie, de littérature et de méthodologie médicale, polygraphies et traités généraux comprenant l'ensemble des sciences médicales). Une des particularités du *Catalogue*

choisie par Dubois d'Amiens consiste dans le placement des livres traitant de l'homme en état de santé avant ceux qui traitent des maladies.

Chaque section est rigoureusement organisée suivant le même principe. Par exemple Ta, l'Anatomie, est abordé sous une perspective générale (histoires, encyclopédies, dictionnaires, généralités, mélanges). Elle est suivie d'une section consacrée à l'anatomie générale, présentée dans l'ordre chronologique des publications. Il y a ensuite les autres sections spécialisées, qui considèrent une par une les différents systèmes et appareils de l'organisme humain jusqu'à l'étude individuelle des organes.

La partie réservée aux traités d'hygiène [Tc] comprend 55 subdivisions qui vont de l'histoire de l'hygiène [Tc1] au chapitre « Des sépultures au point de vue de l'hygiène publique » [Tc55], en passant par « Du célibat et du mariage au point de vue hygiénique » [Tc27]. L'examen du [Tc] présente quelques particularités par rapport aux autres sections : par exemple, à l'intérieur du chapitre consacré à l'hygiène privée, en plus de l'examen de l'hygiène des corps, des aliments, des sexes, des âges, avait été créée une section sur l'hygiène des conditions, qui retenait, entre autres, celle des grands et du clergé » [Tc 32] ou celle des « Gens adonnés aux travaux de l'esprit » [Tc 33].

A partir de la cinquième section, la Pathologie [Td], qui compte 37 295 notices, est l'étude des individus en état de maladie qui domine, partie qui se termine sur la médecine militaire. Les soins sont regroupés dans la section Thérapeutique [Te]. Si l'on s'intéresse par exemple à la lèpre, on trouve tout sur la description de cette affection en [Td 132] (Pathologie, Appareil Tégumentaire, Lèpre) et le traitement à y appliquer en [Te116] (Thérapeutique, Appareil Tégumentaire, Thérapeutique de la lèpre).

La Thérapeutique médicale [Te] constitue la partie la plus développée et la plus diverse du *Catalogue* et comprend 39 562 notices. Y sont réunis non seulement les traités consacrés aux différentes thérapeutiques pratiquées (médicale, chirurgicale) jusqu'aux plus marginales ou nouvelles (astrologique, magique, homéopathique) mais aussi la pharmacie, la physique et la chimie médicale, la zoologie ou botanique médicale et se conclut par l'hydrologie médicale (balnéologie, hydrothérapie, eaux minérales et bains de mer), les établissements et les stations correspondantes. L'examen des cotes [Td] et [Te] permet d'étudier les rapports de la société avec la maladie et la médecine de l'époque. Ainsi certaines sous cotes sont consacrées aux pathologies et aux traitements concernant certaines catégories sociales : le [Td37] est dédié aux maladies du gens du monde ou le [Td39] à celles des gens de mer.

Les deux dernières sections regroupent la Médecine légale, la Jurisprudence médicale, la Toxicologie dans [Tf] et l'Art vétérinaire dans [Tg]. Comme pour l'ensemble du catalogue, le schéma de présentation est similaire : de l'examen anatomique à l'action thérapeutique, en passant par les différentes pathologies.

Le fonds médical conservé à la BnF, dont les ouvrages datent du XVI^e siècle jusqu'à nos jours, permettent de suivre les grandes transformations de la médecine dans ces domaines.

1.2.1.2. *Les thèses médicales*

La BnF conserve une collection très importante de thèses médicales qui complète la division T et qui s'élève à environ 100 000 volumes. Elle comprend des thèses françaises et étrangères du XVII^e siècle et jusqu'à nos jours. Elles ne figurent pas dans le *Catalogue des Sciences médicales* et ne figurent dans la *Catalogue général auteurs* qu'à partir de 1960. Des fichiers classés par ordre alphabétique se divisent en deux sections, l'une jusqu'en 1936 et l'autre de 1937 à 1959. Les thèses étrangères postérieures à 1882 figurent dans les catalogues et fichiers généraux.¹⁶

1.2.1.3. *Fin du classement méthodique*

Le classement méthodique a été poursuivi jusqu'en 1925. A partir de cette date, la lettre T – *Sciences médicales* n'est plus une lettre cataloguée avec des subdivisions mais une lettre unique classée par ordre d'entrée et par format. Les nouvelles entrées sont donc traitées en lettres non cataloguées. En plus des formats-lettrages de lettres cataloguées qui ne s'accroissent plus, il y a huit formats-lettrage de lettres non cataloguées (de Gr. Fol. T. à 8-T. pièce) auxquels se sont ajoutés en 1943 les formats in-16, 16-T et 16-T pièce. Cette décision a été regrettée par certains érudits, pour lesquels les divisions du plan méthodique ne semblaient pas périmées malgré l'évolution des sciences¹⁷.

Le T est donc la seule lettre à comporter à la fois un fonds catalogué (jusqu'en 1925) et un fonds non catalogué. Une confusion n'est pas possible entre les deux fonds car les divisions cataloguées comportent, le plus souvent, une sous-lettre et un exposant [Te39]-Anesthésie.

¹⁶ *Les Catalogues du Département des Livres imprimés*. Paris : Bibliothèque Nationale, 1970. p. 26-27.

¹⁷ LEDOS, E., *op.cit.* p.181

2. Etat de lieux de la BNUe dans le Département Sciences et techniques

Au marché courant Gallica, qui prévoit une « numérisation seule » de 12 000 volumes par an par les prestataires Infotechnique pour la filière « papier » et Jouve pour la filière « microforme », s'ajoutent les deux grands marchés de numérisation en cours de traitement :

- Le « marché des 30 000 », lancé en 2006, concerne la numérisation+OCR+contrôle de 30 000 documents. La sélection des ouvrages, dont fait partie la *Sélection de printemps* et les *Listes « Europ »*, a été accomplie, et l'envoi des lots auprès du prestataire Jouve est en cours. En ce qui concerne les envois de la « filière papier » du Département Science et techniques, un premier lot de 97 UC, parti en mars 2007 reviendra dans les magasins mi-juillet. L'exemplarisation réalisé en grande partie par Van Dijick, est déjà visible dans BN Opale Plus, bien que la consultation des documents ne sera effective que dans quelques mois. Les UC parties sont bloquées à la communication 19 semaines environ, à partir de la date de départ des lots. Les deux lots suivants comprennent respectivement 127 UC et 188 UC et sont déjà partis pour la numérisation.

- Pour le « marché des 100 000 », les sélections dans la Direction des Collections (DCO) sont en cours ainsi que l'appel d'offre pour le prestataire extérieur. Les envois en numérisation commenceront en septembre ou octobre 2007 et il est prévu qu'un lot d'essai, qui inclura des cas particuliers, soit envoyé par le Département Philosophie, Histoire, Sciences de l'Homme avant le début des opérations. Etant donné la volumétrie à accomplir, pour laquelle la DCO ne s'est engagée que sur les minima, (45 000 documents la première année, 60 000 la deuxième et 75 000 la dernière année), les délais sont plus serrés par rapport aux marchés précédents.

2.1. Sélection de printemps

La « sélection de printemps DCO » comprend les propositions de numérisation faites au printemps 2006 par le Département Sciences et techniques et par d'autres départements de la DCO. Ces sélections concernent principalement les collections papier conservées dans les magasins, auxquelles s'ajoutent quelques microformes. En décembre 2006, une partie de ces propositions a été validée, et d'autres ont été repoussées à plus tard par la coordination DCO. Elles font l'objet d'une sélection fine, titre à titre dans le cadre d'un

corpus, à partir du catalogue BN Opale Plus, contrairement à la sélection de masse qui se réalise à partir de tranches de cotes.

Nous présentons un aperçu des propositions qui ont été déjà validées suite aux analyses concertées du Département Sciences et techniques et la DSR, ainsi qu'un dépouillement de ces corpus.

2.1.1. Les expositions universelles

Un corpus est consacré aux expositions universelles, notamment des catalogues généraux officiels des expositions de 1878, 1889 et 1900. Cela complétera les corpus déjà existant dans Gallica sur les « Expositions des produits de l'industrie française » (1798-1849) et sur les « Expositions universelles ». De plus, en étroite collaboration avec la BnF, le CNAM projette la numérisation des textes officiels concernant les expositions tenues à l'étranger, les actes de congrès, les comptes-rendus des délégations ouvrières et les périodiques liés à ces manifestations.

La DCO a conservé 53 cotes dans l'[8-V] dont:

- 7 documents déjà numérisés (disponibles dans Gallica, en général numérisés à partir d'autres exemplaires que ceux de la BnF) ;
- 35 ouvrages déjà reproduits sur microfiche ;
- 3 documents non numérisables : 2 pour des raisons d'état physique, 1 pour l'intérêt documentaire faible car une autre édition a été numérisée ;
- 5 ouvrages pour lesquels un visa positif a été donné.

Les documents microfichés ont fait l'objet de vérifications poussées concernant la qualité des microfiches afin de pouvoir obtenir un OCR correct, nécessaire à une recherche en texte intégral et une numérisation en couleur à partir des originaux.

2.1.2. Les Sciences naturelles

Le corpus de périodiques de sciences naturelles se compose de 33 titres mais le nombre de cotes est plus élevé que le nombre de titres, situation fréquemment rencontrée, en raison des périodiques de l'Inventaire [S], où chaque volume peut porter une cote différente. Les vérifications en amont visaient l'identification de séries exhaustives de périodiques numérisables, ce qui présente un fort intérêt documentaire. Toutefois, un bilan intermédiaire met en évidence 12 titres de périodiques non intégralement numérisables par rapport à un seul titre, *Annales des sciences géologiques* 1869-1889 [8-

S-657] (22 tomes) numérisable intégralement. Nous précisons que ce bilan n'est pas définitif car certains titres n'ont pas encore été intégralement vérifiés.

Le corpus sur les naturalistes français comprend, entre autres : des annales, bulletins ou mémoires de sociétés savantes ou associations (*Mémoires de la Société d'histoire naturelle de Paris* [S-4649], *Feuille des jeunes naturalistes* [4-S-192]) ; des périodiques de botanique comme *La revue générale de botanique* [8-S-1219] ; des revues dédiées à l'agriculture comme *Annales de l'agriculture française* [S-16680 à 16890], *Feuille du cultivateur* [S-4625 à 4634] ; des périodiques de zoologie, généraux (*revue et magasin de zoologie pure et appliqué* [S-11856 à 11885]) ou spécialisés (*L'Abeille : mémoires d'entomologie* [8-S-195]).

2.1.3. Les périodiques médicaux

Le corpus de périodiques médicaux comprend 12 titres, dont 2 numérisables intégralement selon les critères Gallica, à savoir : *Journal de l'anatomie et de physiologie normales et pathologiques de l'homme et des animaux 1864-1914* Paris ([8-T34-12], 50 UC, 34 300 images) et *Archives de physiologie normale et pathologique 1868-1898* [8-T34-15], 33 UC dont deux volumes de tables pour une estimation d'environ 27 700 images.

2.1.4. Autres propositions

Aux corpus présentés précédemment s'ajoutent des propositions de numérisation d'autres départements, dont une proposition pour les *Comptes-rendu de l'Assemblée générale des actionnaires de la Banque de France*, actuellement conservés sur 35 bobines de microfilm, une monographie et un périodique au thème « bibliophilie » proposés par Département de Littérature et Art ; une proposition pour *l'Encyclopédie méthodique* de Panckoucke, recommandé par le Département de la Recherche Bibliographique (DRB).

2.2. Listes « Europ »

Le corpus Bibliothèque française sur l'Europe a été constitué selon des équations de recherche dans le catalogue de la BnF portant sur tous les titres contenant le terme europ*. Il s'agit de documents microformés et de documents originaux des collections du Département Sciences et techniques. Les vérifications menées au sein du département

ont conduit à un taux d'acceptation de 77 % pour les documents microfilmés et de 31 % pour les documents originaux.

Après les vérifications supplémentaires effectuées avec la Direction des Services et Réseaux (DSR), la répartition par format-lettrage des UC numérisables et ayant été acceptées, se présente ainsi :

Format lettrage	% par rapport au nombre total d'UC numérisables	% par rapport au nombre total de pages
Inventaire S	29 %	40 %
8-V et 8-Vpièce	21 %	16 %
T	18 %	15 %
Fol-S et 4-S	8 %	6 %
8-S et 16-S	8 %	16 %
4-V et Fol-V	8 %	6 %
Sp	8 %	1 %

Le tableau démontre le fait que la moitié des documents numérisables, en nombre d'UC et en nombre de pages, provient des inventaires [S] et [8-V], riches en illustrations et planches illustrées en couleur ou en noir et blanc.

Les refus de 16 % pour les documents microfilmés et de 61 % pour les documents originaux des collections scientifiques sont motivés essentiellement par l'état physique des documents, une des causes étant la reliure trop serrée, comme c'est le cas pour plus de 40 % des refus dans l'Inventaire [S]. D'autres causes de refus, toujours liées à l'état physique des documents sont : l'atteinte à l'intégrité du corps d'ouvrage (pages et cahiers déchirés) ou l'état avancé de l'acidité du papier. Ces critères qui concernent 60 % des refus du fonds du [8-S] sont considérés comme bloquants par Gallica.

Les autres refus sont dus aussi bien aux problèmes documentaires (extraits et tirés à part, pertinence documentaire, édition à l'identique, publications étrangères), qu'aux critères juridiques critères (ouvrages sous droit) ou aux problèmes d'accès aux collections (dépoussiérage, manque en place, conservation dans un autre département).

2.3. Justification et choix d'un corpus

Si pour le « marché des 30 000 » la sélection des ouvrages dans le cadre du projet Europeana a été accomplie, pour le « marché des 100 000 » seuls les axes documentaires ont été définis. Les choix portent sur trois thématiques :

- Les sciences naturelles (sciences de la vie et sciences de la terre) au XIX^e siècle : contribution française au développement et à la diffusion des connaissances dans ces domaines ;
- La médecine française au XIX^e siècle : la maladie et son traitement ;
- Les expositions universelles au XIX^e siècle, vitrines du progrès et instruments de diffusion des connaissances.

Les trois corpus rappellent les trois grands axes de développement prédéfinis pour la Bibliothèque numérique Européenne : le patrimoine national, les programmes à vocation internationale et le rayonnement européen. Les corpus mettront en avant la contribution française au développement et à la diffusion du savoir dans les sciences naturelles, les sciences médicales et les sciences appliquées.

L'élargissement du corpus à d'autres domaines des sciences constitue un des objectifs de la politique documentaire de la BNuE afin de mieux faire connaître l'apport français au domaine de la connaissance scientifique.

Tout type de document intellectuel est potentiellement numérisable : des ouvrages de référence, textes, revues, manuels d'enseignement, des cours ou des outils comme les dictionnaires et les annuaires. La collection impressionnante de périodiques, d'extraits de périodiques, de pièces et de recueils constituée au long des siècles, en grande partie grâce au dépôt légal, accompagne l'essor des disciplines scientifiques et leur haute spécialisation à partir de la fin du XVIII^e siècle. Une grande partie des auteurs, en particulier des médecins, ne publient leurs travaux que sous forme d'articles de périodiques.

Notre travail se construit autour d'un élargissement du corpus sur la maladie et son traitement au XIX^e siècle, évoqué précédemment, en proposant un regard historique et philosophique (donc interdisciplinaire) sur la médecine française à travers les collections conservées dans le Département Sciences et techniques. Dans cet objectif d'élargissement, nous nous sommes particulièrement intéressés aux représentations de la douleur et de la maladie, à une réflexion philosophique sur la souffrance, aux théories

médicales élaborées pour en élucider les mécanismes ainsi qu'aux thérapeutiques développées pour soulager le patient. Cela explique notre focalisation sur la division destinée à la Philosophie médicale [T19] et à l'Anesthésie [TE39]. Nous avons également retenu la division [T2], qui offre l'avantage d'une perspective historique limitée à certaines époques, notamment au XIX^e siècle.

La dimension pluridisciplinaire de la thématique abordée s'inscrit dans les directives de la *Charte documentaire de Gallica* et complète les corpus déjà disponibles en ligne en médecine, en histoire et en philosophie.

2.3.1. Philosophie médicale

Dans le *Catalogue des sciences médicales*, 180 notices bibliographiques correspondent à la subdivision Philosophie médicale [T19] de la Section V Philosophie, Méthodologie et Littérature, Erudition et Critique Médicale, du chapitre « Préliminaires et généralités ». Entre 1857, date de parution du *Catalogue* et 1925, la cote [T19] s'est enrichie de 355 notices.

La « philosophie médicale » ne figure pas dans l'index matière de Gallica.

Dans le catalogue en ligne de Gallica, les requêtes peuvent porter sur le « mot du titre », « l'auteur », « le sujet » ou sur le champ d'une « recherche libre ». Il n'y a pas d'opérateurs booléens. Les recherches croisées sont possibles en remplissant plusieurs champs avec les mots-clés.

Le tableau ci-dessous présente un court bilan des résultats obtenus avec Gallica le 1 mars 2007 :

Descripteurs et type de requête	Nombre de notices bibliographiques
philosophie médec* (<i>Titre</i>)	13 notices
« philosophie médicale » (<i>Recherche libre</i>)	11 notices
Philosophie (<i>Sujet</i>) + médec* (<i>Recherche libre</i>)	27 notices
Philosophie (<i>Titre</i>) + médecine ou médec* (<i>Sujet</i>)	9 notices

Les résultats obtenus nous prouvent que la problématique est traitée modestement dans Gallica et qu'une numérisation de ce fonds enrichira les collections patrimoniales déjà présentes en ligne. La philosophie de la médecine constitue une piste intéressante de recherche interdisciplinaire, reflet de l'effervescence scientifique dans les sciences en général et en sciences médicales en particulier, au XIX^e siècle.

Nous essayons par la suite de situer les collections de philosophie et d'histoire médicale conservées à la BnF dans leur temporalité et dans leur contexte historique, du XV^e siècle au début du XX^e siècle.

Au XV^e siècle, la médecine reste associée à la philosophie. Elle n'a pas encore conquis son individualité et les médecins doivent plus leur diplôme à la philosophie, la théologie, les mathématiques, l'astronomie, l'optique qu'à la médecine elle-même.

Parmi les figures médicales françaises de la Renaissance, Ambroise Paré (1509-1590) et Jean Fernel (1497-1558) avec son *Universa Medicina*, très souvent rééditée, occupent des places d'honneur dans la bibliothèque numérique Gallica.

Au début du XVI^e siècle les livres français sont principalement imprimés à Lyon (par exemple *Controversiarum medicarum et philosophicarum Francisci Vallesii Covarruviani editio 3a, ab auctore denuo recognita et aucta. Accessit libellus de Locis manifeste pugnantis apud Galenum, eodem Vallesio auctore*, publié en 1591 (Lugduni) par Francisco Vallés de Covarrubias mais l'influence lyonnaise ne durera pas longtemps, car à partir du XVII^e siècle les principaux ateliers d'impression médicale se situent à Paris.

Si le XVII^e siècle reste encore sous l'empreinte idéologique de l'Eglise, les médecins voient leurs analyses limitées par l'absence d'une instrumentation adéquate et recourent souvent à des constructions intellectuelles qui faussent l'interprétation des faits observés. Le XVIII^e siècle divise le domaine médical entre les mécanistes et les vitalistes et c'est à Montpellier que l'école du vitalisme se développe. Pour Théophile de Bordeu (1722-1776) et Paul-Joseph Barthez (1734-1806), des adeptes du vitalisme, la vie a un caractère irréductible qui ne peut être expliquée ni par animisme, à cause des échanges physiques et chimiques existants, ni par mécanisme, jugé trop simpliste. Ces systèmes de pensée alimentent une riche réflexion philosophique, avec des répercussions sur les époques futures, mais elles restent sans aucun effet sur le traitement des malades. On remarque néanmoins un effort de classification des maladies, inspiré du travail de Linné pour les plantes et les animaux, travail qui influe profondément les savants de l'époque.

Le bilan de l'état de la médecine à la fin du XVIII^e siècle marque un mauvais état sanitaire de la population et une mauvaise réputation de l'enseignement et des enseignants. Diverses décisions prises pendant la Révolution s'avèrent être désastreuses : nationalisation des hôpitaux, fermeture des universités, interdiction des corporations et des sociétés savantes.

Le XIX^e apporte un changement dans le statut de la médecine. La santé publique et la gestion des hôpitaux deviennent une affaire de l'Etat, les Sociétés savantes sont reconstituées, la pédagogie médicale subit une vraie réforme. Trois écoles de médecine sont créées à Paris, Strasbourg et Montpellier. L'Académie de médecine débute ses enseignements en 1820 et la Chaire d'Histoire de la Médecine de la Faculté de Médecine de Paris se réjouit d'être la première créée au monde, bien qu'elle subira des vicissitudes tout au long au XIX^e siècle.

La division [T19] constitue une image de cette évolution historique. Elle comprend par exemple une série de cours de médecine comme *Programme du cours de philosophie médicale, professé à la Faculté de Strasbourg* en 1844 par Dr Charles-Polydore Forget [8-T19-140], des discours ou des lettres prononcés à l'ouverture des enseignements universitaires, comme *Académie de médecine. Séance du 25 août 1874. Discours sur l'imagination dans ses rapports avec la philosophie et la médecine* par Dr Paul Jolly [8-T19-293] ou *Des Rapports de la médecine et de la philosophie, discours prononcé le 21 décembre 1899 à la séance solennelle de rentrée de l'École de médecine et de pharmacie d'Amiens* par (Dr) Louis-Sylvain Dewèvre [8-T19-376].

Au début du XIX^e siècle, les médecins prennent conscience d'une rupture au sein de la médecine, qui devient une discipline véritablement scientifique, grâce au développement de la méthode anatomo-clinique et à l'essor de la bactériologie par laquelle la médecine devient une science de laboratoire. L'évolution de la médecine dépend aussi du développement économique, social et politique de l'époque.

Ce sujet est fructueux d'un point de vue de son étude philosophique car toute rupture ou révolution scientifique est accompagnée d'une remise en question des fondements et des causes d'un tel bouleversement.

2.3.2. L'anesthésie

Dans les dernières décennies du XIX^e siècle un des changements majeurs dans les sciences médicales vient de la reconsidération de l'individualité du patient et des rapports entre celui-ci et le médecin. Cela conduira à l'apparition de l'anesthésie comme

discipline médicale scientifique, qui a pour but de rendre le patient insensible à la douleur engendrée par les actes médicaux. Ce but s'inscrit dans la tradition hippocratique *Primum sedare dolorem* (« Avant tout calmer la douleur »). Grâce à Lavoisier on sait que le poumon est capable d'absorber des gaz qui pénètrent dans l'organisme. Toutefois, l'adoption de l'anesthésie en chirurgie (anesthésie locale ou générale) par l'utilisation de l'éther, chloroforme ou protoxyde d'azote, est un sujet de débat assez controversé à l'Académie des Sciences pendant les années 1847 et 1848, entre Velpeau et Magendie notamment. Ces discussions sont reflétées dans les collections de la BnF par une abondance d'articles publiés sur les pratiques et les modalités d'exercice de l'anesthésie ou sur des recherches effectuées sur les voies d'introduction de l'anesthésie. Cela illustre une phase initiale dans la constitution de la discipline. On retrouve d'ailleurs dans cette division un ouvrage publié à Paris en 1847 qui fera date : *Mémoire sur la découverte du nouvel emploi de l'éther sulfurique* [8-TE39- 6] par William Morton, dentiste à Boston, avec lequel on marque la naissance de l'anesthésie.

Une note présentée par Velpeau et relayée par Paul Broca devant la Société de chirurgie le 5 décembre 1859, intitulée *Sur l'Anesthésie chirurgicale hypnotique* [8-TE39-60] constitue une ouverture vers le sujet controversé de l'hypnotisme et de son statut. La BnF en conserve précieusement les traces et nous la proposons pour la numérisation.

Dans la cote [TE39] il y a 274 notices en BN Opale Plus dont 1 ouvrage numérisé.

Un court bilan sur « l'Anesthésie » dans la bibliothèque numérique Gallica, consultée le 1 mars 2007, nous confirment que le sujet est modestement traité, avec peu de références :

<u>Descripteurs et type de requête</u>	<u>Nombre de notices bibliographiques</u>
anesthésie (<i>Index sujet</i>)	1 notice
anesthésie (<i>Titre</i>)	2 notices

Présentant un rapport avec la douleur et la souffrance, sujets de réflexion philosophique par excellence, l'anesthésie [TE 39] rejoint notre premier choix, Philosophie médicale [T19].

2.3.3. La dimension historique du corpus

La division [T2] (*Histoire des sciences médicales limitée à certaines époques*) est composée de 70 notices bibliographiques dans BN Opale Plus et elle vient intégrer des corpus déjà constitués sur la médecine du XIX^e siècle, comme par exemple *Histoire de la révolution médicale du XIX^e siècle* par Almire Lepelletier (1790-1880), ou celui sur les Expositions universelles : *La Médecine à l'Exposition universelle* de 1878 par (Dr) André-Justin Martin. Cette division offre une dimension historique complémentaire pour notre corpus.

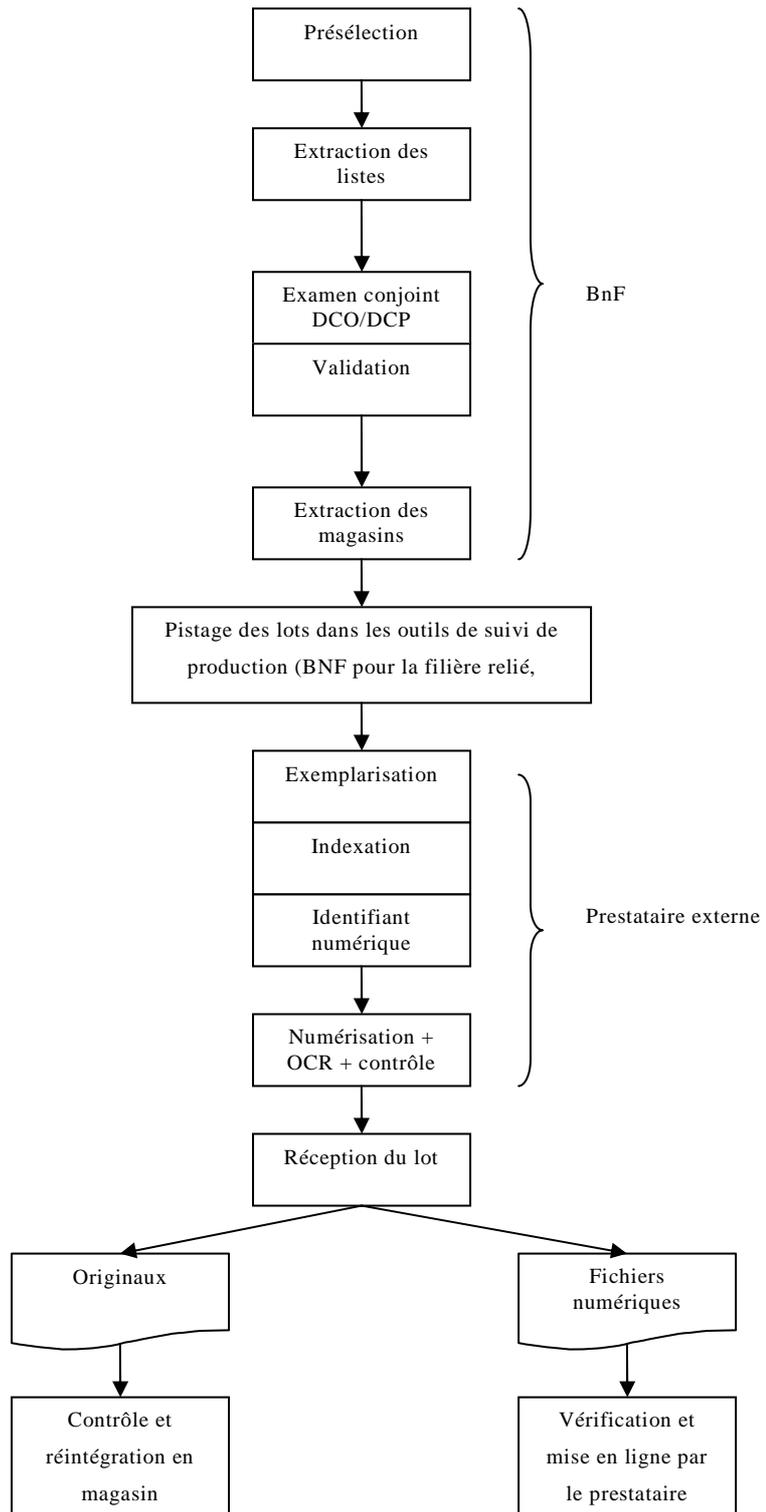
3. Processus de sélection

Les volumétries prévues pour le « marché des 100 000 » (voir Annexe 2) entraînent un rythme de sélection assez soutenu au sein de la DCO. L'outil d'aide à la sélection ADCAT 15 sera disponible fin juin 2007. Entre temps, l'examen de sélection des ouvrages est effectué sur des fichiers Excel¹⁸, générés par le Département des systèmes d'information (DSI). La coordination DCO donne des consignes précises¹⁹ pour leur manipulation. Après le traitement des fichiers, selon des critères juridiques, documentaires et physiques bien définis, la DCO procède aux validations des fichiers de sélection. Le processus de sélection est le fruit d'un intense travail collaboratif via Lotus Notes (la base de production de la BnF) qui supposent un enchaînement de vérification, une répartition des tâches entre le département et le service de coordination ainsi que l'archivage des listes selon leur état (« initial », « en cours de vérification », « validé, prélèvement »).

Le diagramme ci-dessous décrit les étapes du processus de sélection :

¹⁸ Exemple de fichier Excel dans l'Annexe 4.

¹⁹ Annexe 5



3.1. Critères juridiques

L'application des lois sur le droit d'auteur (loi du 11 mars 1957, modifiée en 1985, loi du 1^{er} juillet 1992 portant sur le code de la propriété intellectuelle et loi DADVSI du 1^{er} août 2006) doit faire l'objet d'une vigilance particulière pour la mise en ligne des œuvres. Le choix de numérisation porte essentiellement sur les œuvres du domaine public mais des discussions sont engagées auprès des éditeurs ou des ayant droits car l'un des objectifs de la politique de la BNUe est d'ouvrir la possibilité de consultation des ouvrages sous droit.

L'œuvre tombée dans le domaine public est celle dont le droit patrimonial ne peut plus être exercé parce qu'arrivé à échéance (70 ans après la mort d'auteur). Mais le droit moral demeure toujours sur toute œuvre. Celui-ci ne peut être dénaturé et on doit toujours attribuer l'œuvre à son auteur.

3.1.1. Droit d'auteur pour les monographies

L'article L.123-1 du Code de la Propriété Intellectuelle (CPI) précise la durée des droits patrimoniaux et dispose que « l'auteur jouit, sa vie durant, du droit exclusif d'exploiter son œuvre sous quelque forme que ce soit et d'en tirer un profit pécuniaire. Au décès de l'auteur, ce droit persiste au bénéfice de ses ayants droit pendant l'année civile en cours et les soixante-dix années qui suivent ». Cette durée de protection peut être prolongée suite à des périodes de guerre, comme dispose l'article L.123-8 et L.123-9 : 6 ans et 152 jours accordés en plus pour les œuvres qui ne sont pas tombées dans le domaine public le 3 février 1919 et ont été divulguées entre le 2 août 1914 et le 31 décembre 1919 ; 8 ans et 120 jours accordés pour les œuvres qui n'étaient pas tombées dans le domaine public au 13 août 1941 et qui ont été divulguées avant 1^{er} janvier 1948.

Les créations des auteurs morts pour la France (et reconnus comme tels par décret) sont protégées pendant une période supplémentaire de trente ans, aux termes de l'article L. 123-10 du CPI.

Toutes ces prorogations exceptionnelles peuvent se cumuler. C'est par exemple le cas de l'œuvre d'Alain-Fournier, mort au combat pendant la Première Guerre mondiale, qui bénéficie encore d'une longue période de protection.

Lorsqu'une œuvre est créée à partir des interventions de plusieurs auteurs, le droit d'auteur suit des régimes particuliers selon qu'il s'agisse d'une œuvre de collaboration, d'une œuvre composite ou d'une œuvre collective. Le droit d'auteur concerne également

les auteurs secondaires (pré/postfacier, illustrateur, traducteur), un examen sur ceux-ci est donc également nécessaire.

Les vérifications concernant le droit d'auteur effectuées par Gallica et poursuivies par le service juridique doivent être très strictes bien qu'on ne dispose pas toujours des éléments bibliographiques permettant une telle analyse.

Pourtant, la numérisation de masse impose en pratique une limitation des recherches dans la vérification des droits, sans toutefois prendre de risque inconsidéré :

- les ouvrages publiés antérieurement à 1870 sont par défaut considérés comme faisant partie du domaine public. Pour le « marché des 100 000 », la barrière pourra être étendue aux ouvrages publiés avant 1900, notamment pour les auteurs peu connus. Pour tout autre cas (auteurs ou illustrateurs connus) les vérifications sont plus poussées ;
- pour les monographies ayant plus de quatre auteurs principaux, peu connus, la date de publication du titre fera foi.

Gallica assume toute prise de risque sur ces critères et elle s'engage à ce que toute revendication légitime d'ayant droit donne lieu à la suppression du titre numérisé de la bibliothèque numérique de la BnF.

3.1.2. Droit d'auteur pour les périodiques

Les périodiques, assimilés à la BnF à des œuvres collaboratives, tombent dans le domaine public 70 ans échus après la date de publication des articles. Le problème est complexe essentiellement pour les titres dont les fascicules ont commencé à être publiés après 1930 et pour lesquels une numérisation partielle ne présente pas de réel intérêt documentaire, dans le cas où des exemplaires seraient encore publiés après 1937.

Quelques cas exceptionnels ont cependant été admis par la DCP afin de poursuivre la numérisation au-delà de 1937 :

- périodiques qui disparaissent au moment de la seconde guerre mondiale, pour lesquels il est préférable de numériser la série complète. La mise en ligne se réalisera d'une manière progressive ;
- périodiques qui alimentent le portail Cairn, dont la numérisation et la mise en ligne s'étendra aux périodiques édités jusqu'en 2000. Les titres sont fixés en accord avec les éditeurs (par exemple pour le « marché des 30 000 » : *Etudes*, *Revue d'histoire moderne et contemporaine*) ;

- périodiques pour lesquels la mise en ligne a été autorisée par l'éditeur ou la société rédactrice, qui prend en charge la question des droits d'auteur (*Comptes rendus de l'Académie des sciences ou Revue d'histoire littéraire de la France*).

3.2. Critères documentaires

Les critères documentaires pour BNUE/Gallica 2 ont été définis et leur mise en pratique consiste à mener des vérifications préalables dans le catalogue BN Opale Plus. Les ouvrages à exclure sont ceux qui répondent à un ou plusieurs des critères suivants :

- publication étrangère (le latin et le grec sont acceptés ainsi que les traductions en français des textes de référence, selon des listes d'auteurs établies en commun au préalable) ;
- document déjà numérisé ou pris en charge par d'autres programmes de numérisation (BIUM, CNUM) ;
- extrait de revues, car la politique de la BnF est une numérisation au volume²⁰ et pas à l'article, comme certains autres programmes ;
- réimpressions multiples à l'identique. La numérisation des éditions successives peut s'avérer par contre d'un réel intérêt documentaire car elles sont revues, augmentées ou corrigées donc accompagnées d'un appareil critique et d'un suivi de l'époque.

Concernant les monographies et les périodiques en plusieurs volumes, il est préférable de ne pas numériser des tomes isolés mais la série intégrale, à moins que ce soit dans le but de compléter des lacunes ponctuelles existantes dans Gallica.

3.3. Critères liés à l'état physique du document

3.3.1. Critères physiques bloquants

Dans le cadre du « marché des 30 000 » le cahier des charges impose contractuellement trois critères principaux. Il faut mentionner qu'à l'heure actuelle l'appel d'offre est en cours pour le « marché des 100 000 ». Nous présentons ci-dessous les critères physiques de numérisation pour le « marché des 30 000 » et précisons lorsque c'est possible les décisions prises pour le « marché des 100 000 » :

²⁰ A la BnF un volume est défini par sa page de titre et sa table de matières. Si plusieurs livraisons d'un périodique possèdent chacun leur table des matières et sa page de titre et sont reliés ensemble, l'unité numérique sera liée à chacun des fascicules et non au volume physique (UC de conservation).

- le format des documents est majoritairement de l'ordre du A4 mais il est possible d'aller jusqu'au format A2. Cette limite est valable pour les dépliants à l'intérieur de l'ouvrage et il n'y a pas de contraintes de taille pour les petits formats. Par contre en ce qui concerne le « marché des 100 000 », la limite est repoussée au format A1 et les nains sont exclus de la numérisation ;
- l'ouvrage doit permettre une ouverture de minimum 120°. D'une manière exceptionnelle et pour les documents particulièrement importants une ouverture de 90° peut être acceptée. En effet, dans ces cas là, l'ombre portée, peut altérer la lisibilité du texte et donc la conversion OCR.
- les documents doivent avoir une marge intérieure de chaque côté de minimum 1 cm.

Ces critères sont imposés par un contrat signé entre la BnF et le prestataire extérieur, Jouve en l'occurrence pour le « marché des 30 000 », qui est responsable des éventuelles détériorations physiques des livres subies pendant les manipulations de numérisation. Il refuse en conséquence les documents précieux et les documents déjà abîmés qui pourraient se détériorer davantage pendant l'accomplissement de la mission.

Dans le processus de vérification des documents sélectionnés, ces critères supplémentaires ont pour conséquence d'éliminer des lots de numérisation les ouvrages atteints par l'acidité du papier, lorsqu'elle va jusqu'à l'effritement des pages. Il en est de même pour les ouvrages dont l'intégrité du corps du volume n'est plus assurée (pages ou des cahiers volants ou déchirés).

3.3.2. Critères physiques non bloquants

Il y a des critères physiques non bloquants pour la numérisation qui peuvent entraîner une altération de la qualité des images ou du texte, mais qui sont acceptés par le prestataire s'il s'agit du seul exemplaire disponible dans la bibliothèque.

Ces critères concernent :

- les ouvrages présentant des tâches ou des rousseurs, sauf si le texte est illisible ;
- les ouvrages présentant des contrastes hétérogènes d'une page à l'autre ou au sein d'une même page ;
- les ouvrages ayant des transparences recto-verso, qui nécessitent une numérisation en niveau de gris, plus onéreuse.

Partie 3 : Analyse, évaluation et bilan du corpus étudié

Dans un premier temps, nous proposerons une analyse statistique descriptive du corpus présenté dans la partie précédente – que nous qualifierons dans la suite de « corpus de Philosophie médicale », à chaque étape du processus de sélection : avant tout examen, après l'examen documentaire et après l'examen physique. Cela nous permettra par exemple d'identifier des catégories d'ouvrage qui ont un taux de rejet supérieur à la normale et d'en expliquer les causes.

Par suite, nous analyserons et interpréterons les résultats définitifs de la sélection, après les vérifications supplémentaires de Gallica et la validation des fichiers par la coopération DCO. Une comparaison sur l'avancement du processus de sélection entre les départements de la DCO sera brièvement dressée ainsi qu'un état des lieux des sélections dans le Département Sciences et techniques.

Enfin, nous présenterons les pistes de travail pour un éventuel élargissement du corpus à d'autres sections du *Catalogue des Sciences médicale* ou à des thématiques proches d'autres programmes de numérisation en cours.

1. Bilan des sélections

1.1. Présentation du corpus

Le corpus de Philosophie médicale retenu pour l'analyse, avant tout travail de sélection, est constitué au total de 792 UC, réparti en 3 cotes : [T2] : Histoire des Sciences médicales limitée à certaines époques (8% du total), [T19] : Philosophie Médicale (60% du total) et [Te39] : Anesthésie (32% du total).

L'ouvrage le plus ancien date de 1548 : Petrus de Abano. *Conciliator controversiarum quae inter philosophos et medicos versantur, Petro Abano,... auctore. Libellus de venenis, eodem auctore. Petri Cararii quaestio de venenis, ad terminum. Symphoriani Champerii... in Conciliatorem cribationes*, [FOL-T19-5], et le plus récent de 1931 : Lalanne, Justin. *Notice complémentaire sur les inondations de l'Adour* [8-T19-187].

Nous avons représenté dans le graphique ci-dessous la distribution des ouvrages selon leur date d'édition, en 10 intervalles de même durée :

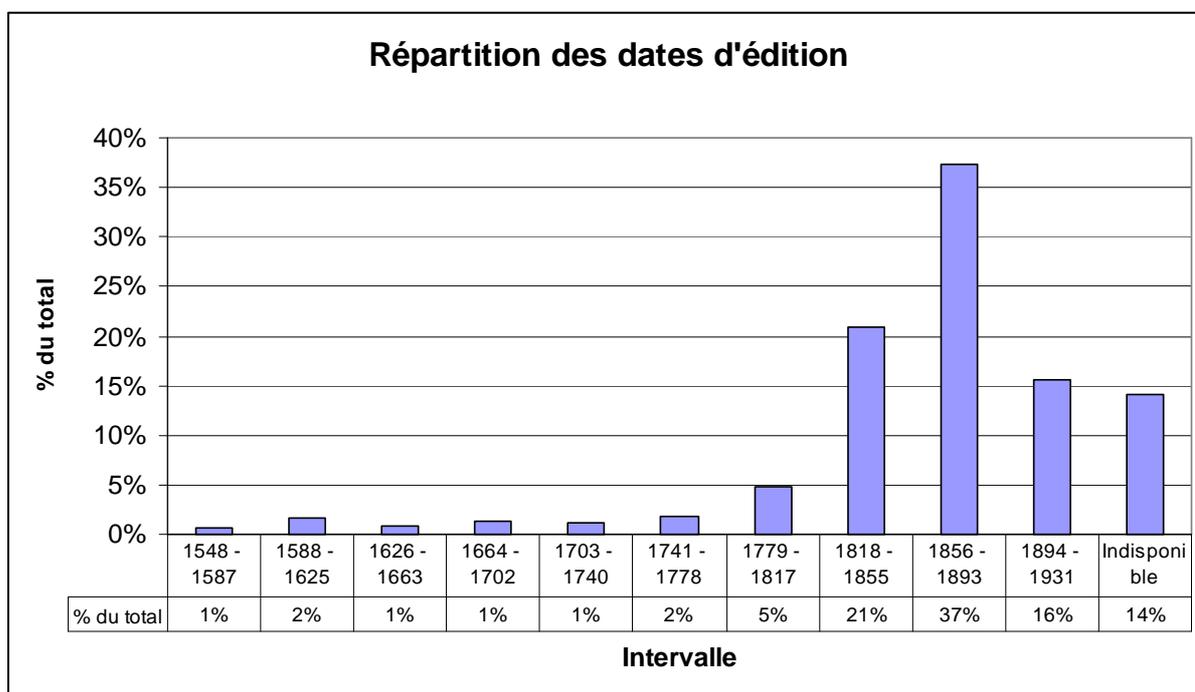


Figure 1

Il y a 14 % des livres dont la date d'édition n'est pas renseignée dans les extractions fournies, ensemble qu'on a dénommé « indisponible ». On constate que 74% des ouvrages sont concentrés entre 1818 et 1931. Cela conforte le choix de numérisation définie dans la *Charte documentaire de Gallica*, qui accorde une priorité aux ouvrages du XIX^e siècle, pour plusieurs arguments :

- un essor des sciences se produit à cette époque dans plusieurs disciplines ; il s'agit donc d'une période particulièrement recherchée par les historiens et prolifique en ouvrages scientifiques, notamment en ce qui concerne la production française ;
- les ouvrages sont dans leur très grande majorité libres de droits ;
- les ouvrages de date d'édition antérieure sont édités dans une typographie ancienne, peu adaptée au processus de reconnaissance optique des caractères, ou voient également une proportion importante de publications étrangères, qui ne rentrent pas dans les critères définis par Gallica 2/Europeana.

Nous nous intéressons plus particulièrement à la période 1818–1931, dont nous donnons dans le graphique suivant une répartition selon 10 intervalles :

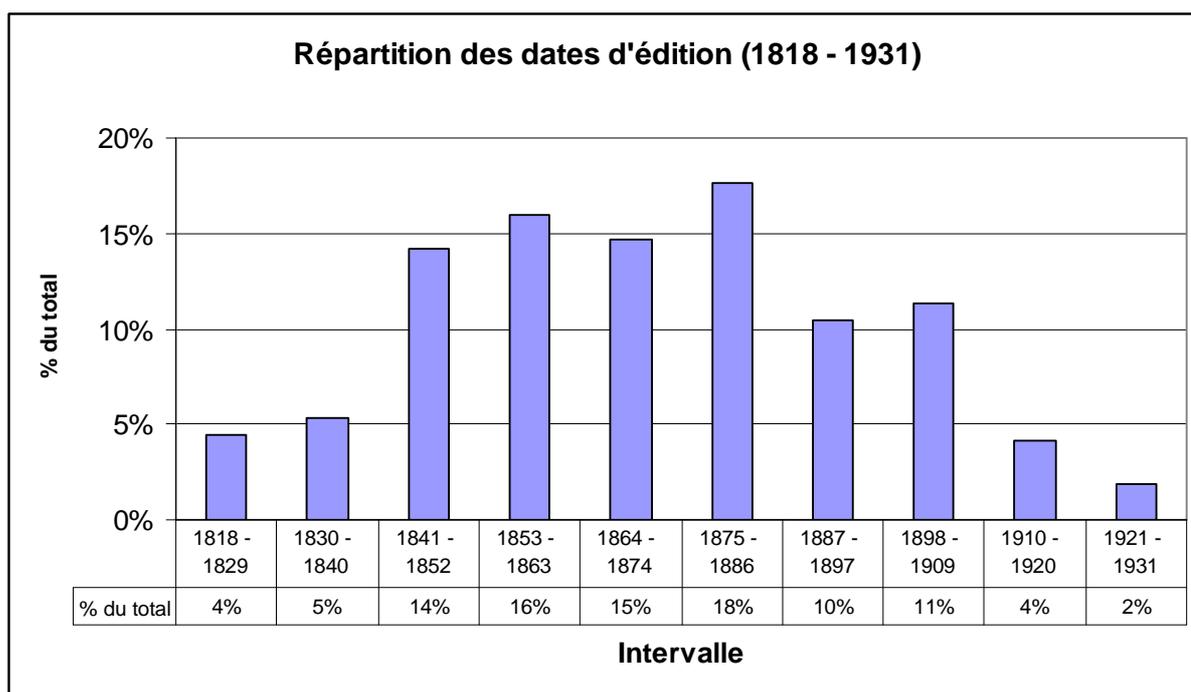


Figure 2

Nous constatons une concentration des dates d'édition dans la deuxième moitié du XIX^e siècle. Comme Henri-Jean Martin le souligne²¹, quelle que soit la série statistique à laquelle on fait appel, la *Bibliographie de la France* ou le *Catalogue de la librairie française*, la production imprimée française dans son ensemble s'enfle entre 1840 et la Première Guerre mondiale. C'est notamment grâce aux développements de la presse mécanique à vapeur, de la machine à papier continu ou de la reliure industrielle qu'on assiste à une industrialisation de la fabrication du livre. Le nombre des titres enregistrés par le Dépôt légal est de trente-trois mille en 1913, ce qui représente une multiplication par 5,5 du chiffre de 1840. Toutefois le début du XX^e siècle enregistre une baisse dans la production éditoriale française, malgré les années isolées de reprise en 1909 et surtout en 1913. Henri-Jean Martin parle même d'une « crise du livre » et d'un déclin de la production entre les deux guerres mondiales.

La Figure 2 reflète ainsi cette trajectoire éditoriale, avec des spécificités propres qui tiennent notamment des transformations qui ont marqué la médecine à cette époque :

- l'unification du doctorat en médecine, permettant de proposer les mêmes ouvrages aux étudiants dans toutes les facultés, d'où un agrandissement massif des tirages ;

- la francisation des livres médicaux, définitivement débarrassés de leurs références latines et grecques ;
- la laïcisation des hôpitaux et du personnel soignant, entraînant l'utilisation d'un langage uniquement scientifique.

1.2. Corpus après l'examen documentaire

Après avoir analysé la composition du corpus avant tout examen, nous nous intéressons aux résultats après la première étape d'analyse, l'examen documentaire.

1.2.1. Répartition par critère documentaire

L'examen documentaire consiste en une revue des caractéristiques non physiques des livres afin de vérifier s'ils satisfont les critères documentaires ou de droits définis pour Gallica 2. L'outil dont nous nous sommes servis pour effectuer le travail d'analyse est le catalogue BN Opale Plus dans son application professionnelle, DCAT-02.

Nous présentons, dans les deux tableaux qui suivent, la répartition des ouvrages acceptés et rejetés en fonction des critères documentaires suivants :

- Publications étrangères (Pub. étr)
- Ouvrages sous droits (Droits)
- Documents déjà numérisés (Doc déjà numérisés)
- Extraits de revues (Extr. De revues)
- Microformes (documents déjà reproduits, qui donnera lieu à un chantier de numérisation à part engagé par Gallica)
- Ruptures d'épine dorsale (RED)

Nombre de Cote	Examen docum.		
	O/N		
Nom type	O	N	Total
Autres	340	2	342
Doc deja numérisés	1	19	20
Droits		5	5
Extr. de revues	6	138	144
Microformes		86	86
Pub. étr.		123	123
RED		72	72
Total	347	445	792

Tableau 1 : répartition des ouvrages acceptés ou rejetés (en nombre)

²¹ CHARTIER, Roger, MARTIN, Henri-Jean. *Histoire de l'édition française*. Paris : Fayard/Cercle de la Librairie, vol. 3, p.107

Nombre de Cote	Examen docum.		
	O/N		
Nom type	O	N	Total
Autres	98%	0%	43%
Doc déjà numérisés	0%	4%	3%
Droits	0%	1%	1%
Extr. de revues	2%	31%	18%
Microformes	0%	19%	11%
Pub. étr.	0%	28%	16%
RED	0%	16%	9%
Total	100%	100%	100%

Tableau 2 : répartition des ouvrages acceptés ou rejetés (pourcentages colonnes)

Nous constatons que le taux global de rejet à cette première étape de sélection est de 56%. La catégorie « Autres » comprend les ouvrages qui ont franchi la première étape d'analyse et qui feront l'objet d'une vérification physique, à l'exception de deux ouvrages rejetés pour diverses causes : format nain [8- T19- 246] et impression à l'identique [8-T19-44 (A)].

En général, les extraits de revue ne sont pas numérisés indépendamment, car on privilégie la numérisation de la revue dans son intégralité ; par ailleurs, étant donné l'objectif important de documents à numériser, le département Sciences et Technique est censé sélectionner plutôt des monographies. Cependant 6 extraits de revue ont été retenus : 5 parmi ces 6 n'ont pu être identifiés comme extrait de revue que lors de l'examen physique, car la notice bibliographique ne le mentionnait pas. Le dernier extrait a été retenu car il présentait un véritable intérêt documentaire. Il s'agit d'Ozanam, Charles. *L'Anesthésie, histoire de la douleur*. Paris : B. Douniol, 1854 [8-T19- 184], extrait du *Correspondant*.

1.2.2. Répartition par cote

Notre corpus initial était constitué de trois tranches de cotes : [T2], [T19] et [TE39]. Le tableau suivant montre le taux d'acceptation/rejet par cote :

Nombre de Cote	Examen docum.		
	O/N		
Ref_cote	O	N	Total
T2	36%	64%	100%
T19	44%	56%	100%
TE39	46%	54%	100%
Total	44%	56%	100%

Tableau 3 : taux d'acceptation/rejet par cote

Nous constatons que les cotes [T19] et [TE39] sont homogènes car elles présentent des taux d'acceptation/rejet comparables à la moyenne. Pour [T2], le taux de rejet apparaît plus élevé que la normale. Sans toutefois effectuer un test d'indépendance qui nous permettrait de conclure plus rigoureusement, on peut penser que cet effet s'explique à la fois par une proportion de publication étrangères plus élevée que la moyenne (28% pour [T2] contre 16% en moyenne), et d'autre part par un échantillon plus faible (64 documents seulement sur un total de 792).

1.2.3. Répartition par date de publication

Nous proposons de regarder l'évolution de la répartition des dates d'édition après la première étape de sélection.

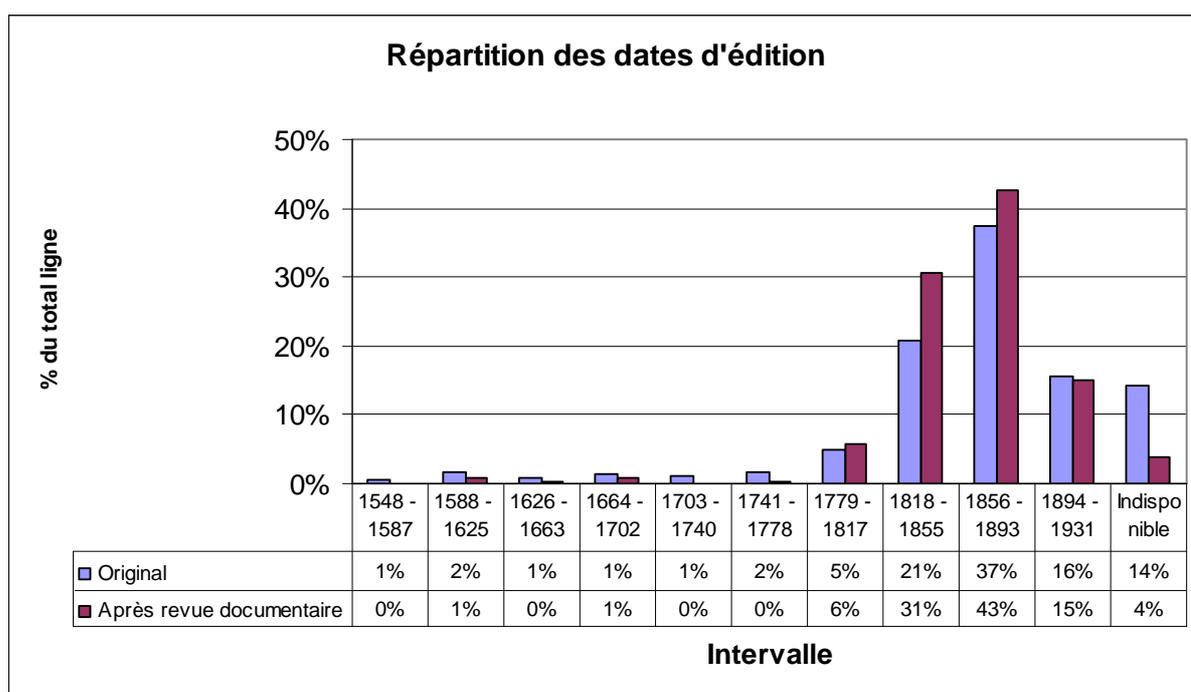


Figure 3

Le graphique ci-dessus montre que la première étape du processus de sélection a en première approche maintenu la répartition par rapport à la situation initiale, avec toutefois un léger renforcement dans la proportion des ouvrages édités dans la première moitié du XIX^e siècle, comme le montre ce graphique du taux de sélection par intervalle de date d'édition (nombre d'ouvrages sélectionnés sur nombre d'ouvrages initiaux pour chaque intervalle) :

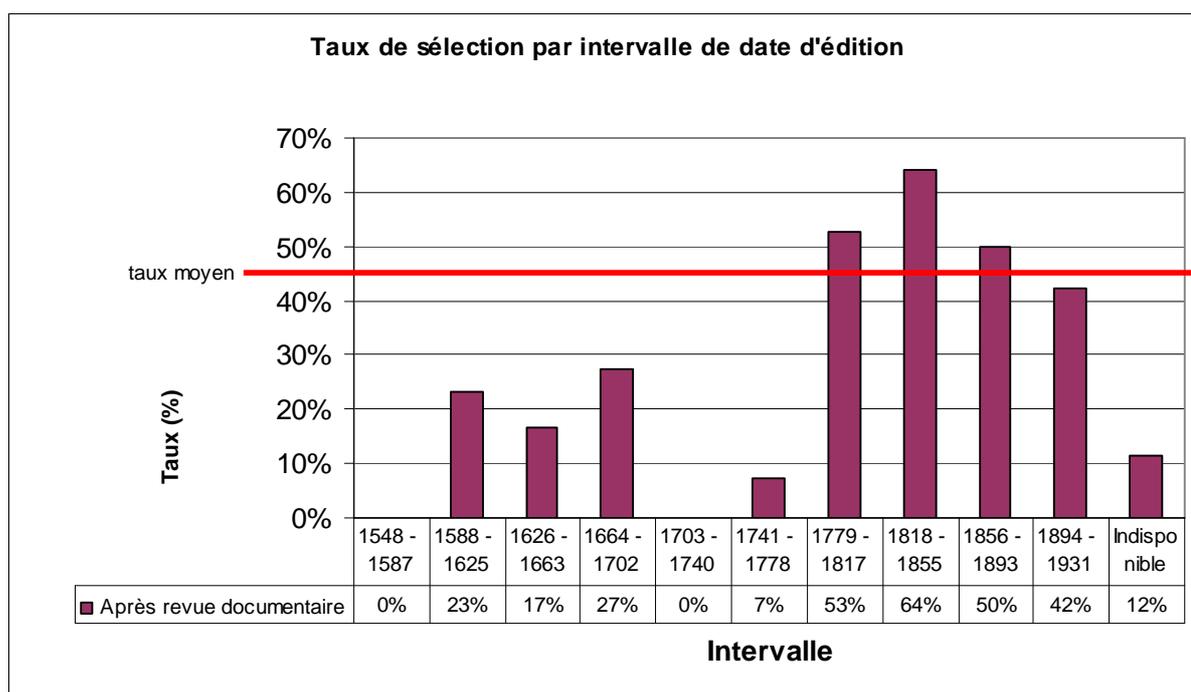


Figure 4

Même après des rejets importants suite à l'analyse documentaire nous remarquons que le taux de sélection est prépondérant pour les collections médicales du XIX^e siècle.

Les historiens de la médecine répartissent ce siècle en deux périodes afin de mieux saisir son essor scientifique et éditorial.

Une première période va du début du siècle où se succèdent les découvertes dans le domaine de la botanique, de la zoologie, de la chimie, de la physique et où sont créées les doctrines physiologiques, jusqu'à l'époque de la pathologie cellulaire et de la bactériologie. Entre 1815 et 1847, il y a un vrai essor de la méthode anatomo-clinique en médecine avec une production éditoriale considérable. L'anatomie pathologique est représentée par Jean Cruveilhier, la physiologie devient expérimentale avec Magendie, une médecine physiologique est imposée par Broussais, professeur au Val-de-Grâce en 1814 et Laennec, l'inventeur du stéthoscope, reste l'un des cliniciens le plus réputé de l'époque. La chirurgie connaît également un rapide progrès, qui se fonde sur les avancées de l'anatomie et de la pathologie. Celles-ci font l'objet d'études particulièrement approfondies, accompagnées de réflexions sur les méthodes et les pratiques médicales adoptées. Nous donnons quelques exemples : Roullier, Auguste (Dr). *Essai sur la philosophie médicale, contenant l'examen des principes qui servent de bases aux diverses théories, et leur application à la pratique*. Paris : Croullebois, 1815

[8- T19- 74] ; Dardonville, Hippolyte (Dr). *Réflexions pratiques sur les dangers des systèmes en médecine*. Paris : Méquignon-Marvis, 1818 [8- T19- 74] ou Caffin, Jacques-François (Dr). *Inductions physiologiques, pathologiques et thérapeutiques, ou Éléments généraux d'anthropologie et de médecine déduits des faits*. Paris : Gabon, 1822 [8- T19- 80]. Ces réflexions constituent de matières de cours enseignés dans les Faculté ou l'Académie de médecine de médecine : Broussais, François-Joseph-Victor (Dr). *Mémoire sur la philosophie de la médecine, lu à l'Académie des sciences, le... 8 octobre 1832*. (Paris,) : impr. De Lachevardière, (s.d.) [8- T19- 99] ou Forget, Charles-Polydore (Dr). *Programme du cours de philosophie médicale, professé à la Faculté de Strasbourg en 1844-45*. Paris : J.-B. Baillière, 1845 [8- T19- 140].

La seconde période débute en 1846 environ, date de la découverte de l'anesthésie chirurgicale et domine le siècle par le triomphe de la bactériologie. Les collections scientifiques de la BnF sont un reflet de cette évolution. Nous reviendrons plus en détail sur cette période dans la section 1.3.3 du troisième partie.

1.3. Corpus après l'examen physique

L'examen physique des documents est une étape indispensable avant toute validation définitive de la sélection, car il permet d'éliminer les ouvrages dont l'état ou les caractéristiques physiques les rendent inaptes à la numérisation. Comme nous le verrons dans le dépouillement des résultats, les facteurs d'échec peuvent provenir des caractéristiques intrinsèques du livre (typographie, taille, etc.) ou de phénomènes extrinsèques, en général l'usure du temps (tâches, déchirures, etc.).

1.3.1. Résultats généraux

Le Tableau 1 nous indiquait que 347 ouvrages ont franchi l'étape de l'examen documentaire. Le tableau ci-dessous montre les résultats de l'examen physique, détaillé par cote :

Nombre de Cote	Examen physique			
	O/N		(vide)	Total
Ref_cote	O	N		
T2	16	6	1	23
T19	156	44	7	207
TE39	104	12	1	117
Total	276	62	9	347

Tableau 4 : Résultats de l'examen physique (nombre d'ouvrages)

Nombre de Cote	Examen physique			
	O/N			
Ref_cote	O	N	(vide)	Total
T2	70%	26%	4%	100%
T19	75%	21%	3%	100%
TE39	89%	10%	1%	100%
Total	80%	18%	3%	100%

Tableau 5 : Résultats de l'examen physique (% par cote)

Nous pouvons remarquer qu'un taux important des ouvrages a été retenu : 80% contre seulement 44% dans l'étape de sélection documentaire. En effet, l'examen physique est consommateur de temps (il faut se rendre en magasin, localiser les ouvrages, les sortir et les examiner), c'est pourquoi il est important de procéder à un filtrage soigneux dans la première étape afin de minimiser le taux de rejet.

Ensuite, nous remarquons que les trois cotes sont relativement homogènes dans les taux d'acceptation : [T2] est un peu en dessous de la moyenne, [TE39] un peu au-dessus, ce qui est tout à fait comparable à la répartition obtenue dans la première étape (cf. Tableau 3).

Enfin, nous remarquons que sur les 347 ouvrages sélectionnés pour un examen physique, 9 n'ont pas pu être examinés (soit 3%) : il s'agit d'ouvrages manquants en place pour 5 d'entre eux, et de doublon de cotes pour les 4 restants (ouvrages déjà référencés sous une autre cote).

1.3.2. Analyse des ouvrages rejetés

Bien que les facteurs de rejet soient spécifiques à chaque monographie, et que parfois ces facteurs sont multiples pour un même ouvrage, nous avons tout de même tenté de regrouper les facteurs de rejet en dix catégories. Le tableau ci-dessous présente les résultats obtenus :

Nombre de Cote		
Nom_rejet	Total	%
Corps d'ouvrage dégradé	17	27%
Typographie	13	21%
Reiure serrée	10	16%
Marges trop faibles	6	10%
Extrait de revue	5	8%
Papier brûlé ou ondulé	4	6%
Cahiers manquants	1	2%
Format	1	2%
Transparence	1	2%
Autres	4	6%
Total	62	100%

Tableau 6 : répartition des facteurs de rejet

Nous constatons que les trois principales causes de rejet sont :

- le mauvais état du corps de l'ouvrage, en général également associé au papier acide ou ondulé. Cela rend la numérisation délicate car la manipulation risque de détériorer davantage l'ouvrage, et le prestataire refuse d'assumer ce risque ;
- la typographie ancienne, non reconnue par les logiciels de reconnaissance optique des caractères ;
- la reliure trop serrée, c'est-à-dire qui ne permet pas une ouverture suffisante de l'ouvrage pour une prise de vue correcte. En général ce facteur est également associé à des marges trop faibles.

Ces trois facteurs pris au sens large, c'est-à-dire incluant les marges trop faibles et le papier brûlé ou ondulé, expliquent plus de 80 % des rejets. Les autres sont liés à d'autres facteurs marginaux (extraits de revue non identifiés comme tels lors de la revue documentaire, et autres facteurs physiques divers et assez fréquents dans les collections médicales : tâches, rousseurs ou contrastes hétérogènes dans le corps de l'ouvrage).

1.3.3. Analyse du corpus sélectionné

Le graphique ci-dessous présente une comparaison dans la répartition des dates d'édition aux trois étapes de la sélection :

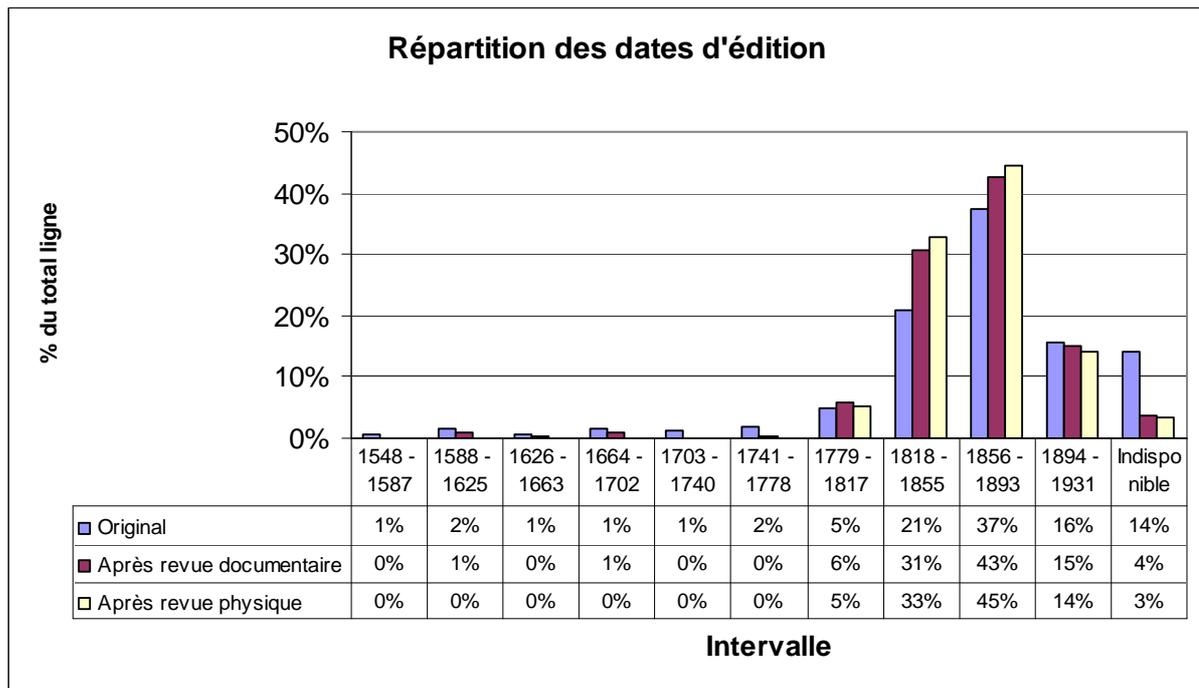


Figure 5

On constate que l'on retrouve la tendance observée lors de la première étape de sélection, c'est-à-dire une concentration importante d'ouvrages dans la deuxième moitié du XIX^e siècle. Les découvertes de l'anesthésie et de l'antisepsie bouleversent la chirurgie de l'époque et ont entraîné la publication massive d'articles scientifiques, bien représentés notamment dans la cote [TE39] où le nombre des extraites de revues et des pièces (ouvrages inférieurs à 50 pages) est conséquent. Cela prouve l'émergence d'une discipline nouvelle dans le champ médical. Le médecin Lucas-Championnière, fondateur du *Journal de médecine et de chirurgie pratique* publiera en 1876 le premier ouvrage didactique sur l'antisepsie chirurgicale. A cette époque, Pasteur publie ses travaux sur la microbiologie ; la physiologie générale est expérimentée par Claude Bernard, élève de Magendie, dont les résultats de son enseignement au Collège de France sont bien représentés dans Gallica (153 résultats dans une recherche simple par auteur).

Un certain nombre d'ouvrages parmi ceux sélectionnés enrichiront le corpus sur Claude Bernard déjà présent en ligne. Il s'agit par exemple de : Dupuy, Paul. *Du Rôle de la raison dans la médecine expérimentale d'après M. Claude Bernard*. Bordeaux : impr. De G. Gounouilhou, 1867 [8- T19- 257] et de Larroque, Patrice. *La Physiologie matérialiste. [Lettre adressée par M. Patrice Larroque à M. Claude Bernard, 22 juin 1870.]*. Paris : impr. De C. Noblet, (s. d.) [8- T19- 267].

Les ouvrages antérieurs à 1779 ont totalement disparu de la sélection, ce qui est dommage d'un point de vue de l'intérêt documentaire et historique. Cela s'explique davantage par des raisons combinant typographie ancienne et reliure plein cuir serré, plutôt que par un mauvais état. Au niveau des corps d'ouvrage, les ouvrages du XVII^e-XVIII^e siècles sont en général en meilleur état que les documents de la fin du XIX^e siècle (papier chiffon de meilleure qualité, cahiers cousus etc).

1.4. Validations du corpus par Gallica et DCO

Suite à la sélection des ouvrages, le fichier de résultats a été transmis à Gallica pour des vérifications supplémentaires et à la DCO pour validation. Le tableau ci-dessous présente les résultats de cette validation.

Nombre de Cote	Accepté Gallica		
	O/N		Total
Examen physique O/N	O	N	
O	254	22	276
N		63	63
(vide)	5	448	453
Total	259	533	792

Tableau 7 : résultats de la validation Gallica

276 ouvrages ont été soumis à Gallica : il s'agit des ouvrages qui ont passé l'étape de l'examen physique. Sur ces 276, 254 ont été acceptés et 22 ont été rejetés. En plus de ces 254 ouvrages, on compte également 5 ouvrages acceptés par Gallica, mais qui n'avaient pas été retenus suite à l'étape de sélection documentaire. Il s'agit d'extraits de publications de *L'Union Médicale*, qui dans un premier temps ont été éliminés car tirés à part, mais acceptés par Gallica pour compléter des lacunes dans la numérisation exhaustive du périodique *Union Médicale*. Cela porte donc le nombre total de sélections validées à 259.

Nous analyserons les 22 ouvrages refusés par Gallica. Le tableau suivant détaille les causes de rejet :

Nombre de Cote	
Rejet_Gallica	Total
Déjà numérisé	11
Numérisation en cours	7
Sous droits	3
Etat physique	1
Total	22

Tableau 8 : détail des rejets Gallica

On constate que plus de 80% des ouvrages rejetés sont déjà numérisés ou en cours de numérisation. En effet Gallica tient à jour un inventaire de tous les programmes de numérisation en cours internes ou externes à la BnF afin d'éviter les doublons de numérisation.

Les autres rejets sont marginaux et concernent surtout des ouvrages dont les auteurs sont décédés après 1937, et donc pas encore libres de droits.

2. Bilan des sélections des documents au sein de la BnF

2.1.1. Département Sciences et techniques

Nous présentons en Annexe 3 l'état actuel de sélection dans la tranche de cote [Td] (Pathologie) avec les taux de rejets répartis entre les différents critères définis. Ce document nous permet de rapporter le travail effectué au travail prévu par la coordination DCO, qui était initialement de 370 monographies, auxquelles s'ajoutent 190 périodiques à sélectionner et à valider par mois (comptage en termes d'unités documentaires).

Le nombre d'UC vérifiées dans la tranche [Td], pour le « marché des 100 000 », s'élève à 5 379.

Dans les fichiers du [Td], 68 % des UC sont rejetées après les premières vérifications dans le catalogue. Les deux premières causes de ces rejets sont la grande proportion de publications étrangères (31%) ainsi que les nombreuses ruptures d'épine dorsale (40 %), qui concernent également les publications étrangères. Les rejets pour cause de documents déjà reproduits s'élèvent à 9 % des rejets documentaires et les extraits de revues représentent 9 %.

Lors des vérifications en magasin, les deux causes principales de rejets sont liées à la présence dans ce fonds d'ouvrages anciens (XVI^e siècle – XVIII^e siècle), qui présentent des problèmes soit de reliure serrée (39 %) soit de typographie (30 %).

Après tous les rejets liés aux causes documentaires et physiques, le nombre des UC sélectionnées est de 980, soit 18 % du nombre total des UC proposées à la numérisation.

2.1.2. Direction des collections

Nous présentons un état des lieux sur l'avancement du travail de sélection d'ouvrages pour la numérisation de masse dans les départements de la Direction des collections, depuis le début des opérations. Cela se révèle utile pour mieux situer la position du Département Sciences et techniques par rapport aux autres départements :

Département	Nombre UC proposés	Nombre d'UC validés	Pourcentage
Sciences et Techniques	11 686	950	8,13 %
Philosophie, Histoire, Science de l'Homme	8 295	4 418	53,26 %
Economie, Politique	17 975	1 160	6,45 %
Littérature et Art	4 214	2 712	64,36 %
Total	42 170	9 240	22 %

Les écarts significatifs entre les départements sont dus essentiellement à l'application des critères documentaires. Les deux premières causes de rejets pour le DEP et SCT sont la présence massive de tirés à part (respectivement 47 % et 33 %) et les RED (respectivement 16 % et 19 %). Ensuite, les publications étrangères sont très fréquentes dans les collections des DEP et SCT : respectivement 15 % et 19 % des rejets sont relatifs à ce critère.

La première cause de rejets du Département Philosophie, Sciences de l'Homme est due aux tirés à part, identifiés en proportion de 22 %. Pour le Département Littérature et Art la reproduction sur microfilms représente la première cause de rejets (17 %). Les publications étrangères et les problèmes de droit sont une cause de rejet très minoritaire au sein de ce département.

2.2. Analyse d'une méthode de sélection expérimentale

Un outil de gestion des documents sélectionnés, intitulé ADCAT 15, sera mis en production à la fin du mois juin 2007. En attendant, les sélections, les vérifications et les validations de documents à numériser dans le cadre du « marché des 100 000 », le travail s'effectue à partir de fichiers constitués de tranches de cotes prédéterminées. Au sein des collections de médecine, la section Pathologie [Td], choisie comme axe documentaire dans Gallica 2, comprend la totalité des UC de la tranche [Td], de [Td 1 à Td 40], soit un nombre de 5 379 UC à vérifier.

Chaque fichier fourni, au format Excel, comprend un grand nombre de colonnes, certaines remplies automatiquement, d'autres à remplir et d'autres qui peuvent rester vides.

Une fois les vérifications documentaires accomplies, on établit la liste des ouvrages à vérifier en magasin, selon les critères physiques²². Les colonnes du fichier Excel sont remplies avec les observations sur l'état physique de livres (état du corps d'ouvrage, type de papier, reliure). Ensuite les fichiers sont envoyés pour des vérifications supplémentaires et pour validation.

Afin d'accélérer le rythme des sélections au sein du département, un changement dans la méthode de travail a été proposée. Au lieu de partir de listes complètes d'ouvrages avec des données bibliographiques, les vérifications commencent directement en magasin, sans listes préalables, à partir d'une tranche de cote définie. Les ouvrages sont vérifiés à la fois selon leur état physique et selon les contraintes documentaires facilement identifiables dans l'ouvrage. Des vérifications supplémentaires sont nécessaires par la suite dans le catalogue afin de s'assurer du respect de tous les critères documentaires.

Dans le cadre de mon stage, j'ai expérimenté en binôme cette méthode afin de tester son efficacité. Sur un total de 1 107 UC vérifiées dans les tranches [Tb] et [Tc], 204 UC ont été retenues, soit 18 %. Les rejets les plus importants sont dus aux publications étrangères (38 %) et aux ouvrages anciens (15 %).

Cette nouvelle méthode de sélection des ouvrages, différente de la méthode classique utilisée dans le département et en partie proche du récolement, présente l'avantage d'un contact direct avec les collections en vue de l'identification des ouvrages numérisables. Cependant, elle demande un temps supplémentaire pour les investigations et des manipulations plus nombreuses des documents en magasin. Nous pensons que

l'utilisation de cette méthode pourra être efficace en faisant un brassage des deux méthodologies. Les vérifications peuvent débuter en magasin, mais seraient avantageusement accompagnées par l'usage des listes des ouvrages à vérifier afin d'éviter des manipulations inutiles. Cela permettrait également d'identifier des documents qui présentent un intérêt documentaire mais qui se retrouvent écartés de la sélection pour cause de rupture d'épine dorsale.

3. Elargissement et enrichissement du corpus

Le corpus de « Philosophie médicale », actuellement constitué essentiellement des collections du Département Sciences et techniques, pourra être enrichi avec des documents intellectuellement proches mais disséminés dans d'autres sections du *Catalogue des Sciences médicales* ou dans d'autres programmes de numérisation en cours.

3.1. Catalogue méthodique des sciences médicales

Nous avons dégagé deux thématiques qui offrent une réelle ouverture pour l'enrichissement de notre corpus : la littérature médicale et l'histoire des sciences médicales. Ces disciplines font partie du premier chapitre du *Catalogue*, « Préliminaires et généralités des Sciences médicales », réservé aux traités de philosophie et de méthodologie médicale et aux traités généraux sur l'ensemble des sciences médicales.

Nous indiquons les trois cotes suivantes :

- [T1] : Littérature, érudition et critiques médicales (272 notices bibliographiques dans BN Opale Plus) ;
- [T5] : Histoire, exposition et critique des doctrines (376 notices bibliographiques dans BN Opale Plus) ;
- [T20] : Méthodologie (113 notices bibliographiques dans BN Opale Plus)
- [T21] : Histoire générale des sciences médicales (1 437 notices bibliographiques dans BN Opale Plus).

Dans ces sections, nous avons relevé des ouvrages ayant trait à la thématique de notre corpus et nous en indiquons quelques exemples :

²² Détaillé dans la page 52 de notre mémoire.

- GASTE, Léonard-Fulcrand (Dr). *Abrégé de l'histoire de la médecine considérée comme science et comme art, dans ses progrès et son exercice, depuis son origine jusqu'au XIXe siècle.* [8- T1- 35] ;
- DAREMBRG, Charles. *Cours sur l'histoire et la littérature des sciences médicales.* [8- T1- 61 (1)] ;
- CHORTET, Jean-François (Dr). *Traité où l'on démontre philosophiquement que le système de l'immortel John Brown est le seul vrai en physiologie.* [8- T5- 57] ;
- LEBESCHU, de La Bastays. *Révolution de la médecine, ou Régénération de l'art de guérir, ouvrage où l'on expose une nouvelle doctrine sur l'organisation et la désorganisation animale.* [8- T5- 61] ;
- GOURAUD, Henri (Dr). *Essai critique sur Broussais, sa doctrine médicale et ses opinions philosophiques.* [8- T5- 97] ;
- GALLE, François. *Observation clinique, précédée et suivie de quelques réflexions sur la véritable situation de la médecine, ou Nouvel examen des doctrines médicales.* [8- T5- 90] ;
- SAGE, Balthasar-Georges. *Théorie de la vitalité. Décomposition de la dépouille mortelle de l'homme. Itératives annotations.* [8- T5- 105] ;
- CHRISTOPHE, César-Auguste (Dr). *L'Évangile médical, ou Traité des causes premières de l'homme, nouvelle doctrine fondée sur la découverte de la vie, de son essence et de ses lois.* [T5- 142] ;
- SAUCEROTTE, Constant. *Histoire critique de la doctrine physiologique, suivie de considérations sur l'histoire philosophique de la médecine, et sur l'hippocratisme moderne.* [8- T5- 150] ;
- FUNCK-BRENTANO, Théophile. *Méthode et principes des sciences naturelles, introduction à l'étude de la médecine.* [8- T20- 70] ;
- LE FRANÇOIS, Alexandre. *Réflexions critiques sur la médecine, où l'on examine ce qu'il y a de vrai et de faux dans les jugements qu'on porte au sujet de cet art.* [8- T21- 130] ;
- CHEVALIER, J.-P. *La Médecine dévoilée, ou Examen critique de la science médicale et démonstration de la nécessité de recourir aux enseignements de la.* [8- T21- 249] ;
- MARMISSE, Géraud. *Merveilles évangéliques éclairées par les sciences médicales.* [8- T21- 251] ;
- MOREAU, L.-D. (Dr). *Du Triomphe de la vérité en médecine.* [8- T21- 252].

3.2. Autre pistes d'élargissement

L'enrichissement de notre corpus peut aussi se poursuivre par une coopération fructueuse avec les programmes de numérisation des diverses institutions afin de mettre à la disposition des lecteurs une offre culturelle encore plus variée. Les autres programmes de numérisation en cours sont menés par les pôles associés de la BnF ou par des institutions en étroite collaboration avec la BnF. Nous citons quelques programmes en cours :

- BIUM : Medic@ prévoit la numérisation de 4000 ouvrages répartis entre les thématiques suivantes : histoire de la médecine, maladies mentales, pédiatrie ;
- le CNRS est lancé dans un ample projet de numérisation de l'intégralité de l'œuvre Buffon, Ampère et Lamarck, Lavoisier ;
- le Service Commun de la Documentation Médicale de l'Université Pierre et Marie Curie enrichira un corpus déjà existant en ligne autour de Charcot ;
- la Médiathèque de la Villette proposera un corpus sur la vulgarisation des sciences et sur les maladies mentales.

La numérisation interne d'une collection de périodiques qui intégrera bientôt Gallica pourra compléter ponctuellement notre corpus. Il s'agit de : *Archives des sciences physiques et naturelles*, *Bulletin de l'Académie de médecine*, *Bulletins de la Société d'Anthropologie de Paris*, *Gazette médicale de Lyon*, fondée et publiée par M. Barrier, chirurgien en chef désigné de l'Hôtel-Dieu.

Enfin, on pourra explorer une mise à disposition d'une offre sous droits plus importante, qui permettra un accès à des documents plus récents.

Conclusion

Dans le cadre du stage, nous avons participé au sein du Département Sciences et techniques au processus effectif de sélection des ouvrages pour la numérisation de masse, en conformité avec la politique documentaire définie par la BnF, afin de constituer un corpus d'ouvrages en sciences médicales. Déjà validé, ce corpus intégrera bientôt les collections en ligne de Gallica et par la suite celles de Gallica 2.

L'accomplissement du travail de sélection, à partir des collections patrimoniales, nous a permis de découvrir la richesse des documents conservés dans les magasins de la BnF, de comprendre leur état matériel et de s'initier au système de cotation. Nous avons également beaucoup appris sur l'organisation de la chaîne de production numérique à la BnF et les acteurs qui y participent, l'application des lois sur le droit d'auteur et la politique documentaire de numérisation.

Les critères stricts de sélection imposés par les services de numérisation et de coopération nous ont conduit à retenir au final 259 ouvrages sur un total de 792 initialement prévus pour la numérisation au sein de notre corpus, soit un taux de sélection de 31%. Des analyses menées à tous les stades du processus de sélection documentaire et physique, aussi bien sur la répartition des dates d'édition que sur les tranches de cotes prévues, nous ont permis de conclure que les contraintes techniques de numérisation, malgré les rejets qu'elles produisent, ne remettent pas en cause la pertinence documentaire du corpus sélectionné. Avant toute sélection, plus de la moitié des ouvrages de notre corpus étaient édités dans le XIX^e siècle, avec une prépondérance pour la deuxième moitié XIX^e siècle. A la fin du processus de sélection, cette même période, particulièrement fructueuse dans la production médicale et éditoriale, reste également massivement représentée.

Le stage nous a permis de mieux comprendre l'enjeu majeur du projet de la Bibliothèque numérique Européenne : la démocratisation du savoir par une mise à la disposition intégrale et cohérente de son patrimoine. Cela conforte la mission traditionnelle des bibliothèques, de véritables cathédrales du savoir, qui jouent un rôle

majeur dans la formation des esprits des générations présentes et futures. Nous avons également saisi l'articulation des coopérations entre la BnF et d'autres institutions ou organismes afin de poursuivre une riche politique documentaire partagée, une des clés de la réussite de ce projet.

Malgré le fait de ne pas encore disposer d'un outil spécial dédié aux sélections, qui sera mis en production à la fin du mois de juin 2007, l'expérience acquise au sein de la BnF s'avère très riche pour mon futur professionnel et je suis très contente d'avoir participé, grâce à la mission de stage qui m'a été confiée, à ce projet européen dont les valeurs me sont très proches.

Au vu de sa participation importante à la richesse matérielle et spirituelle de la planète, de ses 450 millions d'habitants et de ses nombreux partenaires (78 pays de plusieurs continents), l'Europe a la responsabilité de la réussite de ce projet très ambitieux.

Liste des sigles et des acronymes

BNF	Bibliothèque Nationale de France
BNUE	Bibliothèque Numérique Européenne
DCO	Direction des collections
DCP	Département de la coopération
DSC	Département de la conservation
DSR	Direction des Services et Réseaux
DBN	Département de la Bibliothèque numérique
DRB	Département de Recherche Bibliographique
DSI	Département des Systèmes d'Information
TEL	The European Library
CENL	The Conference of European National Librarians
BIUM	Bibliothèque Interuniversitaire de Médecine
INHA	Institut National d'Histoire d'Art
CNAM	Conservatoire National des Arts et Métiers

Glossaire

ADCAT 15

Outil de gestion des documents sélectionnés dans le cadre du « marché des 100 000 », mis en production à la fin du mois de juin 2007.

A-DCM- 26

Application informatique de magasinage pour la « Gestion des unités de conservation »

BN-Opale

Base des documents imprimés destinés aux magasins, opérationnelle de 1987 à 2002 pour le catalogage courant et rétrospectif. Contenait aussi, par chargement, les documents entrés à partir de 1970.

BN-Opale Plus

Catalogue en ligne qui contient la majorité des références des documents conservés sur tous les sites de la BnF.

Cotatran

Système de cotation transitoire des imprimés (1996-juillet 1998). Exemple : 4-D2 MON-25, 8-D4 PER-12.

Cote pure

Cote sans sous-cote, par opposition à la même cote existant avec une sous-cote. Exemple : JO-1074 [cote pure], JO-1074 (BIS) et JO-1074 (TER).

CRI

Conversion Rétrospective des Imprimés (marchés de 1990 à 1999). Fichiers et catalogues des origines à 1969.

ADCAT-2

Application de catalogage et l'interface professionnelle de BN Opale Plus.

Exemplarisation

Création d'un exemplaire numérique, microforme ou libre accès.

GEAC-LA

Base des documents imprimés destinés au libre accès et des documents imprimés numérisés de 1992 à 2002.

OCR

Reconnaissance optique des caractères (en anglais Optical Character Recognition) : application qui permet de convertir en fichier texte un texte sur une image numérique afin de l'exploiter sur un traitement de texte.

Microforme

Reproduction sur support argentique, qui peut se décliner entre microfilms, destinées à la reproduction des périodiques, des ouvrages volumineux ou des ouvrages illustrés et microfiches pour la reproduction des monographies pas volumineuses.

Pôle Associé

Réseau constitué autour de la BnF par des bibliothèques municipales, de bibliothèques universitaires et quelques autres institutions, dans le but de mettre en commun leurs ressources documentaires.

RED

Rupture d'Épine Dorsale. On parle de rupture basse d'épine dorsale (pas d'UC associée à une côte) et de rupture haute d'épine dorsale (pas de notice bibliographique associée à un UC). Pour un exemplaire sans notice bibliographique, on emploie également l'expression « exemplaire orphelin ».

Sycomore

Système informatique utilisé pour la communication et le récolement aux départements des Imprimés et des Périodiques de 1991 à 1998.

Unité de conservation (UC)

Unité de conservation. C'est l'unité physique sur laquelle on appose un code barres. L'UC peut correspondre à une seule notice du catalogue (monographie), à plusieurs notices (recueil de pièces) et inversement une notice de catalogue peut avoir plusieurs UC rattachées à elle (cas de périodiques par exemple).

Uctran

Application utilisée transitoirement pour la création de PEX et d'UC (novembre 1998-2002). Utilisée parfois pour qualifier les cotes millésimées correspondant au nouveau système de cotation de la BnF.

Unité documentaire (UD)

Pour les monographies, l'UD c'est l'ouvrage ou le tome pour les publications en plusieurs tomes mais avec page de titre distincte pour chaque tome. Pour les périodiques, c'est le fascicule avec sa page de titre ou de couverture, sa table des matières et sa pagination. Si plusieurs fascicules ont une pagination continue et une table des matières communes, alors c'est le total de ces fascicules qui constitue l'ud.

Bibliographie

DOCUMENTS PAPIER

ALIX, Yves (dir.). *Le Droit d'auteur et les bibliothèques*. Paris : Cercle de la librairie, 2000. 237 p. (Bibliothèques). ISBN 2-7654-0785-1

ARMS, William Y. *Digital Libraries*. Cambridge Mass. : M.I.T. Press, 2000. 287 p. (Digital libraries and electronic publishing). ISBN 0-262-01180-8

AROT, Dominique. La bibliothèque numérique patrimoniale : paradoxes, missions, typologie et évaluation. In **LE MOAL, Jean-Claude, HIDOINE, Bernard** (dir.). *Bibliothèques numériques : cours INRIA, 9-13 octobre 2000*. Paris : ADBS éditions, 2000. p.11-52.

BARBIER, Frédéric. Les Imprimeurs. In **CHARTIER, Roger, MARTIN, Henri-Jean** (dir.). *Histoire de l'édition française : le temps des éditeurs, du Romantisme à la Belle Epoque*. Paris : Fayard – Cercle de la Librairie, 1985, p. 67-92.

BIBLIOTHEQUE IMPERIALE. Département des Imprimés. *Catalogue des Sciences médicales*. Paris : Firmin-Didot, 1857-1880.

BURESI, Charlette, CEDELLE-JOUBERT, Laure. *Conduire un projet de numérisation*. Villeurbanne : ENSSIB ; Paris : Ed. Tec & Doc, 2002. 326 p.

CACALY, Serge (dir.). *Dictionnaire encyclopédique de l'information et de la documentation*. Paris : Nathan, 1997. 634 p. ISBN 2-09-190528-3

CHARLE, Christophe. Le champ de la production littéraire. In **CHARTIER, Roger, MARTIN, Henri-Jean** (dir.). *Histoire de l'édition française : le temps des éditeurs, du Romantisme à la Belle Epoque*. Paris : Fayard – Cercle de la Librairie, 1985, p. 137-176.

DUMAITRE, Paul. *Histoire de la médecine et du livre médical*. Paris : Pygmalion, 1978. 430 p. ISBN 2-85704-038-5

FARCHY, Joëlle, ROLLET, Alain (eds.) *Droit d'auteur et numérique*. Paris : Hermès Science Publications, 2002. (Réseaux). 215 p. ISBN 2-7462-0429-0

FOUCHE, Pascal, PECHOIN, Daniel, SCHUWER, Philippe, et al. *Dictionnaire encyclopédique du livre*. Paris : Cercle de la Librairie, 2005. 900 p. ISBN 2-7654-0841-6

FREYRE, Elisabeth. La Bibliothèque européenne : un portail européen. *BIBLIothèque(s) : revue de l'Association des bibliothécaires français*, 2005, n° 22, p.19-21.

GAME, Valérie. La constitution d'un fonds d'images et d'ouvrages numérisés par la BNF. *Bulletin d'informations de l'Association des bibliothécaires français*, 1996, n° 172, p.89-92.

HUARD, Pierre, IMBAULT-HUART, Marie-José. Histoire de la médecine des origines au XX^e siècle. In *Encyclopédie de la Pléiade : Médecine*, vol. 45. Paris : Gallimard, 1980, p. 601-670.

JACQUSSON, Alain, RIVER, Alexis. *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*. Paris : Cercle de la librairie, 2005. 573 p. (Bibliothèques). ISBN 2-7654-0915-3

JAOUEN, Lisa, LE DIVIDICH, Aude, LEMAIRE, Alice. La Bibliothèque Nationale et les sciences au XIX^e et XX^e siècle : premiers résultats d'une enquête. *Revue de la Bibliothèque nationale de France*, 2003, n° 14, p. 71-76.

JEANNENEY, Jean-Noël. *Quand Google défie l'Europe : plaidoyer pour un sursaut*. Paris : Mille et une nuits, 2005. 113 p. ISBN 2-842-05912-3

LAFAY, Arlette (dir.). *La douleur : approches pluridisciplinaires*. Paris : Editions l'Harmattan, 1992. 175 p. ISBN 2-7384-1281-5

LEDOS, Gabriel. *Histoire des catalogues des livres imprimés de la Bibliothèque Nationale*. Paris : Edition des bibliothèques nationales, 1936.

LEMAIRE, Alice. Les collections scientifiques et techniques de la Bibliothèque nationale de France. *Bulletin de la société des Amis de la Bibliothèque Polytechnique*, 2002, n° 30, p. 78-83.

LUPOVICI, Christian. L'évolution du rôle des bibliothécaires et documentalistes dans le cadre de la bibliothèque numérique. In **PAPY, Fabrice** (dir.) *Les bibliothèques numériques*. Paris : Lavoisier, 2005, p. 17-33. ISBN 2-7462-1036-3

MICHAUD, Yves (dir.). *Les Maladies et la Médecine*. Paris : Editions Odile Jacob, 2002. 346 p. ISBN 2-7381-1094-0

MORTON, L.T., GODBOLT, S. (eds.). *Information Sources in the Medical Sciences*. London : Butterworths, 1984. 534 p. ISBN 0-408-11473-8

MOUILLET, Evelyne. *La recherche bibliographique en médecine et santé publique : guide d'accès*. Paris : Elsevier, 2005. 155 p. ISBN 2-84299-684-4

PASQUIGNON, Anne. Les sciences médicales dans les catalogues de livres imprimés de la BN. *Revue de la Bibliothèque Nationale*, 1990, n° 36, p. 44-50.

PREVOTEAU, Marie-Hélène, UTARD, Jean-Claude. Manuel de bibliographie générale. Paris : Cercle de la librairie, 2005. 524 p. (Bibliothèques). ISBN 2-7654-0912-9

RENOULT, Daniel. La bibliothèque numérique de la BNF : conception, méthodes, résultats. *Document numérique*, 1998, vol. 2, n° 3-4, p. 72-87.

RENOULT, Daniel, MELET-SANSON, Jacqueline (dir.). *La Bibliothèque nationale de France : collections, services, publics*. Paris : Cercle de la librairie, 2001, 238 p. (Bibliothèques) ISBN 2-7654-0820-3

REY, Roselyne. *Histoire de la douleur*. Paris : La Découverte, 2000. 420 p. ISBN 2-7071-3196-2

RICHARD, Hélène. La mémoire des Etats européens : pour une conservation concertée des Cartes et plans. *BIBLIOTHÈQUE(s) : revue de l'Association des bibliothécaires français*, 2005, n° 22, p. 22-24.

SANSOY, Nicolas. Anesthésie. In *Encyclopédie de la Pléiade : Médecine*, vol. 45. Paris : Gallimard, 1980, p. 841-865.

SOURNIA, Jean-Charles. *Histoire de la médecine*. Paris : La Découverte, 1997. 358 p. ISBN 2-7071-2783-3

STASSE, François. *La véritable histoire de la grande bibliothèque*. Paris : Ed du Seuil, 2002. 205 p.

DOCUMENTS EN LIGNE

BIBLIOTHEQUE NATIONALE DE FRANCE. *Numérisation : charte technique* [en ligne]. [réf du 21 mai 2007]. Disponible sur : <http://bibnum.bnf.fr/numerisation/charte_technique_standard_num_mars2005.pdf>

BIBLIOTHEQUE NATIONALE DE FRANCE. *Charte documentaire de Gallica* [en ligne]. [réf du 23 mai 2006]. Disponible sur : <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancree=numerisation/po_chartegallica.htm>

GOUSTARD, Alain. Vers la Bibliothèque numérique européenne. *Chroniques de la Bibliothèque nationale de France* [en ligne]. 2007, no 38, p. 3-7. [réf. du 8 juin 2007]. Disponible sur : <http://chroniques.bnf.fr/default.php?src1=numero_courant/cooperation/menu_gauche.php&src2=numero_courant/cooperation/biblio_numerique.htm&m1=e1One&m3=E1&m3=1>

LUPOVICI, Catherine, CLOAREC, Thierry, CHARENTENAY, France de. Les usages de Gallica. *Bulletin des Bibliothèques de France* [en ligne]. 2003, n° 4, p. 40-44. [réf du 4 juin 2007]. Disponible sur : <<http://bbf.enssib.fr/sdx/BBF/frontoffice/2003/04/document.xsp?id=bbf-2003-04-0040-007/2003/04/fam-dossier/dossier&nDoc=8&statutMaitre=non&statutFils=non&tri=>>>

LUPOVICI, Catherine. Les stratégies de gestion et de conservation préventive des documents électroniques. *Bulletin des Bibliothèques de France* [en ligne]. 2000, n°4, p. 43-54. [réf du 4 juin 2007]. Disponible sur : <<http://bbf.enssib.fr/sdx/BBF/frontoffice/2000/04/document.xsp?id=bbf-2000-04-0043-004/2000/04/fam-dossier/dossier&nDoc=5&statutMaitre=non&statutFils=non&tri=>>>

LUPOVICI, Christian. La chaîne de traitement des documents numériques : caractéristiques et mise en oeuvre. *Bulletin des Bibliothèques de France* [en ligne]. 2002, n°1, p. 86-91. [réf. du 7 juin 2007]. Disponible sur : <<http://bbf.enssib.fr/sdx/BBF/frontoffice/2002/01/document.xsp?id=bbf-2002-01-0086-001/2002/01/fam-apropos/outil&nDoc=2&statutMaitre=non&statutFils=non&tri=>>>

POIRIER-BRECHE, Véronique. Stratégies de numérisation. *Bulletin des Bibliothèques de France* [en ligne]. 2001, n°6, pp. 24-28. [réf. du 5 juin 2007]. Disponible sur : <http://bbf.enssib.fr/bbf/html/2001_46_6/2001-6-p24-poirier-breche.xml.asp>

TESNIERE, Valérie, LESQUINS, Noémie. La bibliothèque numérique européenne. Une stratégie culturelle de la Toile. *Bulletin des bibliothèques de France* [en ligne]. 2006, n°3, p. 68-80. [réf. du 4 juin 2007]. Disponible sur : <<http://bbf.enssib.fr/sdx/BBF/frontoffice/2006/03/document.xsp?id=bbf-2006-03-0068-012/2006/03/fam-dossier/dossier&nDoc=2&statutMaitre=non&statutFils=non&tri=>>>

TESNIERE, Valérie. Numérisation partagée : quelle offre et quels accès ? Etat de l'art et portée des expérimentations en cours. *7e Journée des pôles associés* [en ligne]. 2003, Paris : Bibliothèque Nationale de France. [réf. du 5 juin 2007]. Disponible sur <http://www.bnf.fr/pages/infopro/journeespro/pdf/poles_pdf/po2003_Tesniere.pdf >

RESSOURCES ELECTRONIQUES DANS LE DOMAINE MEDICAL

ACADEMIE NATIONALE DE MEDECINE. Bibliothèque. *Catalogue de l'Académie* [en ligne]. Paris : Bibliothèque de l'Académie nationale de Médecine. [réf. du 5 juin]. Disponible sur : <<http://www.academie-medecine.fr/bibliotheque.asp>>

BANQUE DE DONNES SANTE PUBLIQUE. Annuaire critique de sites web [en ligne]. [réf. du 7 juin 2007]. Disponible sur : <<http://www.bdsp.tm.fr/webs/default.asp>>

BIBLIOTHEQUE INTER UNIVERSITAIRE DE MEDECINE. *Medic@* [en ligne]. [réf. du 4 juin 2007]. Disponible sur : <<http://www.bium.univ-paris5.fr/histmed/medica.htm>>

CHU-HOPITAUX DE ROUEN. *Catalogue et Index des Sites Médicaux Francophones de CHU de Rouen* [en ligne]. [réf. du 7 juin 2007]. Disponible sur : www.chu-rouen.fr/cismef/

INSTITUT PASTEUR. Médiathèque. *Catalogue informatique de la Médiathèque.* [en ligne]. Paris : Médiathèque de l'Institut Pasteur. [réf. du 5 juin 2007]. Disponible sur : <<http://www.pasteur.fr/externe.html>>

INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE. Medical Subject Headings bilingue [en ligne]. [réf. du 7 juin 2007]. Disponible sur : <<http://ist.inserm.fr/basismesh/mesh.html>>

REPertoire DE LIENS

BIBLIOTHEQUE NATIONALE DE FRANCE. *Europeana : la contribution française à la Bibliothèque Numérique Européenne* [en ligne]. [réf. du 4 juin 2007]. Disponible sur : <<http://www.europeana.eu>.>

BIBLIOTHEQUE NATIONALE DE FRANCE. *Les signets de la BnF : bibliothèques numérique* [en ligne]. [réf. du 4 juin 2007]. Disponible sur : <http://signets.bnf.fr/html/categories/c_011textes_num.html>

COMMISSION EUROPEENNE. *European Digital Library Project* [en ligne]. [réf. du 4 juin 2007]. Disponible sur : <<http://edlproject.eu>>

CONFERENCE OF EUROPEAN NATIONAL LIBRARIANS. *The European Library* [en ligne]. [réf. du 4 juin 2007]. Disponible sur : <<http://www.theeuropeanlibrary.org>>

CONSERVATOIRE NATIONAL DES ARTS ET METIERS. *Conservatoire numérique des arts et métiers* [en ligne]. [réf. du 5 juin 2007]. Disponible sur : <<http://cnum.cnam.fr/>>

DIGITAL LIBRARY FEDERATION. *Digital Library Federation* [en ligne]. [réf. du 5 juin 2007]. Disponible sur : <<http://www.diglib.org/>>

MINISTERE DE LA CULTURE ET DE LA COMMUNICATION. *Culture.fr, le portail de la culture : bibliothèque numériques* [en ligne]. [réf. du 5 juin 2007]. Disponible sur : <http://www.culture.fr/fr/sections/themes/bibliotheque_mediathèque/sous_themes/c391>

MINISTERE DE LA CULTURE ET DE LA COMMUNICATION. *Actualité de la numérisation du patrimoine culturel* [en ligne]. [réf. du 5 juin 2007]. Disponible sur : <<http://www.coulture.gouv.fr/culture/doc/index.html>>

NORMALISATION

ASSOCIATION FRANCAISE DE NORMALISATION. *Formation des bibliothécaires et documentalistes : normes pour l'épreuve de catalogage : références bibliographiques, partie composantes, cartes.* Paris : AFNOR, 1998. Z 44-005 : décembre 1987 : Documentation – Références bibliographiques : Contenu, forme et structure, p.1-13.

ASSOCIATION FRANCAISE DE NORMALISATION. *Formation des bibliothécaires et documentalistes : normes pour l'épreuve de catalogage : références bibliographiques, partie composantes, cartes.* Paris : AFNOR, 1998. Z 44-005 : décembre 1987 : Documentation – Références bibliographiques : Contenu, forme et structure, p.1-13.

Diagramme de Gantt

Diagramme de Gant du projet		19/2	26/2	5/3	12/3	19/3	26/3	2/4	9/4	16/4	23/4	30/4	7/5	14/5	21/5	28/5	4/6
Tâche / Semaine		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Présentation BNF																
2	Cotation Clément																
3	Définition du corpus																
4	Recherches BN Opale Plus et Gallica																
5	Mise à la disposition des dossiers à traiter																
6	Présentation Europeana																
7	Formation "Etat matériel des documents"																
8	Formation interne BnF "Internet-Veille Documentaire"																
9	Traitement des fichiers (documentaire, juridique et physique)																
10	Analyse du corpus																

Table des annexes

ANNEXE 1.....	84
ANNEXE 2.....	89
ANNEXE 3.....	90
ANNEXE 4.....	91
ANNEXE 5.....	92

Annexe 1 : fonctionnalités Europeana

La recherche

Europeana est un prototype de bibliothèque en ligne développé par la Bibliothèque nationale de France, dans le cadre du projet de Bibliothèque numérique européenne.

Europeana rassemble environ 12 000 documents libres de droits issus des collections de la BnF, de la Bibliothèque Nationale Széchényi de Hongrie et de la Bibliothèque nationale du Portugal.

Chercher, Consulter, Personnaliser : Europeana offre une large gamme d'outils et de services pour faire une recherche sur ou dans un livre, lire, imprimer ou télécharger des documents, créer sa propre bibliothèque personnelle.

Pour en savoir plus, consultez les rubriques À propos et Aide en ligne.

Pour donner votre avis, remplissez le questionnaire en ligne.

Plusieurs modes de recherche dans les collections d'*Europeana* sont possibles : la recherche par mots ou expressions, dite recherche plein texte, la recherche par critères et la recherche par thèmes.

La recherche par critères permet une découverte des collections par époque de publication, langue des documents ou bibliothèque de provenance.

La recherche par thèmes est organisée selon la classification Dewey, conçue au XIX^{ème} siècle, régulièrement mise à jour depuis et utilisée par un grand nombre de bibliothèques dans le monde pour classer leurs collections. Une variété de thèmes regroupés en 10 grands ensembles permet d'accéder à des listes de documents portant sur un même sujet.

Comment affiner les résultats ?

The screenshot shows the Europeana search interface. At the top, the Europeana logo is on the left, and a search bar contains the text 'carnets de voyages'. Below the search bar, a navigation bar shows '1071366 Résultats' and 'Aller Page 1 / 137'. On the left, a sidebar titled 'Affiner' (Refine) is open, showing filters for 'Provenance' (Bibliothèque nationale de France), 'Date' (XIXe siècle), 'Langue' (Français, Français indien, Latin, Allemand, Anglais), and 'Auteur' (Shenchi, Diderot de La Harpe, Jean-Louis, Flaubert, Gustave, Féval, Paul, Huygens, Christian). The main results area displays a list of documents with their titles, authors, and dates. Each result includes an 'Extrait' (Snippet) and a 'Sujet' (Subject) line. The results are:

- Carnet de notes d'un voyageur en France / par A. C. Poiré - 1903**: Extrait: Sujet: France — Descriptions et voyages. 100 %
- Carnets de route d'un cultivateur champenois / par M. Charles Giraut - 1912**: Extrait: Sujet: Italie — Descriptions et voyages — 19e siècle. 92 %
- Journal d'un voyage à Paris au mois d'août 1802 / par Sir John Dean Paul ; traduit et annoté, pour la Société d'histoire contemporaine, par Paul Lacombe - 1913**: Extrait: Table des matières: Accidents, Police, Fièvres, Cabriolets, 35-37, 56, 64, 72, 99 — voitures de voyage. Barouche. Sujet: Paris (France) — Descriptions et voyages. 96 %
- Voyages en tous pays / par Armand Grébauval - 1899-1903**: Extrait: Sujet: Italie — Descriptions et voyages — 1870-1914. Sujet: France — Descriptions et voyages — 1870-1914. 95 %
- Du Tchad au Dahomey en ballon : voyage aérien au long cours / Léo Dex (Deburau) - 1903**

Une fois la recherche lancée, la palette « Affiner » regroupe les résultats trouvés en sous-ensembles classés par grandes catégories :

« Provenance » correspond à la bibliothèque d'origine des documents, « Date » correspond à la date de publication des documents, « Langue » correspond à la langue des documents, « Auteur » correspond aux principaux auteurs des documents.

En cliquant sur l'un des liens proposés, l'utilisateur restreint les résultats à ce sous-ensemble. Il peut ainsi affiner progressivement sa recherche.

Cette technique est appelée « la recherche par facettes ».

Comment s'effectue la recherche dans un document ?

The screenshot displays a digital document viewer. On the left, a sidebar titled "Rechercher dans ce document" (Search in this document) contains a search input field with the word "rome" entered. Below the input, it lists "37 pages trouvées" (37 pages found) with a scrollable list of page numbers and occurrence counts: p. 140 (1), p. 155 (1), p. 158 (1), p. 171 (1), p. 177 (2), p. 178 (1), p. 169 (1), and p. 212 (2). The page number "178" is highlighted in the list. Below the list are two buttons: "Naviguer" (Navigate) and "Utiliser" (Use), each with a plus sign and a square icon. The main document area shows the title "Voyages en tous pays / par Armand Grébauval / Grébauval, Armand, 1899-1903" and a page number "178" in the top right corner. The page content includes the page number "178" and the section title "AU PAYS LATIN". The first paragraph begins with "hésitation. **A Rome**, nous avons juste le temps d'avaler un café chaud, tandis qu'on rattache notre wagon à l'autre train, le nôtre allant à Milan, un second à Gênes. Puis la marche reprend, monotone, à peine égayée par des divinations, dans la nuit étoilée." The word "Rome" is highlighted in yellow. The second paragraph starts with "Nous redescendons la vallée du Tibre, vers la Méditerranée." The third paragraph discusses the Simplon tunnel and the P.-L.-M. railway. The fourth paragraph mentions the Doublage des voies, de Calos à Modane. The fifth paragraph discusses the prestige of the railway and the barriers it overcame.

Il est possible de lancer une recherche d'un ou plusieurs mots à l'intérieur d'un document grâce à palette « Rechercher dans ce document ».

Chaque page du document contenant l'un des termes recherchés est indiquée, ainsi que le nombre d'occurrences trouvées dans celle-ci. Pour visualiser une page, il suffit de cliquer sur le lien vers cette page dans la palette. Le terme recherché est surligné en jaune dans la page.

« Ma bibliothèque »

A quoi sert « Ma bibliothèque » ?

« Ma bibliothèque » est un espace personnel qui permet d'enregistrer des références de documents ou d'indexer des pages afin de les retrouver sans avoir à renouveler une recherche.

« Ma bibliothèque » est composée de trois pages différentes : « Ma page d'accueil », qui donne une vue d'ensemble sur les données personnelles enregistrées, « Mes documents », dédiée aux documents enregistrés, « Mes étiquettes » qui comprend les pages indexées.

Tout internaute peut créer un compte à partir de la palette « Ma bibliothèque », à l'aide d'un alias (nom d'utilisateur) et d'un mot de passe.

A quoi sert « Mes documents » ?



La fonction « Mes documents » de « Ma bibliothèque » permet d'enregistrer dans l'espace personnel des références de documents et de les organiser par dossiers.

L'utilisateur peut sélectionner les documents qui l'intéressent depuis une page de résultats ou depuis l'espace de consultation en cliquant sur l'icône « Ajouter à mes documents ».

Tous les documents ajoutés apparaissent dans la page d'accueil de « Ma bibliothèque » et sur la page « Mes documents ». Pour consulter le document, il suffit de cliquer sur la référence enregistrée.

Depuis la page « Mes documents » de « Ma bibliothèque », l'utilisateur peut créer des dossiers et trier, déplacer ou supprimer les références des documents à l'intérieur de ces dossiers.

Qu'est-ce qu'une étiquette ?

The screenshot shows the Europeana user interface. At the top left is the Europeana logo with the tagline 'la contribution française à la Bibliothèque Numérique Européenne'. The main header is 'Ma bibliothèque'. On the left, there is a user profile section for 'Intemate' with an email address and a 'Modifier' button. Below that is an 'Organiser' section with a trash icon. The central area is titled 'Mes étiquettes (29)' and displays a 'tag cloud' of city names: Alexandrie, Barcelone, Belfast, Bruges, Bruxelles, Budapest, budapest, Dublin, Edimbourg, Faro, Florence, Funchal, Le Caire, Lisbonne, Liverpool, Londres, Madagascar, Madrid, Manchester, Melbourne, Moscou, Naples, New York, Porto, Rome, Tamatave, Toulon, Valence, Venise. Below the tag cloud, there is a section for 'Résultats pour l'étiquette 'Porto': 3' containing three document entries, each with a checkbox, a title, author information, and a book icon.

Le terme « étiquette » est la traduction française de l'anglais « tag ».

Une étiquette est un mot clé ou un ensemble de mots clés qui, tel un *post-it*, serait collé par l'utilisateur sur la page d'un document.

Tout utilisateur disposant d'un espace personnel peut créer une ou plusieurs étiquettes sur l'ensemble des pages des documents qu'il consulte.

Les étiquettes créées sur un même document en cours de consultation sont visibles depuis la palette « Ma bibliothèque » et permettent à l'utilisateur de naviguer dans le document selon ses propres mots clés.

Sur la page « Mes étiquettes » de son espace personnel, l'utilisateur retrouve la totalité de ses étiquettes organisées en « nuage » (« tag cloud » en anglais).

Le nuage consiste à représenter graphiquement les étiquettes en fonction du nombre de pages auxquelles chacune renvoie : plus le nombre de pages indexées est important, plus la taille de l'étiquette est grande.

En cliquant sur un mot du nuage, l'utilisateur a accès aux références des pages portant cette étiquette. Chaque référence est un lien vers la page indiquée.

Annexe 2 : volumétrie Gallica 2

Européana : Volumétrie du marché 100 000 toutes filières							
	Année du marché	Nombre de pages minimum	Nombre de pages maximum	Nombre de documents numériques	Nombre de documents numériques	Nombre d'objets minimum	Nombre d'objets maximum
Papier massicoté	1	2 040 000	2 400 000	8 500	10 000	8 000	9 500
	2	960 000	1 200 000	4 000	5 000	3 800	4 700
	3	0	0	0	0	0	0
Papier relié	1	6 500 000	8 200 000	45 000	60 000	36 500	52 000
	2	8 000 000	10 500 000	60 000	80 000	47 000	70 000
	3	10 000 000	12 000 000	75 000	90 000	57 500	77 000
Microformes	1	4 000 000	4 500 000	40 000	45 000	30 000	40 000
	2	3 000 000	3 500 000	30 000	35 000	20 000	30 000
	3	1 500 000	2 500 000	15 000	25 000	9 500	20 000

EUROPEANA : Estimation marché 100 000 collections reliées pour la DCO											
		1ère année		Total	2ème année		Total	3ème année		Total	Total gén.
		ouvrages	périodiques		ouvrages	périodiques		ouvrages	périodiques		
Unités doc. (id est exemplarisation)	mini	30 000	15 000	45 000	40 000	20 000	60 000	50 000	25 000	75 000	180 000
	maxi	40 000	20 000	60 000	50 000	30 000	80 000	55 000	35 000	90 000	230 000
UC (id est support physiq)	mini	35 000	1 500	36 500	45 000	2 000	47 000	55 000	2 500	57 500	141 000
	maxi	45 000	7 000	52 000	55 000	15 000	70 000	60 000	17 000	77 000	199 000
pages (id est numérisation) Ref : 50 pages/perio, 180 pages/mono	mini	5 400 000	750 000	6 500 000	7 200 000	1 000 000	8 000 000	9 000 000	1 000 000	10 000 000	24 500 000
	maxi	7 200 000	1 000 000	8 200 000	900 000	1 500 000	10 500 000	10 000 000	2 000 000	12 000 000	30 700 000

Annexe 3 : bilan des sélections de la tranche de cote Td

Fichier	Nb uc dans le fichier	Rejets									Total
		RED	Publications étrangères (non reproduites)	Publications étrangères reproduites	Documents microformés	Documents déjà numérisés	Absence constatée	Extraits de revues	Droits	Autres	
Td1 à 4	228	114	40	0	3	1	1	2	0	0	161
Td5 à 6	268	56	105		7	0	0	1	2	0	171
Td7 à 9	201	82	43	1	7	6	0	0	1	1	141
Td10	155	16	45	0	16	0	0	25	0	1	103
Td11	371	40	88	1	40	3	0	58	8	4	242
Td12 à 15	313	67	62	1	22	12	4	24	27		219
Td16 à 20	312	29	47	1	25	2	3	28	20	8	163
Td21 à 29	411	165	100	0	17	4	2	21	3	1	313
Td30-1 à 314	500	254	136		4	3	1			4	402
Td30-315 à 415	553	0	48		4	38		2	95	2	189
Td31 à Td33	361	115	100		22	3		41	2	2	285
Td34-401 à 670	517	212	97	3	35	2	5	40	0	27	421
Td34-671 à Td3	493	202	90		61			35	12	25	425
Td35 à 37	505	105	103	6	54	12	5	14	25	9	333
Td38 à 40	191	20	26	4	17	5	0	33	2	3	110
Total	5379	1477	1130	17	334	91	21	324	197	87	3678
% / total uc		27%	21%	0%	6%	2%	0%	6%	4%	2%	68%
% / total rejets mag											0
% / total rejets catalogue		40%	31%	0%	9%	2%	1%	9%	5%	2%	

Tableau 9 : rejets documentaires

Fichier	Rejets en magasin									Total
	Pb reliure (serrée, marges)	Corps d'ouvrage dégradé / fragile	Typographie et ouvrage <1750	Manque en place	Format	Transparences, impression	Intérêt du contenu	Autre exemplaire sélectionné	Divers	
Td1 à 4	7	1	32			6	1	1		48
Td5 à 6	12	1	52	1		2				68
Td7 à 9	18	6	1						1	26
Td10	4		2		1	1				8
Td11	34	12	2				2	1		51
Td12 à 15	15	13	19		1	1		1	3	53
Td16 à 20	15	22	34					1		72
Td21 à 29	32	7	27							66
Td30-1 à 314	76		8							84
Td30-315 à 415	13	5		1			1			20
Td31 à Td33	7					2				9
Td34-401 à 670	8	12	7			4	1			32
Td34-671 à Td3	1	6				1				8
Td35 à 37	35	63	23	1		5		1	3	131
Td38 à 40	4	28	11	1		0		1	0	45
										0
Total	281	176	218	4	2	22	5	6	7	721
% / total uc	5%	3%	4%	0%	0%	0%	0%	0%	0%	13%
% / total rejets mag	39%	24%	30%	1%	0%	3%	1%	1%		
% / total rejets catalogue										0

Tableau 10 : rejets physiques

Annexe 4 : exemple de fichier de travail

Séle- ction	Thème	Cote	sous- cote	Auteur principal	Bibli	Date édition	Qualif volumen	volum e	titre	nombre fasc.	nombre de pages	année affinée	Départ UC	Prés.	Date de	Code Com.	cote microforme de document (p. accessus)	Filigr papier/ micro	Type de document monopériod.	Examen docum. CN	Autre ex. num. CN	Droits CN	Examen physique CN	Accepté CN	Observations (voir association CCP)		
		FOL-T19-6		Petrus de Abano (1257-1315?)	Conciliator controversiarum quae inter philosophos et medicos versantur. Petrus Abano... auctore. Libellus de venenis, eodem auctore. Petri Carani quaestio de venenis, ad terminum. Symphoriani Champierii... in Conciliatore errationibus... - Petrus de Aban	1548														N		O	N	N	Publication étrangère (Venetis)		
		FOL-T19-6	A	Petrus de Abano (1257-1315?)	Conciliator controversiarum quae inter philosophos et medicos versantur. Petrus Abano... auctore. Quibus nunc in margine per numerum adijunctis rationes, quae ad partes confirmandas, confutandas, atque conciliandas pertinent. Libellus de venenis, eodem au	1566														N		O	N	N	Publication étrangère (Venetis)		
		4-T19-6		Petrus de Abano (1257-1315?)	Consiliator enucleatus... - Petrus de Abano (1257-1315?) - 1615	1615														N		O	N	N	Publication étrangère (Giessee Hassonum)		
		8-T19-6	A	Petrus de Abano (1257-1315?)	Consiliator enucleatus... - Petrus de Abano (1257-1315?) - 1621	1621														N		O	N	N	Publication étrangère (Giessee)		
		8-T19-6	BIS	Porzio, Simone	De Dolore Simonis Portii... liber - Porzio, Simone - 1554	1554														N		O	N	N	Publication étrangère (Florentiae)		
		FOL-T19-7	A	Vallés de Covarrubias, Francisco (médecin)	Controversiarum medicarum et philosophicarum Francisci Vallesii Covarrubiani editio 3a, ab eodem auctore juxta recognita et aucta. Accessit liber de Simplicium medicamentorum facultate (Galenus, cum commentariis) ejusdem auctoris... - Vallés de Covarrubias,	1583														N		O	N	N	Publication étrangère (Complut)		
		8-T19-7	B	Vallés de Covarrubias, Francisco (médecin)	Controversiarum medicarum et philosophicarum Francisci Vallesii Covarrubiani editio 3a, ab auctore denuo recognita et aucta. Accessit libellus de Locis manifeste pugnantibus apud Galenum, eodem Vallesio auctore... - Vallés de Covarrubias, Francisco (médec	1591															N		O	N	N	typographie ancienne, reliure serié	
		FOL-T19-7	C	Vallés de Covarrubias, Francisco	Controversiarum medicarum et philosophicarum Francisci Vallesii Covarrubiani libri X, accessit libellus de locis manifeste pugnantibus apud Galenum, eodem Vallesio auctore. Editio quarta... - Vallés de Covarrubias, Francisco - 1606	1606														N		O	N	N	Publication étrangère (Hanoiae)		
		4-T19-8		Vallés de Covarrubias, Francisco (médecin)	Francisci Vallesii Covarrubiani... Controversiarum medicarum et philosophicarum libri decem, quibus accessit libellus de Locis manifeste pugnantibus apud Galenum, eodem Vallesio auctore. Editio postrema... praecedentibus multo correctior... - Vallés de C.	1625															N		O	N	N	Typographie ancienne	
		4-T19-9		Marcucci, Antonio	De Principibus animae facultatibus, deque eorum sensibus ac instrumentis, Antonio Marcucci... libri duo... - Marcucci, Antonio - 1586	1586															N		O	N	N	Publication étrangère (Lucae)	
		4-T19-10		Chiocci, Andrea	Quaestionum philosophicarum et medicarum libri tres... auctore Andrea Chiocci... - Chiocci, Andrea - 1593	1593															N		O	N	N	Publication étrangère (Veronae)	
		8-T19-11		Libavus, Andreas Libau, dit	Andreas Libavi... Schediasmata medica et philosophica... [Praemissa est epistola P. Scherlii] - Libavius, Andreas Libau, dit - 1596	1596															N		O	N	N	Publication étrangère (Frankfurt)	
		8-T19-12		Horst, Jacob	Jacobi Horstii... Epistolae philosophicae et medicae, cum via Gregori Horstii... a Reinero Reinuccio. Addita sunt consolatoria Joannis Caselli et Nathani Chytraei ad Jacobum Horstium fratrem... - Horst, Jacob - 1596	1596															N		O	N	N	Publication étrangère (Lipsiae)	
		4-T19-13		Lucius, Cyriacus	De Medicina philosophica. Cyriacus Lucius... ostenditur quod scientia medica, liberibus disciplinis omnibus adeoque toti literario sapientiae studio... usui esse queat, talisque in usum hoc tempore nostro converti debeat... - Lucius, Cyriacus - 1597	1597															N		O	N	N	Publication étrangère (Ingolstadt)	
		8-T19-14	1		libro biblo non disponibus															N				N	RED		
		8-T19-14	3		libro biblo non disponibus																N				N	RED	
		8-T19-14	4		libro biblo non disponibus																N				N	RED	
		4-T19-15		Persono, Giovanni Battista	Jo. Baptistae Personae... In Galeni lib- ri titulus est Quod animi mores corporis temperiem sequantur commentarius... - Persono, Giovanni Battista - 1602	1602																N		O	N	N	Publication étrangère (Bergomi)
		8-T19-16		Gubelet, Jourdan	Trois discours philosophiques : le I, de la comparaison de l'homme avec le monde ; le II, du principe de la génération de l'homme ; le III, de l'humour mélancolique, mis de nouveau en lumière par Jourdan Gubelet - Gubelet, Jourdan - 1603	1603															N		O	N	N	Microforme	
		8-T19-17		Burée Duide, Pierre	La Défence de l'escolle de médecine et de Galien contre M. J. Gubelet, d'Evreux, docteur en médecine, et advers sur ses trois discours philosophiques, par Pierre Burée... - Burée Duide, Pierre - 1605	1605																N		O	N	N	typographie ancienne, marge < 1 cm
		4-T19-18		Sala, Giovanni Domenico	De Natura medicinae libellus Jo. Domenici Sala... In quo prolegomena explicuntur... - Sala, Giovanni Domenico - 1625	1625																N		O	N	N	Publication étrangère (Patavi)
		4-T19-19		Grandi, Adriano	Quaestiones philosophicae et medicae... Adriani Grandi... - Grandi, Adriano - 1629	1629															N		O	N	N	Publication étrangère (Veronae)	
		4-T19-20		Laurenberg, Peter Villumsen	Petri Laurenbergii... Porticus Aesculapi, seu Generalis artis medicae constitutio, in qua de medicinae genere, subjecto, fine, definitione, divisione controversiae phaeraeque omnes examinantur... - Laurenberg, Peter Villumsen - 1630	1630															N		O	N	N	Publication étrangère (Rosochi-Rostoc k)	
		4-T19-21		Meysonnier, Lazaro	Lazari Meysonnier... Pentagonum philosophico-medium, sive Ars nova remisscentiae, cum institutionibus philosophicae naturalis et medicae sublimioris et secretioris... ; clavis... omnium arcanorum naturalium Macrocosmi et Microcosmi... Meysonnier, La	1639																N		O	N	N	Typographie ancienne
		8-T19-22		Chiaramont, Scipione	De Atra Bile, quoad mores atinet, libri tres Scipionis Chiaramontii... ad clarissimum virum Gabrielem Naudaeum... - Chiaramont, Scipione - 1641	1641																N		O	N	N	Microforme
		8-T19-23		Leichner, Eoccard	Eoccardi Leichneri... de Individuabili et totali cujusque animae in toto suo corpore et singulari ejus partibus existentia, dissertatio tripartita... - Leichner, Eoccard - 1660	1660																N		O	N	N	Publication étrangère (Erfurt)

Annexe 5 : Consignes pour le traitement des fichiers de sélection

Bibliothèque nationale de France
Département Sciences et techniques

Bnue – Marché des 100 000

Consignes pour le traitement des fichiers de sélection

I. A savoir

En attendant la mise en place de l'outil de sélection spécialement conçu pour les sélections / vérifications / validations des documents à numériser dans le cadre du marché des 100 000, le travail s'effectue à partir de fichiers Excel.

! le format de ces fichiers Excel ayant été défini en accord avec le Département de la coopération, leur structure (nombre de colonnes ...) ne doit pas être modifiée. Il est cependant possible d'appliquer des filtres ou de masquer des colonnes, notamment pour l'impression des listes.

Ces fichiers sont constitués à partir de requêtes faites sur les uc comprises dans une tranche de cotes prédéterminée (par exemple, toutes les uc de la tranche 8-S-1 à 8-S-500), sur le modèle des batch de récolement ; ils comprennent donc une ligne par uc.

A noter qu'en cas de rupture basse d'épine dorsale (pas d'uc associée à une cote), la cote en question n'apparaît pas dans le fichier ; au contraire en cas de rupture

haute d'épine dorsale (pas de notice bibliographique associée à l'uc), l'uc apparaît bien dans le fichier mais on retrouve dans la colonne « biblio » la mention « infos non disponibles ».

Chaque fichier comprend un grand nombre de colonnes, certaines remplies automatiquement lors du batch, d'autres à remplir et enfin d'autres qui peuvent rester vides.

Les données figurant dans les fichiers

Les fichiers qui nous sont transmis comprennent donc un certain nombre de données :

❖ Données bibliographiques et d'exemplaire

- cote
- sous-cote éventuellement
- auteur principal
- biblio
- date d'édition
- qualif volumaison
- volume
- tome

Ces données sont extraites de ce qui figure dans les notices bibliographiques et les données d'exemplaire ; elles peuvent donc être incomplètes.

! le classement « excel » fait qu'en général la sous-cote figure dans la ligne avant la cote pure (donc on a par exemple en première ligne 8-Td37-4 (A) et dans la ligne en dessous 8-Td37-4).

Exemple de données :

Cote	sous-cote	Auteur principal	Biblio	Date édition	Qualif volumaison	volume	tome
------	-----------	------------------	--------	--------------	-------------------	--------	------

8-TD18-4	A,1	Laënnec, René-Théophile-Hyacinthe (1781-1826)	Traité de l'auscultation médiate et des maladies des poumons et du coeur - Laënnec, René-Théophile-Hyacinthe (1781-1826) - 1837	1837			
8-TD18-4	A,2	Laënnec, René-Théophile-Hyacinthe (1781-1826)	Traité de l'auscultation médiate et des maladies des poumons et du coeur - Laënnec, René-Théophile-Hyacinthe (1781-1826) - 1837	1837			
8-TD18-4	A,3	Laënnec, René-Théophile-Hyacinthe (1781-1826)	Traité de l'auscultation médiate et des maladies des poumons et du coeur - Laënnec, René-Théophile-Hyacinthe (1781-1826) - 1837	1837			
8-TD18-4	B	Laënnec, René-Théophile-Hyacinthe (1781-1826)	Traité de l'auscultation médiate et des maladies des poumons et du coeur, par R.-T.-H. Laënnec,... - Laënnec, René-Théophile-Hyacinthe (1781-1826) - 1879	1879			
8-TD18-4	C	Laënnec, René-Théophile-Hyacinthe (1781-1826)	Traité de l'auscultation médiate et des maladies des poumons et du c\0153ur - Laënnec, René-Théophile-Hyacinthe (1781-1826) - 1828	1828			

8-TD18-4	D,1		Infos biblio non disponibles				
8-TD18-4	D,2		Infos biblio non disponibles				

❖ Données d'uc

- détail uc : pour les périodiques
- année d'affinage : également pour les périodiques
- code présence et date d'absence constatée (si le cas se présente)
- code com (permet de repérer les ouvrages reproduits sur microfiches ou microfilms, le code étant alors à « 13 »)
- code-barres
- adresse

La coordination DCO, avant de nous transmettre les fichiers, effectue un premier nettoyage et signale notamment :

- les ouvrages en rupture d'épine dorsale (indication RED dans la colonne observations).
- les documents ne posant aucun problème de droits (par défaut, tous les ouvrages publiés avant 1870 sont réputés libres de droits)
- les uc reproduites : à partir de la présence du nombre 13 dans la colonne « code droit ».

II. Consignes

Les vérifications à effectuer dans le catalogue Bn-opale plus

La première opération consiste à « éliminer » les ouvrages ne répondant pas aux critères documentaires ou de droits définis pour Gallica 2 afin de ne garder que les ouvrages à aller vérifier en magasin. On exclut donc des vérifications en magasins les :

- publications étrangères
- ouvrages sous droits
- documents déjà numérisés
- extraits de revues, sauf exceptions
- réimpressions multiples à l'identique.

On écartera également dans un premier temps les uc déjà reproduites, sur microfiche ou microfilm (car des campagnes massives de numérisation à partir des microformes sont actuellement engagées par Gallica).

Afin d'effectuer ce travail, l'outil de travail principal est le catalogue Bn-Opale plus. Pour chaque cote, il convient d'interroger le catalogue et de vérifier :

- le lieu d'édition. En cas de publication étrangère, mettre N dans la colonne « Examen docum. » (pour examen documentaire), indiquer « publication étrangère » dans la colonne « Observations » et ne pas poursuivre les vérifications au-delà

! pour toutes les colonnes de vérifications : « O » s'entend comme OK, il s'agit d'une acceptation de la validation de l'uc au regard du critère examiné (donc si on met « N » dans la colonne « Autre exe num », il faut comprendre que le document est rejeté à cause du critère « autre exe num » car il existe un exemplaire numérique).
Le fond de case se colore automatiquement en vert quand on entre la valeur « O » et en orange quand on tape « N ».

- la date d'édition : la période de publication privilégiée est dans un premier temps 1750 – 1930 ; on peut envisager de numériser des ouvrages antérieurs à 1750 mais il conviendra lors des vérifications en magasin d'être particulièrement vigilant et de s'assurer que la typographie de ces ouvrages ne posera pas de problèmes pour le passage de l'OCR (s ressemblant à des f par exemple)
- si la date de décès de l'auteur (ou des auteurs) n'est pas mentionnée, laisser la colonne « Droits » vide : Gallica se chargera par la suite d'effectuer des vérifications complémentaires. Si la date de décès figure dans le catalogue, soit elle est antérieure à 1937 et dans ce cas mettre « O » dans la colonne « Droits », soit elle est postérieure et alors mettre « N ». Pour les ouvrages sous droits, arrêter là les vérifications.
- l'existence d'autres exemplaires :
 - o microforme : en cas de document reproduit, mettre « N » dans la colonne « examen documentaire », préciser « microforme » dans la colonne « observations » et compléter si possible la colonne « cote microforme du document » avec le code prestataire (s'il figure dans le catalogue – on

- peut compléter la recherche en vérifiant dans A-dcm-26 – en cas de recherche infructueuse dans le catalogue et A-dcm-26, laisser la case vide)
- exemplaire numérique : en cas d'ouvrage déjà numérisé, mettre « N » dans la colonne « Autre exe num », indiquer de préférence la cote de cet exemplaire numérique dans la colonne « observations » et stopper là les vérifications
 - autres exemplaires papier : il peut être intéressant de mentionner l'existence d'autres exemplaires papier dans les collections de la BnF (au cas où l'exemplaire figurant dans le fichier que l'on est en train de vérifier ne serait pas numérisable).

Concernant l'existence d'autres éditions : on n'exclut pas d'office les différentes éditions d'un ouvrage (surtout qu'il peut être intéressant d'avoir les éditions successives d'un document, quand il s'agit de véritables nouvelles éditions, donc revues, augmentées, corrigées ...); on exclut par contre les réimpressions à l'identique quand cela est mentionné dans la notice.

On exclura également a priori les périodiques ne figurant pas dans la liste de sélections établie au printemps. Si un ou plusieurs titres figurant dans le fichier semblent présenter un réel intérêt documentaire, bien qu'absents de la sélection de printemps, ils pourront être conservés dans le fichier et vérifiés.

Pour les documents n'ayant pas été exclus lors de ces premières vérifications, indiquer dans la colonne « type de document », s'il s'agit d'une monographie ou d'un périodique.

Les vérifications à effectuer en magasin

Une fois la liste des Uc à aller vérifier établie, la seconde étape consiste à :

- vérifier en magasin la numérisabilité des ouvrages, toujours selon les mêmes critères physiques (ouverture des volumes, état du corps d'ouvrage ...). Pour les ouvrages anciens, il faut également vérifier la typographie. Le marché des 100 000 permet de numériser des documents jusqu'au format A1 (donc désormais ne pas exclure de documents pour cause de dépliant de format A2).
 - En cas de refus, mettre « N » dans la colonne « Examen physique » et indiquer la raison du refus dans la colonne « observations »

- Si le document est numérisable, mettre « O » dans la colonne « Examen physique »

- indiquer, pour les uc numérisables, le nombre de pages de l'uc

- noter pour les uc de périodiques le nombre de fascicules contenus dans l'uc. De même pour les uc de monographies où plusieurs sous-cotes ou cotes sont reliées ensemble, indiquer dans la colonne « nb fascicules » le nombre de monographies reliées ensemble.

Si, lors des vérifications en magasin, vous trouvez un document qui aurait été écarté pour cause de rupture d'épine dorsale mais qui vous paraît intéressant (et à condition qu'il soit numérisable), vous pouvez le signaler dans le fichier : les corrections nécessaires seront effectuées afin de pouvoir l'envoyer en numérisation.