



HAL
open science

La dématérialisation des dossiers documentaires : les enjeux et les techniques : étude de faisabilité réalisée pour le centre de documentation des Arts Décoratifs

Ludovic Desaubry

► To cite this version:

Ludovic Desaubry. La dématérialisation des dossiers documentaires : les enjeux et les techniques : étude de faisabilité réalisée pour le centre de documentation des Arts Décoratifs. domain_shs.info.docu. 2009. mem_00523899

HAL Id: mem_00523899

https://memic.ccsd.cnrs.fr/mem_00523899v1

Submitted on 6 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



sciences
et
techniques
de
l'information

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE pour obtenir le
Titre professionnel "Chef de projet en ingénierie documentaire" INTD
Niveau I

présenté et soutenu par

Ludovic DESAUBRY

le 16 novembre 2009

La dématérialisation des dossiers documentaires :
les enjeux et les techniques

Etude de faisabilité réalisée pour le centre
de documentation des Arts Décoratifs

Jury composé de :

Nadia RAÏS

Véronique SEVESTRE

Promotion XXXIX

Remerciements

Je tiens à remercier Véronique Sevestre et Michèle Jasnin pour leur accueil chaleureux au centre de documentation des musées des Arts Décoratifs.

J'adresse également mes remerciements à Séverine Denys et à Ana Parodi qui m'ont accordé un peu de leur temps pour répondre à mes questions et partager leur expérience.

Enfin je remercie vivement toute l'équipe pédagogique de l'INTD, et particulièrement Nadia Raïs pour ses judicieux conseils et ses encouragements au cours de la préparation et de la rédaction de ce mémoire.

Table des matières

Introduction	8
La dématérialisation	12
1.1 Définitions et principes	13
1.2 Les étapes clés de la dématérialisation.....	15
1.3 Les contraintes liées à la spécificité des documents numériques.....	19
1.4 La mise en place d'un projet : méthodologie, gestion de projet, freins humains	20
La Gestion Electronique de Documents.....	25
2.1 Définitions et principes	26
2.2 La mise en place et les enjeux d'une GED	29
2.3 Les étapes de la chaîne de traitement du document.....	31
2.3.1 L'acquisition numérique	31
2.3.2 Le formatage.....	37
2.3.3 Le traitement	41
2.3.4 L'indexation.....	45
2.3.5 Le stockage.....	49
2.3.6 La recherche	52
2.3.7 La consultation	53
2.3.8 La diffusion.....	55
2.4 L'offre logicielle de GED.....	56
Les dossiers documentaires au centre de documentation des musées des Arts Décoratifs	59
3.1 Présentation des Arts Décoratifs.....	60
3.1.1 L'institution	60
3.1.2 Le musée de la Publicité	60
3.1.3 Le centre de documentation des musées des Arts Décoratifs.....	61

3.1.4	La documentation de la publicité	62
3.2	L'analyse de l'existant	64
3.2.1	Les dossiers thématiques de la publicité	64
3.2.2	La chaîne documentaire	65
3.2.3	Recensement de l'existant et typologie des dossiers thématiques de la publicité.....	66
3.2.4	Recensement de l'existant et typologie des dossiers dans les autres documentations.....	69
3.2.5	L'initiative d'un dépouillement électronique de certains périodiques à l'aide du logiciel Mobytext	70
3.3	Analyse de contenus de boîtes d'archives	72
3.4	L'inadéquation des logiciels Mobydoc pour la gestion de dossiers thématiques dématérialisés	77
3.5	L'analyse des besoins et les bénéfices attendus	80
3.5.1	Amélioration du produit documentaire et du service aux utilisateurs.....	81
3.5.2	Réduction de la consommation de papier	81
3.5.3	Gain de place	82
3.5.4	Gain de temps	83
3.5.4.1	Alimentation des dossiers : lecture et dépouillement du périodique, numérisation physique des pages, découpage numérique de l'article, indexation des articles, métadonnées, classement et rangement des documents.....	83
3.5.4.2	Impact sur le temps passé à renseigner les consultants internes	89
3.5.4.3	Impact sur le temps passé à renseigner le public extérieur.....	89
3.5.5	La pérennité du fonds numérisé, et la question du stockage physique des données	91
3.5.6	La gestion des processus, ou « workflow ».....	93
3.5.7	L'abolition des contraintes spatiales, et de nouvelles formes de collaborations et d'échanges entre les conservateurs et les documentalistes	95
3.5.8	L'opportunité d'un désherbage et d'une redécouverte du fonds documentaire	96

3.5.9	Garantir la sécurité de la partie du fonds documentaire ayant une très grande valeur de rareté	96
3.6	Comptes rendus des entretiens avec les documentalistes des autres documentations.....	98
3.6.1	Compte rendu de l’entretien avec la documentaliste responsable de la documentation de la mode et du textile.....	98
3.6.2	Compte rendu de l’entretien avec la documentaliste responsable de la documentation du 20 ^{ème} siècle (le design, les créateurs, le bijou, la céramique)	102
3.6.3	Compte rendu de l’entretien avec la documentaliste responsable de la documentation Art nouveau Art déco.....	104
3.6.4	Compte rendu de l’entretien avec la documentaliste responsable de la documentation du 17 ^{ème} et du 18 ^{ème}	105
3.7	Préconisations techniques complémentaires.....	107
3.7.1	En matière de scanner	107
3.7.1.1	Le choix du meilleur segment de scanners.....	108
3.7.1.2	La technologie des capteurs : généralisation des CCD en matière de GED	109
3.7.1.3	Les caractéristiques fondamentales.....	109
3.7.1.4	Logiciel d’acquisition fourni et pilotes standardisés.....	112
3.7.2	En matière de logiciel de GED	112
3.7.2.1	L’architecture de la solution	113
3.7.2.2	L’administration de la solution.....	113
3.7.2.3	Le gestionnaire de base et les capacités de stockage	114
3.7.2.4	La déclinaison de l’interface utilisateur	114
3.7.2.5	Une configuration multi-utilisateurs	115
3.7.2.6	La question de la sécurité et le cryptage	117
3.7.2.7	Un outil de création de plan de classement.....	117
3.7.2.8	Le bulletinage	117
3.7.2.9	Le standard XML (eXtensible Markup Language) pour l’échange de données, et la souplesse offerte par la Déclaration de Type de Document (DTD)	117

3.7.2.10	La problématique de la migration des langages d'indexation documentaire et des métadonnées	118
3.7.2.11	L'utilisation des codes à barres.....	120
3.7.2.12	La formation aux nouveaux outils.....	120
3.7.2.13	L'évolutivité comme critère de choix de la solution.....	120
3.8	Les contraintes et les stratégies de la dématérialisation.....	122
3.8.1	La numérisation rétrospective du stock des dossiers thématiques de la publicité.....	122
3.8.1.1	Contraintes et spécificités techniques liées à la dématérialisation	123
3.8.1.2	Organisation et planification de la dématérialisation rétrospective du stock	127
3.8.1.3	Cas d'une dématérialisation rétrospective en interne	127
3.8.1.4	Cas d'une dématérialisation par un prestataire en externe.....	128
3.8.2	Le contexte d'un musée, et le prestige du support	130
3.8.3	S'abonner aux versions électroniques de certains périodiques .	131
3.8.4	La perspective d'une ouverture plus large au public externe....	132
3.9	Une démarche qualité, des indicateurs de qualité.....	134
3.10	Les aspects juridiques du projet	137
3.10.1	Le droit de reproduction	137
3.10.2	Les dernières modifications apportées au « droit de reprographie »	139
3.10.3	Les « panoramas de presse » de la documentation publicité	142
3.10.4	Les aspects juridiques des panoramas de presse électroniques	144
Conclusion.....		149
Bibliographie		153
Annexes		160
Annexe 1	: Numérisation et numériseurs.....	161
Annexe 2	: Les supports de stockage	167
Annexe 3	: Présentation de l'offre logicielle EverSuite.....	168
Annexe 4	: Extrait de la liste des publications dont le CFC gère les droits pour les reproductions et représentations numériques.....	178

Annexe 5 : Organigramme des Arts Décoratifs	179
Annexe 6 : Le guide de bonne conduite de l'Aproged	180
Annexe 7 : Extrait du plan de classement des dossiers thématiques de la publicité.....	182
Annexe 8 : La structure des boîtes d'archives.....	184
Annexe 9 : Comparaison entre un système de GED et un SAE.....	188
Annexe 10 : Tests de numérisation et tests OCR	189
Répertoire des sigles utilisés.....	217

Introduction

Les continus progrès technologiques accélèrent sans cesse le développement et la prédominance de ce que l'on nomme communément la "société d'information", au cœur de laquelle se trouvent tous les centres de documentation.

Dans le cadre d'une utilisation exclusivement en interne, ou au contraire d'une mise à disposition pour un public extérieur, de nombreux centres de documentation constituent des dossiers documentaires afin de pouvoir répondre à des besoins d'informations. Apparus à la fin du 19^{ème} siècle, leur élaboration occupe aujourd'hui encore une place centrale dans le travail des documentalistes, qu'il s'agisse de dossiers outils – régulièrement tenus à jour pour répondre aux besoins permanents des usagers – ou de dossiers produits – élaborés sur mesure pour répondre à la demande précise d'un usager particulier.

Non seulement le développement rapide du numérique et l'avènement des nouvelles technologies de l'information et de la communication n'ont pas remis en cause la fonction et la finalité de ce produit documentaire - qui reste plus que jamais d'actualité et dont les formes sont désormais encore plus variées -, mais ils ont permis surtout de lui apporter une forte valeur ajoutée.

Alors que la plupart des centres de documentation sont déjà informatisés, tous ceux qui sont dotés de logiciels documentaires n'ont pas encore numérisé leur fonds de dossiers documentaires. Néanmoins, un grand mouvement de numérisation est en marche, et les projets de dématérialisation fleurissent actuellement.

Parmi les raisons qui expliquent ce progressif mais régulier passage des dossiers papier au dossiers électroniques, on citera en tout premier lieu l'évolution rapide et importante de la Gestion Electronique de Documents (GED) et de toutes les technologies associées, permettant de travailler efficacement sur des documents en texte intégral, de mettre en place un archivage électronique des documents, et d'instaurer de nouvelles habitudes de travail collaboratif grâce notamment au concept du workflow. Comme autre facteur favorable à la dématérialisation, on citera ensuite la création de normes et de techniques de structuration des documents qui permettent de se dégager du support et du logiciel – par exemple le eXtensible Markup Language (XML) et les métadonnées. Enfin, plus généralement, l'impact d'Internet et des intranets amène logiquement les documentalistes à proposer leurs produits et services sous une forme électronique.

Ainsi, parmi les nombreux avantages et bénéfices d'une exploitation des dossiers documentaires au format électronique, on mentionnera notamment : une amélioration du produit documentaire – dossiers mieux classés, mieux décrits, mieux indexés, etc - et du

service aux utilisateurs, la mise à disposition plus directe et plus rapide de la documentation, la disponibilité et une accessibilité hors des contraintes temps/espace, et le partage de l'information en temps réel. Pour certains projets, le passage aux dossiers numériques représentera aussi l'opportunité d'une réflexion sur le positionnement et la stratégie du centre de documentation en termes de produits, de services et de cibles.

Ce mémoire offre une étude et une réflexion qui se focalisent essentiellement sur les techniques et les enjeux du processus de dématérialisation, plutôt que sur l'objet de cette dématérialisation en lui-même. Notamment, les possibilités nouvelles qu'apporte la dématérialisation pour la conception, l'élaboration ou la présentation des dossiers documentaires, en termes de contenus, n'ont pas été incluses dans le périmètre du sujet de ce mémoire.

La dématérialisation ne s'inscrit pas seulement dans une démarche purement technique, comme on pourrait le penser au premier abord, mais elle comporte aussi une forte dimension humaine et organisationnelle. Elle n'est pas seulement synonyme de changement de support de l'information, elle offre aussi la possibilité d'une mise en place de nouvelles procédures et de nouveaux outils de travail. Et le passage des dossiers papier au dossiers numériques nécessite et induit presque toujours une refonte méthodologique du travail documentaire, une réorganisation du système de travail et du mode de fonctionnement, une forte implication et un changement de mentalité.

En effet, dans le cadre de tels projets, les documentalistes sont ainsi confrontés à toute une série de questions, d'autant que les évolutions sont rapides en matière de dématérialisation et qu'il n'est pas aisé – lorsqu'on ne possède ni formation ni expérience préalable dans le domaine – de connaître les procédures, les techniques et les normes les plus adéquates au projet que l'on souhaite réaliser.

Ce mémoire se base sur le cas du centre de documentation des musées des Arts Décoratifs, où j'ai mené une étude de faisabilité et d'opportunité pour un projet de dématérialisation des dossiers thématiques de la documentation publicitaire.

La dématérialisation est un processus qui nécessite et implique l'utilisation d'outils, et l'étude proposée ici porte donc très largement sur la GED. Ainsi, la première partie de ce mémoire, consacrée à la dématérialisation en tant que processus et projet, revêt un caractère plutôt théorique, alors que la deuxième partie, présentant en quelque sorte la GED comme l'instrument de cette dématérialisation, offre un aspect plus technique et pratique.

Ces deux parties se basent toutefois sur une structure relativement similaire, à savoir : un rappel des principes et des définitions, une présentation des différentes étapes, et la proposition de préconisations, d'éléments d'aide et d'informations pour la mise en place et la

méthodologie. Elles décrivent ainsi les principaux outils, techniques, méthodes et normes actuellement disponibles et utilisés par les organismes se lançant dans des projets de numérisation de leur fonds documentaire.

Néanmoins, certains aspects de la GED - sujet vaste et touchant à beaucoup de domaines de l'infodoc - n'ont pas été développés dans ce mémoire, notamment : la question des traitements linguistiques et sémantiques, les caractéristiques des moteurs de recherche proposés par les différents éditeurs, la problématique de la mise en place et de l'exploitation des portails, et la migration des métadonnées vers les applications de GED.

La troisième partie est, quant à elle, consacrée au projet du centre de documentation des musées des Arts Décoratifs. Et dans le cadre de ce mémoire, destiné à des non-spécialistes de la dématérialisation, le choix a donc été fait de réaliser une étude relativement ciblée, la plus concrète et pragmatique possible, en imaginant que la documentation de la publicité pourrait servir de projet pilote à une future extension de la dématérialisation aux autres documentations du centre. Cette troisième partie expose donc les analyses détaillées – l'analyse des besoins, l'analyse de l'existant et des contraintes, les craintes et les attentes, etc - et propose toute une série de préconisations – techniques, organisationnelles. Les dossiers thématiques de la documentation publicité étant alimentés essentiellement par des panoramas de presse, ce mémoire traite également des aspects juridiques du projet liés au droit d'auteur et au droit de reproduction, et souligne les différences significatives qui existent entre le contrat d'exploitation pour une version papier et celui pour une version électronique. Enfin, la problématique des résistances humaines et culturelles à la numérisation a aussi été abordée dans cette troisième partie, puisque, notamment, l'analyse de l'existant et des contraintes pour les autres documentations du centre a très largement porté sur les appréhensions des documentalistes et des conservateurs, plus que sur la typologie et l'état des ressources de ces documentations. Les comptes rendus des entretiens qui ont été menés avec tous les documentalistes responsables de ces différentes documentations ont ainsi fait l'objet d'un chapitre entier de ce mémoire.

Première partie

La dématérialisation

1.1 Définitions et principes

La dématérialisation est aujourd'hui une réalité qui touche de nombreux organismes, dont les centres de documentation qui se lancent dans des opérations de numérisation de leurs fonds documentaires. Et l'on constate une grande variété de ces projets de dématérialisation qui résultent d'objectifs différents et de choix techniques variés.

Mais que signifie exactement la notion de « dématérialisation » ?

Dématérialiser signifie transférer sur support numérique des types d'informations qui existaient jusque-là sous forme analogique, c'est-à-dire sur des supports dits traditionnels, le plus souvent le papier, mais aussi le film ou le microfilm.

Fruit de l'évolution naturelle des technologies de l'information, elle peut être à la fois active dans le sens où elle représente l'action de dématérialiser un document à l'origine sous forme papier pour le rendre numérique, et statique en n'opérant aucune action mais simplement par la transposition de processus jusque-là destinés à traiter des documents papier en processus ne relevant plus que de documents électroniques, immatériels.

L'information numérique présente deux caractéristiques fondamentales :

- elle n'est pas lisible ou visible à l'œil nu ;
- elle est exclusivement codée à l'aide de 0 ou de 1 et est indépendante du support ;

Ses principaux avantages sont :

- la possibilité de partager aisément l'information ;
- le peu de place physique qu'elle exige en comparaison des supports traditionnels ;
- l'utilisation de puissants et efficaces outils de recherche, d'accès et de gestion ;

Son inconvénient principal est qu'elle rend tributaire d'un équipement souvent complexe.

Ainsi, depuis quelques années, le choix de certains centres de documentation de numériser une partie de leurs documents s'explique par la convergence de plusieurs facteurs :

- des possibilités offertes par des technologies performantes ;
- l'augmentation continue des volumes papier depuis des dizaines d'années avec les problèmes de gestion et de stockage afférents ;
- des besoins d'échange et d'accès partagé à l'information ;

Les arguments de faisabilité et d'opportunité des projets de dématérialisation au sein des centres de documentation découlent donc tous des potentialités offertes par la numérisation.

Mais ils peuvent naturellement différer selon les centres de documentation, de par :

- les contextes et les organisations de ces derniers ;

- la nature des documents concernés par cette numérisation ;

Néanmoins, les arguments les plus fréquents et les bénéfiques attendus – au niveau des documents, des documentalistes et des utilisateurs - qui reviennent le plus souvent sont :

- un gain de place ;
- une plus grande facilité de consultation ;
- une meilleure présentation ;
- une communication des documents ;
- un gain de temps dans la recherche, qui sera plus facile et plus rapide (recherche multicritères, recherche en texte intégral, plan de classement électronique, etc) ;
- un gain de temps grâce à une plus grande autonomie des utilisateurs ;
- la conservation (notamment la préservation de documents fragiles) ;

Lorsqu'il est établi que la dématérialisation paraît opportune et rentable, se pose alors la question de la stratégie à adopter : une numérisation rétrospective du stock des dossiers – et dans ce cas faut-il confier l'opération à un prestataire spécialisé ? -, ou une numérisation à période fixe - tous les mois, tous les trimestres -, ou plutôt numérisation au fil de l'eau - notamment dans le cours normal de traitement de l'information.

Voici un certain nombre de concepts que recouvre la dématérialisation :

- transformer des objets physiques en en objets numériques ;
- automatiser, par des logiciels, des traitements réalisés par l'homme ;
- améliorer la qualité et la rapidité des échanges entre personnes par l'utilisation de moyens techniques de communication tels que les espaces de travail partagés ;

Pour chacun de ces concepts il convient :

- d'identifier les outils nécessaires ;
- de définir les modes de mise en place adéquats ;
- d'établir une véritable gestion de projet ;
- de cerner toutes les contraintes ;
- d'essayer le plus objectivement possible de jauger de l'utilité ;

Enfin, quel que soit le choix technologique retenu au final, on peut citer un certain nombre de principes communs à tous les projets de dématérialisation :

- la capture (technique de scannage) ;
- le stockage ;
- l'interprétation du document dématérialisé (production de caractères indexables) ;
- la gestion des métadonnées et identifiants ;
- l'exploitation et la valorisation du document numérisé ;

1.2 Les étapes clés de la dématérialisation

Tout d'abord il est important de rappeler que la dématérialisation est une des facettes de la Gestion Electronique de Documents (GED) – l'autre étant la gestion des processus ou workflow – et que tout projet de ce type induit donc la mise en place d'un véritable outil de GED - celle-ci sera définie et exposée dans la deuxième partie de cette étude. La réussite du projet est aussi conditionnée par une prise en compte pertinente de facteurs très divers, de l'existant documentaire au réseau informatique.

La transformation des objets physiques en une version numérique implique un certain nombre d'étapes, qui sont nécessaires et imposées :

- préparation des documents ;
- processus physique de numérisation ;
- indexation des documents ;
- stockage ;
- contrôle ;

Voici un certain nombre de repères autour desquels on pourra développer et bâtir un projet de dématérialisation.

• La connaissance de l'existant

La connaissance de l'existant passe par l'établissement d'une liste, la plus exhaustive possible, des divers types de documents à numériser. Pour chaque type de document il est également nécessaire de donner une évaluation de la qualité et de l'état de conservation. A partir de ce constat il est alors possible de fixer des critères de choix pour le numériseur de documents. Ces critères ont été regroupés dans un tableau que l'on peut consulter en annexe 1, à la page 161 de ce document.

Pour chaque type de document :

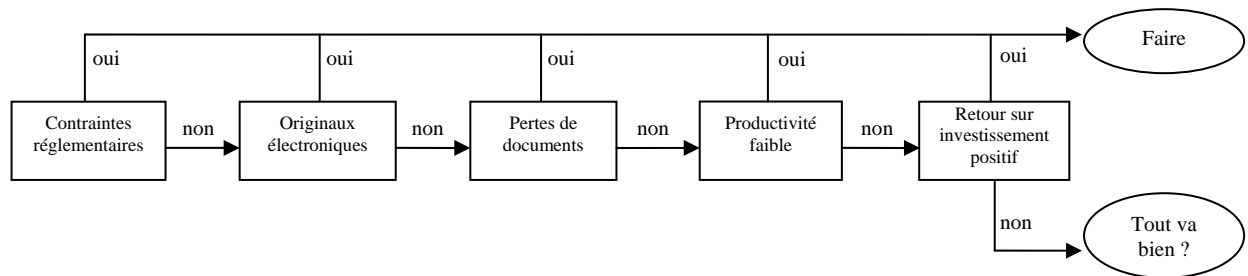
- on définira les éléments permettant de sélectionner l'appareil le plus adapté ;
- on précisera les informations nécessaires à leur manipulation (description, recherche) ;

On doit aussi lister les métadonnées ou les index qui seront saisis lors de la numérisation. En fonction de l'état du document, de la nature des métadonnées, il faudra par exemple essayer de déterminer si une technologie de lecture automatique est recommandée et si elle est réellement utilisable avec profit.

• L'évaluation de l'opportunité du projet

L'évaluation de l'opportunité du projet passe notamment par la prise en compte de trois facteurs primordiaux que sont le retour sur investissement, les risques et la réglementation. Dans le cas d'un centre de documentation, et plus particulièrement celui de la numérisation de dossiers documentaires, il semble que ce soit plutôt le retour sur investissement qui sera le facteur prépondérant et déterminant dans la prise de décision finale.

Voici un petit schéma décrivant le processus de décision :



Toujours dans le cas d'un centre de documentation, qui concerne notre étude, l'analyse de ce schéma fait apparaître que les deux points qui feront toujours pencher du côté du « oui » sont :

- la perte (égarement, vol) de documents ;
- le retour sur investissement, sous forme d'une amélioration de la qualité du service et du produit documentaire, ainsi que d'un gain de temps ;

Sur le plan de la qualité – autrement dit des normes, et qui rejoint aussi la notion de « risques » abordée plus haut -, il faudra ensuite s'interroger pour essayer de déterminer si le contexte du projet de dématérialisation fait qu'il s'inscrit totalement ou non dans le cadre de la norme NF Z42-013¹ pour le stockage permanent de documents électroniques sur support optique non réinscriptible. Le respect des recommandations de cette norme, publiée par l'AFNOR², relatives à la gestion, au stockage et à la restitution des documents électroniques garantit la sécurité, l'intégrité, la conformité réglementaire et la pérennité des documents stockés dans le système.

A priori, et en toute logique, tout investissement et tout effort déployés pour dématérialiser et capitaliser un fonds documentaire devraient aussi s'accompagner d'une solution d'archivage électronique sécurisé et fiable. Ceci afin d'éviter toute perte accidentelle de données, et préserver l'intégrité des documents. On reviendra plus en détails sur ce point dans la deuxième partie de cette étude consacrée aux solutions de GED, pour montrer que ces

¹ Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes.

² Association Française de Normalisation.

dernières autorisent et permettent presque toutes de sauvegarder des données sur des supports optiques non réinscriptibles de type WORM³, mais dans le cadre d'un stockage classique et non d'un véritable archivage électronique – au sens de l'archivage des archivistes. On fera donc la distinction entre un système de gestion électronique de documents et un système d'archivage électronique (SAE). Et les responsables d'un projet de numérisation doivent ainsi se demander si ce dernier implique ou non la conception et la mise en place d'un véritable système d'archivage électronique (SAE), qui pourra accueillir les documents numérisés immédiatement après leur transformation.

Dans le cadre d'une étude de faisabilité de dématérialisation, un autre point très important sur lequel il est nécessaire de se prononcer est le suivant : faut-il procéder à la numérisation en interne, ou plutôt en externe ?

• **La réalisation du projet**

Après avoir réalisé les études et évaluations nécessaires à une bonne connaissance de l'existant, et après avoir analysé tous les facteurs clés qui ont conduit à s'assurer de la faisabilité du projet de dématérialisation, il faut s'attacher à bien étudier et surveiller les points importants qui conditionnent le bon déroulement de la réalisation du projet.

Parmi ces points déterminants, on peut citer les exemples suivants :

- le réseau va-t-il supporter les transferts et la circulation des documents électroniques ?
- les postes de travail sont-ils équipés de mémoires vives suffisantes, et d'écrans adaptés à la taille des documents numériques ?
- la base de données et les supports de stockage sont-ils suffisants pour accepter la volumétrie des objets numériques ?

• **La conduite du changement**

L'utilisation d'un matériel de numérisation et d'un éventuel système d'archivage électronique peut amener à changer certaines habitudes de travail, et nécessiter la définition de certaines procédures ou règles pour la numérisation.

Par exemple il est indispensable de répondre aux questions suivantes :

- qui numérise ?
- y a-t-il un contrôle qualité au niveau de la phase de numérisation ?
- les documents papiers sont ils détruits après leur numérisation ? (certains ?, tous ?, selon quels critères ?, etc)

³ Write Once Read Many.

- **Le suivi des archives**

Sur le plan du suivi des archives, en cas d'adoption d'un SAE, il est nécessaire de mettre en place deux types de procédures:

- Des contrôles périodiques de la qualité des archives conservées ;
- Des migrations des archives sur de nouveaux supports de stockage ou de nouveaux formats informatiques. Cet aspect est abordé plus loin dans l'étude avec notamment une présentation des atouts du format PDF/A⁴ ;

- **La dématérialisation des traitements**

L'intérêt et l'avantage d'une dématérialisation des documents papier au sein des centres de documentation peuvent ne pas se limiter à un simple gain de place ou à un gain de performance pour la recherche de ce document. Elle peut aussi permettre dans certains cas de réaliser informatiquement des opérations jusque-là réalisées manuellement. Il convient donc également pour le choix du système et de la solution de recenser les traitements qui peuvent être dématérialisés et d'évaluer le bénéfice qui peut en découler.

⁴ Portable Document Format / Archive.

1.3 Les contraintes liées à la spécificité des documents numériques

La problématique de la gestion des documents numériques est en partie liée à leur spécificité. Un document est le résultat de la combinaison d'une information - connaissance matérialisée selon un code - et d'un support matériel, moyen de sa communication. Dans le système papier, ces deux éléments ne forment qu'un et sont indissociables. En revanche, dans l'environnement numérique, l'information est codée sous la forme de chaînes de caractères en mode binaire, elles-mêmes inscrites sur des supports informatiques.

- **La viabilité**

L'exploitation de fichiers informatiques, quels qu'ils soient, est dépendante d'outils technologiques qui ne sont pas forcément compatibles entre eux et qui évoluent très rapidement. Inévitablement, cela engendre des problèmes d'échange et d'accès aux documents – le cycle d'évolution des applications qui les exploitent est parfois très court -, et l'on trouve aujourd'hui plusieurs dizaines de formats de fichiers en circulation.

De plus, on l'a vu, un point stratégique pour les documents numérisés est également la conservation sur le long terme. Or celle-ci passe non seulement par la garantie d'une intégrité physique des documents, mais surtout par la possibilité de pouvoir y accéder à tout moment dans le futur malgré cette inévitable évolution des outils qui les exploitent. Ainsi, lors de chacune de ces évolutions technologiques majeures, les documentalistes devront donc faire migrer – une opération informatique souvent complexe - leurs bases de documents électroniques, en stockant ces derniers sur de nouveaux formats.

- **La fiabilité**

Une autre particularité des documents numériques concerne la fiabilité. En effet, contrairement aux documents papier, les documents numériques sont très facilement modifiables, sans que ces modifications soient forcément intentionnelles - erreur de manipulation, défaillance matérielle. Or l'intégrité des corpus documentaires est sous la responsabilité des documentalistes et informaticiens, et elle passe par la mise en oeuvre de procédures contraignantes – par exemple la sauvegarde systématique des données (back up).

1.4 La mise en place d'un projet : méthodologie, gestion de projet, freins humains

De manière générale, les documentalistes sont les premiers acteurs des projets de dématérialisation au sein des centres de documentation. En amont d'un tel projet, ils doivent être capables d'analyser les besoins et les contraintes initiales – c'est-à-dire les problèmes techniques et organisationnels -, et ce avec la plus grande attention puisque de la qualité de ce travail préliminaire dépendra la pertinence du cahier des charges qui suivra. Même si ces projets font intervenir d'autres acteurs, comme les informaticiens qui interviendront sur certaines clauses techniques du cahier des charges, les documentalistes restent seuls responsables de la gestion des documents dont ils ont la charge, et ils occupent donc un rôle essentiel au sein des équipes projets. Il leur est par conséquent de plus en plus indispensable de se doter de compétences techniques solides, ainsi que d'une connaissance minimum des règles en termes de gestion de projet.

En effet, tout projet de dématérialisation de dossiers documentaires doit se baser naturellement et logiquement sur une démarche classique de gestion de projet, qui comprend 3 phases principales : une phase de définition du projet, une phase de développement, et une dernière dite d'exploitation et d'évolution.

Détaillons les différentes étapes de chacune de ces phases :

- **Phase de définition :**

- étude de faisabilité et d'opportunité du projet (définition des objectifs, expression des attentes, finalité et périmètre du projet)
- désignation d'un chef de projet qui a la charge de constituer le groupe projet ;
- analyse de l'existant et analyse des besoins (typologie des documents à numériser, typologie des utilisateurs, cartographie des besoins, modélisation des flux documentaires, etc) ;
- analyse des contraintes techniques et organisationnelles à prendre en compte pour mener à bien le projet ;
- la dernière étape de cette phase consiste à cerner les spécifications et les spécificités du projet de GED (sans pour autant arrêter des choix techniques très précis) et à faire une première analyse de l'offre et des solutions existantes sur le marché ;

Lors de cette phase d'étude, destinée à bien définir le projet, il faut souligner encore une fois la nécessité d'une collaboration et coopération étroites entre les documentalistes et le service informatique, qui conditionneront au final la réussite d'un tel projet toujours technique. Les

contraintes techniques au sein de l'environnement informatique d'un centre de documentation étant parfois difficiles à cerner pour les documentalistes, l'expertise et le soutien des informaticiens s'avèrent donc indispensables.

Par conséquent, il est aussi important, dès le démarrage du projet, de définir précisément le maître d'ouvrage et le maître d'œuvre, soit distinguer clairement le représentant des documentalistes et des utilisateurs de celui qui apporte une dimension technique au projet, permet sa conception et surtout garantit son intégration dans le système d'information du centre de documentation.

- **Phase de développement :**

- rédaction du cahier des charges fonctionnel ;
- choix techniques sur le traitement des documents (type de numérisation, résolution, indexation, stockage) ;
- lancement éventuel d'un appel d'offre, et choix des prestataires ;
- réalisation du projet ;

- **Phase d'exploitation :**

- mise en place du système ;
- formation des utilisateurs ;
- exploitation du système ;
- modifications et évolutions du système ;
- validation du système ;
- maintenance du système ;

En plus de la gestion des différentes étapes citées plus haut, le principal défi que doit relever tout bon chef de projet est donc la bonne coordination entre les différents acteurs du projet. Ces derniers, intervenant ponctuellement sur des phases précises, ou au contraire de manière permanente, pourront être par exemple :

- les documentalistes pour : la définition des besoins et de la typologie documentaire, les clés de recherche et d'indexation, l'élaboration des nouvelles procédures de gestion de l'information ;
- le service informatique pour l'acquisition et la maintenance de matériels et logiciels, à intégrer dans le système d'information ;
- un groupe d'utilisateurs pour le test de la maquette ou du prototype ;

- éventuellement un assistant à maîtrise d'ouvrage expert en GED qui accompagnera le chef de projet et le responsable informatique dans des phases critiques du projet sur les aspects techniques, organisationnels et normatifs ;
- le service archives, s'il en existe un, pour organiser l'archivage et la restitution des documents originaux ;
- le service juridique, s'il en existe un, pour les aspects liés au droit d'auteur et au droit de reproduction ;
- le service formation, s'il en existe un, pour la formation au changement et à l'utilisation des nouveaux outils ;

La réussite d'un projet de dématérialisation est non seulement conditionnée par une bonne prise en compte de tous ces aspects cités plus haut - d'ordre organisationnel et technique -, mais également par d'autres d'ordre humain, qui sont tout aussi importants et qui doivent faire l'objet d'une grande attention.

En effet, la numérisation implique la mise en place de nouveaux outils – comme la GED – qui peuvent venir bouleverser les habitudes de travail et le rapport du documentaliste au document. Un projet de ce type peut donc être ressenti et considéré de plusieurs façons selon les documentalistes. Certains voient dans la dématérialisation une possibilité de résoudre les problèmes liés au papier tandis que d'autres y voient au contraire une source potentielle de dégradation de leur condition de travail.

Parmi ces freins et ces craintes qui peuvent être induits par le passage du support papier au support numérique, on peut en citer quelques-uns.

Tout d'abord, les documentalistes ont encore majoritairement une culture papier. Ils restent donc très attachés à ce support et préfèrent souvent la consultation par feuilletage manuel plutôt que la lecture sur ordinateur. Cette attitude devrait s'estomper au fil du temps, et l'apparition de nouvelles générations d'écrans informatiques – plus grande taille, meilleure contraste, meilleure résolution – autorise d'ores et déjà une consultation suffisamment confortable pour l'œil.

Néanmoins, sur ce point, il faut bien avouer que la notion de « zéro papier », dont on a tant parlé, semble aujourd'hui inatteignable et presque utopique. Ainsi, alors que l'avènement de l'informatique dans le monde professionnel a depuis ses débuts été associé à une forte réduction de l'utilisation du papier, et alors que beaucoup d'organismes cherchent et se disent proches du « zéro papier », force est de constater qu'il en est souvent autrement dans la réalité. Des études ont même démontré que c'est parfois le constat inverse qui intervient : chaque personne impliquée dans l'élaboration ou le traitement d'un document informatique imprime très souvent ce document pour les étapes clé de contrôle, lecture ou relecture. Ainsi

un document électronique est encore parfois imprimé un grand nombre de fois, sans qu'il n'y en ait une véritable nécessité ou obligation. Cela met en exergue la faiblesse majeure de la « dématérialisation du papier » car le document numérisé est souvent moins bien appréhendé dans son ensemble mais aussi dans ses détails.

De la même façon, le passage à un système numérique et à une gestion électronique des documents – notamment l'indexation électronique - est en quelque sorte synonyme d'un travail documentaire exclusivement sur ordinateur. Néanmoins, si cela peut encore induire quelques appréhensions chez certains, la majorité des documentalistes des centres de documentation ne sont pas gênés par ce changement puisqu'ils sont déjà familiers de multiples tâches informatiques – bulletinage, consultation de bases de données, etc. En revanche, une crainte récurrente et touchant absolument tous les documentalistes est celle de la dépendance à l'outil informatique, avec le risque de voir son activité ralentie ou arrêtée en cas de problèmes techniques.

En termes d'impact sur la charge de travail des documentalistes des centres de documentation, dans le cadre du passage de dossiers papier à des dossiers numériques, il est évident que la phase de description et d'indexation électronique – qui est une des étapes cruciales de la GED – alourdit forcément le travail des documentalistes. Et l'on peut dire, plus globalement, que la numérisation des dossiers documentaires engendre une réorganisation du travail de ces derniers, mais sans conduire forcément à un allègement – voire même parfois à un certain alourdissement. Et alors même que l'objectif de tels projets est, pour les documentalistes, de pouvoir consacrer plus de temps à d'autres tâches, l'expérience montre que la dématérialisation ne diminue finalement pas le nombre de manipulations nécessaires à l'alimentation des dossiers documentaires. Au contraire, elle peut même demander plus de temps qu'avec l'ancien système papier. D'où l'expression de quelques appréhensions et interrogations de la part de certains documentalistes lors de la mise en place de certains projets.

Pour lutter contre les freins humains qui viennent d'être cités, il existe des moyens d'action classiques mais fondamentaux :

- prendre en compte les besoins et les attentes le plus précisément possible ;
- rédiger un cahier des charges très détaillé ;
- unir et fédérer les documentalistes autour du projet, le plus étroitement possible ;
- prévoir des formations sur les nouveaux outils le plus précocement possible ;
- rassurer en planifiant une implantation très progressive du nouveau système, avec l'établissement d'un calendrier précis ;

Par exemple, il est préconisé que des actions de communication soient menées par le chef de projet, pour attribuer des responsabilités de gestion aux différents intervenants sur chaque tâche ou étape du projet, harmoniser les relations humaines au sein de l'équipe du projet, motiver chaque acteur et participant, formaliser les informations pour une bonne exécution des tâches et informer de leur état d'avancement, etc.

Il est aussi recommandé d'organiser une implantation progressive des nouveaux outils et méthodes, dans le contexte d'un processus de dématérialisation qui se ferait ainsi en douceur. Une fois l'outil de GED choisi, une période devra permettre une intégration pas à pas du nouveau système, ce qui passera notamment par des tests de toutes les fonctions attendues dans l'environnement de travail – procédures de traitement des documents, construction des outils documentaires, définition des profils utilisateurs, gestion des droits d'accès, implantation technologique, etc.

Enfin, on soulignera encore une fois la nécessité d'impliquer un maximum d'acteurs de l'organisme dans le projet : les documentalistes et les utilisateurs, en premier lieu évidemment, mais de façon tout aussi significative la direction et le service informatique du centre. Tout le monde doit « jouer le jeu » pour que le système soit une pleine réussite, afin qu'il puisse faire profiter le centre de documentation de tous ses atouts et de tout son potentiel, actuel et futur.

Deuxième partie

La Gestion Electronique de Documents

2.1 Définitions et principes

Toutes les organisations qui se lancent dans des projets de numérisation de documents, ou « dématérialisation », sont directement touchées par les enjeux d'une gestion électronique de documents.

La GED (Gestion électronique de documents) ou GEIDE (Gestion électronique d'informations et de documents existants) représente un ensemble d'outils et de techniques qui permettent, à partir d'applications informatiques, de dématérialiser, organiser, gérer, stocker et distribuer des informations documentaires sous forme électronique.

On peut distinguer deux époques marquantes de l'histoire de la GED :

- Une première période – allant de l'apparition de la GED au milieu des années 80 jusqu'aux environs de 1995 – où les premières solutions développées correspondaient à des besoins d'archivage électronique. La justification d'une GED était alors d'ordre légal et administratif. Les solutions techniques de GED étaient alors très lourdes, très coûteuses, et proposaient une alternative à la micrographie.
- Une deuxième période, survenant vers 1995, au moment de l'expansion des réseaux et du multimédia. La justification d'une GED est alors devenue économique et stratégique, avec une recherche de gisements de productivité, de facteurs d'innovation et de qualité, à travers la mise en place de processus de travail collaboratif, de capitalisation et d'échanges informationnels.

Au cours des dernières années, l'essor de la GED a été favorisé par l'augmentation des capacités de stockage, le développement des techniques de numérisation et de reconnaissance optique de caractères, ainsi que celui des réseaux de télécommunications à haut débit. Sur ce dernier point, il est évident que la généralisation et la continue optimisation des réseaux Internet et intranet a permis des évolutions technologiques majeures en termes de solutions logicielles. Ainsi, alors qu'au début les éditeurs se sont contentés de mettre en place des passerelles entre leurs solutions et la consultation à travers ces réseaux, aujourd'hui un nombre croissant de logiciels repose à 100% sur une technologie web. Et de nouveaux standards se sont ainsi imposés, comme le XML (eXtensible Markup Language) ou encore le J2EE (Java 2 Enterprise Edition), qui favorisent le traitement de l'information en ligne ou le développement d'applications interconnectées.

La GED est aujourd'hui à vocation universelle, trouvant de multiples applications dans toutes les organisations où les documents abondent. Et, alors qu'à ses débuts ses principes se résumaient à numériser un document et à le caractériser par des champs descriptifs afin de

faciliter son accès et sa recherche, la GED s'est depuis considérablement développée en intégrant notamment des notions fondamentales comme la gestion de contenu (content management) ou la gestion de processus (workflow).

La GED a pour vocation de rendre l'information accessible :

- plus facilement avec les indexations et les moteurs de recherche ;
- plus rapidement grâce l'informatique qui abolit la distance entre l'utilisateur et le lieu où se trouve physiquement l'information ;
- plus sûrement car les accès sont contrôlés et les documents ne risquent pas d'être déclassés par un utilisateur négligent ;
- simultanément par plusieurs utilisateurs ;

Les 5 grandes catégories de GED existantes sont :

- La GED administrative permet de numériser puis de classer les documents administratifs (factures, fiches techniques, formulaires, devis ...).
- La GED bureautique regroupe l'offre de progiciels de travail collaboratif permettant d'échanger des documents, de les lire dans leur format d'origine (Word, Excel, Powerpoint, Outlook ...).
- La GED COLD (Computer Output on Laser Disc) qui permet d'archiver sous une forme électronique les états produits par l'informatique d'un organisme (relevés de compte, factures, etc).
- La GED technique ou GED métier qui concerne la manipulation de documents dont le format et le contenu sont propres à un métier (plans, schémas etc..).
- **La GED documentaire** consiste à indexer un grand nombre de fichiers numériques aux formats les plus divers (texte, image, ..) selon des critères définis par et pour l'organisme. Ce mode de gestion suppose de mettre en place de puissants moteurs de recherche afin que chaque fonction de l'organisme puisse exploiter efficacement les multiples ressources mises à sa disposition. Cette catégorie de GED a pour objectif la diffusion à distance des documents, et recouvre des applications variées : la fourniture électronique de documents, la consultation de fonds documentaires numérisés, les dossiers documentaires (accessibles à distance et de manière partagée), les banques d'images, etc.

Dans le cadre de cette étude, ce sera cette dernière application qui nous concernera directement.

Les fonctions de la GED peuvent se décomposer en trois classes - ou trois pôles - complémentaires et successives : l'acquisition de documents, la production de documents, la consultation de documents.

Ces fonctions comportent des étapes – on parlera donc d'étapes fonctionnelles – qui s'enchaînent selon le processus en boucle suivant – on parlera de chaîne GED :

- acquisition du document au format numérique ;
- formatage ;
- traitement (compression, retouche et correction graphique, reconnaissance optique de caractères) ;
- indexation ;
- stockage ;
- recherche ;
- consultation ;
- diffusion ;

Il est important de préciser que lorsque les GED incluent des fonctionnalités dites « d'archivage », c'est bien souvent au sens restreint du terme, c'est-à-dire au sens de stockage et de gestion des fichiers dans le système. Ce n'est donc pas l'archivage des archivistes car il s'agit le plus souvent de copies dépourvues de valeur probante et qui n'ont donc pas à être gérés comme des archives proprement dites. C'est précisément le cas du centre de documentation des Arts Décoratifs, et ce point sera abordé dans la troisième partie de ce mémoire.

Une distinction doit donc être faite entre un système de GED et un système d'archivage électronique (SAE), qui peuvent présenter des fonctionnalités différentes puisqu'ils poursuivent des buts différents.

Le choix d'un système par un organisme repose sur le statut et la fonction des fichiers concernés : il est important de savoir si ces derniers sont figés en vue de tracer des responsabilités ou de défendre des intérêts, ou au contraire s'ils représentent des objets sans valeur probante et dont le contenu est mis à disposition des équipes et des utilisateurs de l'organisme qui peuvent éventuellement le modifier.

Un tableau tiré du modèle européen MoReq⁵ pour l'archivage électronique, comparant les caractéristiques respectives du « document management » (équivalent de la GED) et celles du « e-records management » (équivalent de l'archivage électronique), est consultable en annexe 9 à la page 188 de ce document.

⁵ Model Requirements for the management of electronic documents and records.

2.2 La mise en place et les enjeux d'une GED

La mise en place d'une GED nécessite tout d'abord de choisir judicieusement les technologies adéquates intégrables et évolutives à court et moyen terme. Ensuite, dans le cadre d'un projet de dématérialisation, dès qu'un projet de GED dépasse les frontières de la simple station personnelle, il comporte des composantes humaines et organisationnelles sur lesquelles la plus grande attention doit être portée. Notamment, l'introduction de la GED dans le système d'information d'un centre de documentation nécessite d'associer étroitement les utilisateurs finaux au projet, pour améliorer les chances de réussite et de rapidité d'intégration des solutions, et donc du retour sur investissement.

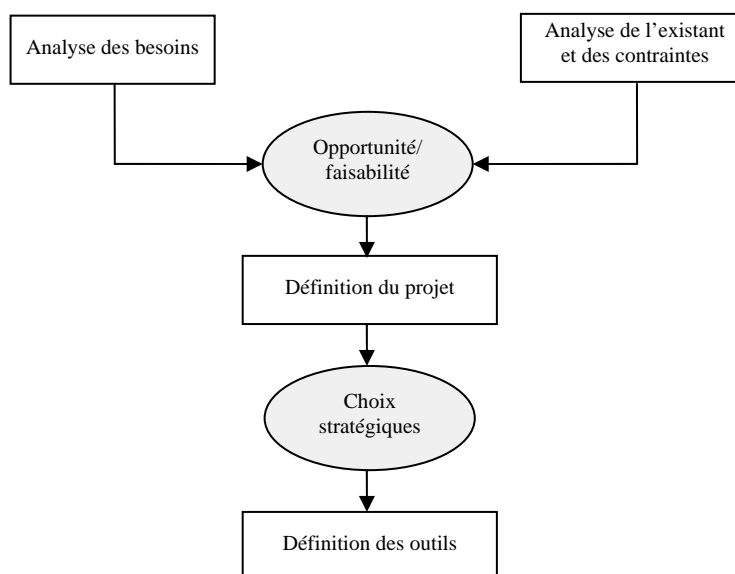
Face aux résistances humaines et culturelles qu'elle suscite parfois, et malgré les modifications des modes de travail qu'elle engendre, la GED – dont les deux facettes sont la dématérialisation et le travail collaboratif - oppose pourtant un solide et fondamental atout qui est celui d'apporter de la souplesse là où le papier est au contraire extrêmement figé : circulation quasi instantanée de l'information, recherche rapide et fiable sur des centaines de milliers de documents à l'aide de mots-clés, sécurisation de l'accès et assurance de la bonne conservation des documents. Et même si la mise en place d'une GED signifie généralement un investissement et un effort conséquents, il est souvent constaté que les utilisateurs de ces nouveaux outils et services auraient maintenant du mal à s'en passer. L'information est présentée plus efficacement, mieux ciblée, et circule plus vite.

Les configurations de la GED peuvent être multiples, orientées en fonction de multiples paramètres :

- configuration physique des options d'acquisition : pour le cas d'un centre de documentation, par exemple, il s'agira plutôt d'une configuration centralisée ;
- nature des documents : des articles papier de périodiques par exemple ;
- législation : en termes de droit de reproduction, par exemple, celle-ci est sensiblement différente pour le paysage papier et pour le paysage électronique ;
- confidentialité et sécurité ;
- usage attendu : modes de recherche, modes d'archivage, etc ;

La mise en place d'une GED doit être menée comme un projet - avec une méthodologie et un mode de pilotage appropriés. Un projet de GED ne se résume pas à une simple acquisition et installation d'un outil, il est donc indispensable de faire précéder la mise en place de l'outil de GED par une étude de faisabilité et d'opportunité – sur le plan technique, organisationnel et économique. L'analyse stratégique doit également être bien conciliée avec l'analyse

technique. La réflexion doit notamment partir d'une observation des pratiques des utilisateurs et des documentalistes, et s'appuyer sur les méthodes de gestion de projet dont le schéma peut être représenté avec la figure ci-dessous.



De manière générale, les risques d'échec d'un projet GED sont :

- un manque d'implication des utilisateurs et de la direction ;
- le choix d'un outil sans études préalables, et en omettant les aspects organisationnels et humains ;
- une mauvaise définition des objectifs en début de projet, se traduisant par des changements de direction au cours des phases de mise en oeuvre ;
- une mauvaise analyse des contraintes internes conduisant au choix de solutions qui ne s'intègrent pas dans le système d'information existant ;
- le choix de documents non représentatifs entraînant des remises en cause des matériels d'acquisition ;
- une mauvaise définition des critères d'indexation et de recherche rendant le système mal adapté aux utilisateurs et par conséquent un manque d'adhésion au système ;
- un mauvais choix technique par manque d'évolutivité et de pérennité des systèmes choisis ;

2.3 Les étapes de la chaîne de traitement du document

Avant de décrire plus en détails le processus de dématérialisation et les différentes étapes de la chaîne de traitement du document – ou chaîne GED –, qui va de l’acquisition numérique à la diffusion et dans lesquelles les documentalistes sont susceptibles d’intervenir en appliquant des processus, essayons de préciser la notion de « numérisation ». Pour cela rappelons la définition qui en est donnée par le Dictionnaire encyclopédique de l’information et de la documentation : « procédé électronique de production de signaux électriques numériques soit à partir d’un document ou d’un objet physique, soit à partir d’un signal électrique analogique. Le fichier numérique permet des traitements informatiques et, notamment, la réplique illimitée et sans perte de qualité indispensable à l’archivage et à la diffusion des documents ».

Dans le cadre de cette étude qui traite de la dématérialisation de dossiers documentaires, l’objet de la numérisation est essentiellement représenté par des documents textuels ou des images.

La mise en place d’un système de GED s’accompagne d’une nécessité de se poser toute une série de questions qui permettront de s’orienter vers des choix technologiques. Certaines d’entre elles ont été regroupées dans un tableau consultable en annexe 1, à la page 161 de ce document. Les réponses à ces questions passent, dans tout projet, par une analyse approfondie de toutes les étapes de la chaîne de GED, qui va de l’acquisition numérique jusqu’à la diffusion du document numérisé.

2.3.1 L’acquisition numérique

Le premier maillon de la chaîne est l’acquisition numérique. Pour que l’information soit gérable par ordinateur, elle doit être disponible sous forme numérique (suite de bits à 0 ou 1).

Il existe trois possibilités d’acquisition numérique - dont une seule concerne notre étude :

- l’acquisition directe de l’information en mode numérique (saisie directe du texte sur traitement de texte, acquisition d’images à partir d’un appareil photo, etc) ;
- la collecte et l’assemblage de documents déjà numériques (importations de documents et opérations de conversions de formats) ;
- la conversion en numérique de documents analogiques ;

C’est ce dernier cas de figure qui nous intéresse pour cette étude, avec pour support analogique le papier et pour technologie le scanner. Le scanner est un périphérique qui est capable de capturer un document en différenciant les zones de ce dernier selon l’intensité

lumineuse qu'elles réfléchissent, et ce grâce à des cellules photoélectriques. La numérisation d'un document est donc rendue possible grâce à ces capteurs ultra sensibles à la lumière rediffusée par les documents.

La composition d'un scanner est la suivante :

- une source lumineuse ;
- un miroir qui réfléchit la lumière renvoyée par le document ;
- des capteurs qui captent le rayonnement lumineux et le transforment en impulsions électriques qui sont envoyées vers un convertisseur analogique/numérique ;
- un logiciel pilote ;

Enumérons maintenant les principales caractéristiques d'un scanner.

• **Les capteurs**

Premier élément très important de la numérisation : la technologie employée pour les capteurs du scanner. Il existe à l'heure actuelle deux types de capteurs : les capteurs dits à technologie CCD⁶, et les autres dits CMOS⁷.

- Les capteurs CCD ou « capteurs à transfert de charges » :

Avec ce système optoélectronique, le document est éclairé par une lampe. L'image virtuelle est alors acheminée par un ensemble de miroirs et d'objectifs. Ce dispositif système optique réduit au final l'image virtuelle à la dimension du capteur CCD.

Ces capteurs permettent une performance extrême (imagerie médicale, astronomie, caméras vidéo), une numérisation plus rapide - haute fréquence de fonctionnement - et une grande netteté. Ils sont recommandés pour le scannage de photos, de livres ou d'illustrations.

- Les capteurs CMOS ou « capteurs d'images par contacts » :

La technologie CMOS repose sur des semi-conducteurs à oxyde de métal complémentaire. Avec ce système optoélectronique, le document est éclairé par un système de LED⁸ qui se trouve au contact même du document. D'où un trajet optique inexistant et une acquisition immédiate.

Ces capteurs sont moins encombrants, moins chers, moins consommateurs en électricité. Ils sont utilisés sur les produits d'entrée de gamme. Ils offrent une qualité d'image moins bonne que pour les CCD, et surtout ils sont plus lents. Ils sont suffisants pour scanner par exemple des schémas ou des cartes.

⁶ Charge-Coupled Device.

⁷ Complementary Metal Oxide Semiconductor.

⁸ Light-Emitting Diode, ou diode électroluminescente.

Actuellement, tous les scanners dédiés à la GED documentaire utilisent des capteurs CCD dernière génération.

- **Le format de document**

Pour numériser des dossiers documentaires, les formats qui doivent être pris en charge sont le plus souvent le A4 et le A3, sachant que la grande majorité des documents contenus dans les dossiers outils au sein des centres de documentation sont généralement au format A4. Or ce dernier est pris en charge par tous les scanners appartenant à la catégorie dite des « scanners de bureau ». Cette catégorie sera d'ailleurs celle retenue pour la grande majorité des projets de GED documentaire, qui ne nécessitent généralement pas des traitements de type industriels ou très pointus. En revanche, le format A3 n'est pas pris en charge par tous les scanners de bureau, et il faudra donc être vigilant sur ce point.

- **La résolution**

Les capteurs divisent l'image en une grille numérique de points élémentaires nommés pixels. La résolution de numérisation est donc une mesure théorique exprimée en points par pouce⁹ (dpi¹⁰), qui définit la finesse du quadrillage et fait directement référence aux performances des capteurs. Pour exemple, la numérisation de photos d'art exige une résolution très élevée, autour de 8000 dpi, afin de conserver le grain.

En revanche, pour appliquer un traitement OCR¹¹ à un texte dans le cadre de la GED, la résolution doit être d'environ 300 dpi, ou un peu plus si les caractères sont vraiment petits.

Une résolution élevée, si elle garantit une restitution conforme à l'original, génère en contrepartie des fichiers qui peuvent être très lourds, ce qui pose des problèmes de stockage et de diffusion. Les documentalistes doivent donc trouver un compromis entre la qualité de l'image et l'espace qu'elle occupe. Ce compromis est à déterminer en fonction de la nature des documents et de leur usage.

Un tableau récapitulant les définitions de numérisation adéquates en fonction des modes numérisation et des types de documents est consultable en annexe 1, à la page 162 de ce document.

- **Le mode couleur**

Les scanners actuels utilisés en GED sont capables de numériser dans les trois modes suivants : noir et blanc, niveau de gris, couleur.

⁹ Le pouce (« inch » en anglais) est une unité de mesure de longueur. 1 pouce = 2,54 cm.

¹⁰ Dot per inch, ou point par pouce.

¹¹ Optical character recognition, ou reconnaissance optique de caractères.

- Le mode noir et blanc est le plus économique sur le plan de l'espace de stockage. Il peut être utilisé pour numériser des documents exclusivement textuels mais n'est pas adapté pour le traitement des images.

- Le mode niveau de gris est moins économique, mais permet un bon rendu des pages numérisés. Son rendu esthétique lui permet même de numériser assez fidèlement des images grâce à ses 256 niveaux de gris.

- Le mode couleur repose sur le même principe que le niveau de gris, mais en superposant cette fois trois numérisations en trois couleurs différentes : rouge, vert, bleu. Il est très gourmand en mémoire de stockage, mais c'est de loin le mode qui donne les meilleurs résultats en termes de rendu des détails, de lisibilité et d'esthétisme.

Pour la numérisation de dossiers documentaires composés de documents textuels et d'images – par exemple un article de revue – ce sont les modes niveau de gris ou couleur qui sont préconisés.

- **La linéarité/ la précision**

C'est la capacité du scanner à restituer des lignes parallèles sans déformation, ni perte de distinction. Cette caractéristique est intrinsèque au scanner, gérée automatiquement, et n'est pas paramétrable par l'utilisateur. Aucune indication de mesure n'est fournie par les constructeurs, et seuls des tests effectués en laboratoire permettent de faire des comparaisons entre les scanners – un peu à la manière des dossiers de la Fnac pour le matériel grand public.

- **La restitution chromatique**

Elle détermine le rendu des couleurs. Même situation que pour la linéarité. Seuls des tests en laboratoire permettent des comparaisons objectives entre les matériels disponibles sur le marché.

- **Les niveaux de gris**

Ils déterminent la capacité du scanner à restituer les contrastes. Idem que pour les deux critères cités ci-dessus.

Les trois critères de qualité qui viennent d'être cités plus haut sont fortement liés à la qualité des capteurs du scanner, mais aussi à l'électronique dont dispose ce dernier. En réalité, tous les scanners sur le marché et qui se destinent à la GED sont dotés des mêmes générations de capteurs, fournis par les mêmes constructeurs. Seule la taille de ces capteurs varie véritablement, en fonction des usages attendus (exemple de la photo d'art cité plus haut).

La précision, la restitution chromatique et le contraste sont trois facteurs qui sont pris en charge par le scanner, sur un plan logiciel, et ceci à chaque numérisation. Ils ne nécessitent aucune compétence particulière de la part de l'opérateur, excepté si celui-ci souhaite un résultat très particulier comme un rendu artistique. Ce qui n'est pas le cas pour des projets de numérisation de dossiers documentaires.

- **La vitesse d'acquisition**

Elle s'exprime en pages par minute (ppm). Elle est directement liée au format du document traité ainsi qu'à la résolution choisie pour la numérisation. Ce critère de vitesse est très important dans le domaine de la dématérialisation, et constitue souvent le critère de choix par excellence dans l'achat d'un scanner. Cette vitesse est dépendante de la mécanique du scanner (vitesse de défilement de la lampe et des capteurs) mais aussi de la mémoire du scanner (qui stocke les fichiers bruts avant compression).

On doit souligner ici que cette vitesse d'acquisition est la principale caractéristique généralement mise en avant par les constructeurs dans leurs fiches commerciales. Et il est établi que, couramment, la vitesse « commerciale » se mesure dans les conditions suivantes : numérisation en A4 standard à 200 dpi en paysage (idem pour les scanners supportant le A3), dans un mode binaire (c'est-à-dire en noir et blanc), et sans aucune autre fonctionnalité.

- **La volumétrie quotidienne du scanner**

Dans le cadre des projets de dématérialisation de dossiers documentaires, il n'est en général pas exigé une volumétrie extrêmement élevée – surtout s'il s'agit uniquement d'une dématérialisation au fil de l'eau et non d'une dématérialisation rétrospective massive. Une valeur comprise entre 500 et 1500 pages/jours est ainsi souvent suffisante.

Néanmoins, il faut tout de même faire preuve de vigilance et bien « décrypter » là encore lors de la lecture des fiches commerciales. En effet, ce critère de volumétrie quotidienne renseigne en fait aussi directement sur la « durabilité » du scanner, et donne indirectement le nombre total de numérisations rapporté à la durée de vie du scanner pour une utilisation normale du scanner et dans le respect du programme de maintenance préconisée par le constructeur. La durabilité est donc un élément très important de la fiche commerciale d'un scanner, puisqu'elle renseigne sur la robustesse des éléments mécaniques et électroniques composant le scanner.

- **La connectique**

Les deux standards utilisés actuellement sont le Firewire et l'USB¹². Le Firewire est l'interface de prédilection pour les projets de dématérialisation, car son débit est particulièrement adapté pour les périphériques de capture. L'USB est quant à lui présent sur tous les ordinateurs récents, et constitue l'interface standard.

Il est préconisé que l'ordinateur pilotant le scanner soit doté d'un port Firewire. Si tel n'est pas le cas, l'achat et l'ajout d'une carte Firewire se fait très facilement et pour un coût extrêmement modique.

- **Le protocole utilisé pour le pilotage du scanner**

Enfin, dernière caractéristique importante d'un scanner, et qui peut intervenir comme critère de choix lors d'un achat : le protocole de contrôle. Il permet d'établir le lien entre le scanner et le système d'exploitation de l'ordinateur, pour assurer la transmission des données au logiciel de traitement.

Actuellement on dénombre l'utilisation de 3 protocoles :

- TWAIN, « Technology Without An Interesting Name » est un protocole standard, ouvert, mis au point en 1992 par un organisme regroupant plusieurs fabricants majeurs.
- ISIS, qui est lui un protocole propriétaire (de la société du même nom) non ouvert, et non gratuit.
- SANE, « Scanner Access Now Easy » est un protocole open source, pour Linux ou Unix, mais aussi pour Mac et Windows. Il permet le fonctionnement de scanners en réseau.

Tous les scanners vendus pour des opérations de dématérialisation sont fournis avec les pilotes TWAIN et ISIS, qui sont adaptés et préconisés pour des projets de GED documentaires classiques.

- **L'adaptation fonctionnelle du numériseur au type de document**

Un tableau résumant l'adaptation fonctionnelle du numériseur au type de document à traiter est consultable en annexe 1, à la page 163 de ce document.

Parmi la liste des scanners décrits dans le tableau, on voit que le type qui correspond le mieux aux besoins d'une numérisation de dossiers documentaires outils - par exemple des dossiers thématiques alimentés par des articles de presse - est un scanner permettant non seulement la numérisation à plat grâce à une vitre d'exposition, mais aussi une numérisation avec défilement grâce à un chargeur.

¹² Universal Serial Bus.

2.3.2 Le formatage

Après la numérisation physique des documents, et avant les opérations d'interprétation, se situe la phase d'enregistrement sous un format de fichier – ou formatage. Ces formats de fichiers se divisent en trois grandes familles : les formats de fichiers textes, les formats de fichiers images et les formats de description de page. Dans cette partie consacrée au formatage, on s'intéressera seulement aux formats de fichiers images, et l'on abordera le cas particulier du format PDF qui est en fait une sorte de format mixte des trois formats de fichiers : texte, image et description de page.

Ce format de fichier - ou mode - « image » consiste en fait en une photographie numérique du document : le codage informatique retranscrit l'image du document, avec le style de caractères, la mise en page et les illustrations. A ce stade, et avant un quelconque traitement de reconnaissance de caractères, il n'y a aucune possibilité d'intervention ou d'interrogation directement sur le texte numérisé.

Le choix d'un format de fichiers peut être plus ou moins critique selon la pérennité que l'on souhaite donner au fonds numérisé. Par exemple, si l'on prend le cas particulier d'un fonds d'archive patrimoniale, le choix du format de fichier est primordial car il engage l'avenir du fonds à long terme. Et de manière plus générale, pour tous les projets, en cas de mauvais choix au départ la nécessité de numériser rétrospectivement le fonds peut être une opération très coûteuse. Il faut donc faire le bon choix au moment de la numérisation, pour ne pas devoir renumériser quelques années plus tard.

Voici les principaux critères qui doivent dicter le choix du format de fichier :

- la garantie de l'intégrité des données ;
- la qualité de compression, qui conditionne la future exploitation de l'image (OCR, impression en grand format, etc) ;
- la rapidité de numérisation ;
- le poids du fichier, qui conditionne le volume de stockage et la vitesse d'affichage du document ;
- la compatibilité avec les logiciels applicatifs ;
- l'anticipation sur la possibilité de réaliser, dans un futur plus ou moins proche, des traitements ou des conversions qui n'existent pas encore, ou dont on a pas l'utilité aujourd'hui ;

Il apparaît donc comme évident que les besoins et les exigences pour la numérisation de simples dossiers documentaires au sein d'un centre de documentation - et qui ne s'inscrirait

pas dans le contexte d'une conservation patrimoniale -, ne seront pas les mêmes que ceux, par exemple, de la Bibliothèque Nationale de France (BNF) dans le cadre du projet Gallica¹³. L'attention et l'importance accordées à la qualité de numérisation et à la pérennité sont naturellement bien supérieures pour un projet de numérisation d'ouvrages anciens que pour la simple mise en ligne de dossiers documentaires outils.

Si on fait le choix de privilégier la qualité de reproduction et la pérennité - ce qu'a fait la BNF par exemple - il est alors généralement recommandé de choisir le format le plus qualitatif, c'est-à-dire le plus lourd - le TIFF¹⁴ étant préféré la plupart du temps. Même si cela doit se faire au prix d'une numérisation plus lente. A partir de là, il est alors possible ensuite, en fonction des besoins, de convertir les fichiers issus de la numérisation dans le format le plus adapté à l'utilisation envisagée. Par exemple, dans le cas d'une mise en ligne sur intranet ou Internet, il est nécessaire de convertir à partir du format de stockage vers le format adéquat (par exemple le PNG¹⁵ ou le GIF¹⁶).

A l'inverse, un centre de documentation aura plutôt tendance et intérêt à numériser directement dans un format adaptée - on verra que le PDF est le plus souvent adopté - à l'utilisation - c'est-à-dire la mise en ligne, et ce sans avoir procédé au préalable à une numérisation au format TIFF. Ceci permettant d'écourter le temps de numérisation, et surtout d'alléger considérablement les volumes de stockage - puisque les documents sont stockés dans un seul format, et qui est de plus moins gourmand en octets.

Il existe en fait deux modes de représentation numérique d'image, ou catégories de formats de fichiers images : le vectoriel et le bitmap. Chacun convient à des utilisations propres, et il est primordial d'avoir une claire compréhension de leurs caractéristiques. Sans entrer dans des détails d'ordre trop technique, présentons donc brièvement les avantages et les inconvénients de ces deux modes.

Le bitmap s'applique en particulier à toutes les images numérisées par un scanner. Il convient non seulement aux images photographiques, mais s'applique également à la représentation de pages de texte. Dans ce cas, bien évidemment, il s'agit d'un document consultable - sans recherche en texte intégral - mais non modifiable - autrement dit non compréhensible par la machine, et non d'un réel document en mode texte - ou mode caractère. Pour passer à un mode texte, il faut utiliser la technologie dite de « reconnaissance

¹³ <http://gallica.bnf.fr>

¹⁴ Tagged Image File Format.

¹⁵ Portable Network Graphics.

¹⁶ Graphic Interchange Format.

optique de caractères » (OCR), qui est un traitement du document et qui sera décrit plus loin dans l'étude.

Le vectoriel propose lui un champ de représentation infiniment plus limité, mais beaucoup plus flexible. Il convient notamment aux dessins au trait, aux graphiques et présentation, ainsi qu'au domaine graphique technique. Mais il n'est pas adapté à la représentation d'images avec un rendu photographique.

Les avantages du bitmap :

- il permet une copie exactement conforme à l'original, et peut représenter absolument toutes les images possibles. Il a donc un caractère plutôt universel ;
- il convient notamment aux photographies ;

Les inconvénients du bitmap :

- il est peu adapté aux textes et tracés très fins ;
- les fichiers obtenus sont très volumineux ;
- il limite énormément la flexibilité du traitement des images, il survient notamment une certaine dégradation de la qualité lors des changements d'échelle ;

Les avantages du vectoriel :

- les tracés de courbes sont très précis ;
- le fichier est léger ;
- les changements d'échelle et les modifications sont faciles ;

Les inconvénients du vectoriel :

- il saisit avec plus de difficulté les images et les photographies marquées par des dégradés importants ;
- l'affichage nécessite un peu plus de puissance de calcul ;

Il existe deux principaux formats de fichier pour le bitmap : le TIFF et le JPEG¹⁷.

Il existe trois principaux formats de fichier pour le vectoriel : le PostScript, le PDF, le SVG¹⁸.

Le mode vectoriel consiste à décrire une image ou une police de caractères selon un ensemble de formes géométriques. Et les traits définissant ces formes sont décrits par les coordonnées de leurs points terminaux.

Des tableaux récapitulant les caractéristiques et les recommandations pour chacun des formats de fichier sont consultables en annexe 1, aux pages 164 et 165 de ce document.

¹⁷ Joint Photographic Experts Group.

¹⁸ Scalable Vector Graphics.

On constate donc que le format TIFF est très souvent utilisé au titre d'une conservation patrimoniale, et qu'il assure l'archivage du fichier de référence. Il donne une copie conforme à l'original et garde toutes les informations, en vue d'un traitement ultérieur ou d'une future conversion de fichier. Pour la dématérialisation de dossiers documentaires outils, ce genre de format n'est pas indispensable et peu approprié.

En revanche, il apparaît clairement que le format PDF est nativement destiné à l'exploitation du document numérique, et que c'est sous ce format que les documents sont généralement consultés en ligne sur les postes informatiques. Le PDF est donc très majoritairement préconisé pour les projets de dématérialisation de dossiers documentaires.

Décrivons maintenant plus en détails les avantages et les atouts de ce format devenu récemment une norme sous sa version du PDF/A.

Le Portable Document Format (PDF)

- Le format PDF réunit en un seul fichier tous les fichiers composants de la mise en page d'un document électronique (texte, images, polices, objets graphiques, informations de mise en forme). Ainsi, les fichiers numérisés sont fidèles aux documents originaux, quelles que soient l'application et la plate-forme utilisées pour le créer. Et ils s'afficheront de la même manière sur tout PC, et ce quel que soit le système d'exploitation utilisé. Cette portabilité en fait le format idéal pour l'archivage numérique.
- Le PDF est dit « multiformat » puisqu'il peut aujourd'hui intégrer une couche de texte, des images, mais également des vidéos, du son et des formulaires.
- Il peut être enrichi de pièces jointes et de commentaires.
- Il organise le document en pages physiques au format papier A4, ce qui permet notamment d'imprimer directement le document.
- Il est constitué non seulement du corps du document mais aussi d'options de visualisation et de navigation, comme les signets par exemple, qui permettent de disposer de l'arborescence du document et ainsi d'accéder à ce dernier de manière hiérarchique et non linéaire.
- Il autorise une recherche en texte intégral, fonctionnalité inexistante pour le TIFF ou le JPEG.
- Il permet une forme de navigation via des liens hypertextes et des boutons de navigation.
- Il est possible de pré programmer un niveau de zoom pour faciliter la lecture à l'écran.
- Il présente également l'avantage de pouvoir sécuriser les documents et de préserver ainsi leur intégrité. Avec notamment des interdictions de saisie d'écran, de modification,

d'impression, mais également une authentification des transactions documentaires grâce à la signature électronique.

- Tout document électronique est convertible en PDF.

En 2005 est né le PDF/A ou PDF Archive, basé sur le format PDF v1.4. Et c'est la norme ISO¹⁹ 19005 qui a défini les spécifications de ce format - sous ensemble du PDF centré sur l'archivage - de même que celle de l'outil de visualisation qui lui est associé. Conçu à l'origine par la société Acrobat, ce format est donc désormais ouvert – ainsi les fichiers des composantes sont en formats ouverts (XML, JPEG, etc) – et normalisé ISO.

Cette norme optimise l'indépendance matérielle et logicielle, et permet de s'assurer que tous les documents électroniques au format PDF/A pourront toujours être reconstitués. Cette normalisation garantit aussi la sécurité et la fiabilité de la diffusion et des échanges de documents électroniques.

En 2008, l'ISO a normalisé le PDF v1.7 comme référence ISO 32000.

Le PDF/A étant particulièrement adapté à la GED documentaire, tous les packages actuels (scanner + logiciel) proposent et gèrent ce format.

2.3.3 Le traitement

Dans la chaîne GED, après le formatage vient la phase dite de « traitement » des données. Cette phase comporte deux opérations principales qui sont la compression et la reconnaissance de caractères.

• La compression

La compression permet, en réduisant la taille des documents, d'en optimiser le stockage et la diffusion. Dans cette étude nous ne traiterons pas cette opération en détails.

Mode bitmap

Voici les principales méthodes de compression qui peuvent être implémentées dans le format TIFF :

- Le Run Length Compression (RLC) : C'est une méthode élémentaire qui consiste à remplacer une série de points répétés de même ton par la valeur du ton et son occurrence, et qui utilise donc le principe du codage de la répétition.
- La compression CCIT groupe IV pour les images en noir et blanc.
- La compression LZW qui fournit des ratios de compression moyens de 3.

¹⁹ International Organization for Standardization.

Les algorithmes qui viennent d'être cités sont dits « réversibles », c'est-à-dire qu'ils n'engendrent aucune perte de données.

Le format JPEG n'est pas qu'un format de fichier images, c'est en fait un puissant algorithme de compression dédié à la compression d'images noir et blanc ou couleurs de type photographique. C'est une méthode de compression dite « irréversible », c'est-à-dire entraînant une perte de données, qui permet d'obtenir un taux de compression de 30 :1 sans provoquer de dégradation visible.

Mode vectoriel

A l'origine, le format PDF est fondé sur le format PostScript, mais il représente une nette évolution par rapport à ce dernier notamment parce qu'il utilise des algorithmes de compression de données permettant d'obtenir des fichiers infiniment moins lourds.

• Retouche et correction graphique

Le fichier numérique obtenu après dématérialisation d'un document peut parfois révéler des imperfections susceptibles de nuire à son traitement ou à sa consultation. Voici quelques exemples d'imperfections qui peuvent être réparées ou compensées par un traitement numérique.

- Un mauvais contraste : Le mauvais contraste d'un document textuel peut considérablement diminuer l'efficacité et la fiabilité du traitement OCR.
- La présence de tâches : Certains documents peuvent présenter des tâches - tout en ayant par ailleurs un contraste satisfaisant – dues par exemple au vieillissement du papier. Dans ce cas, ce peut être à la fois le traitement OCR et le confort des lecteurs qui s'en trouvent affectés.
- Des lignes de textes désalignées : certains documents peuvent présenter des lignes de textes désalignées, dues par exemple à un manque de soin apporté lors d'une photocopie. Si cette inclinaison, souvent relativement légère, est assez bien tolérée par l'OCR (dont l'algorithme est plus sensible au contraste qu'à l'alignement), elle est en revanche beaucoup moins bien acceptée par les lecteurs.
- La présence d'un fond de page coloré.

Des traitements numériques adéquats existent et peuvent être appliqués, de façon manuelle ou automatique, pendant ou après la numérisation physique. Tout ceci dépend des options disponibles dans le module d'acquisition du logiciel de GED utilisé.

Ces traitements sont d'autant plus importants et ne doivent pas être négligés lorsqu'il est prévu de détruire le fonds papier après la dématérialisation.

En général, la suppression du fond de page coloré, l'amélioration du contraste et le « désalignement » - redressement des textes et des images inclinés - peuvent être pris en charge automatiquement par les applications de GED, en donnant de bons résultats dans la majorité des cas.

En revanche, la suppression des tâches, ou encore une correction colorimétrique assez fine, sont des traitements qui peuvent nécessiter une intervention et un ajustement manuels, au cas par cas, de la part du documentaliste. Ces corrections exigent en effet le plus souvent une certaine subtilité à laquelle ne peut prétendre la machine et son procédé automatique.

Il est assez difficile de se prononcer sur l'allongement potentiel du temps de numérisation d'un document suite à l'application des traitements cités plus haut. Mais il est certain que le niveau de qualité qui sera exigé, ainsi que la proportion de documents exigeant des traitements pour amélioration – parfois difficilement quantifiable lorsque le fonds est très hétérogène - auront au final un impact sur la durée totale de la dématérialisation.

- **La reconnaissance de caractères (OCR et ICR)**

Rappel des définitions de l'OCR et de l'ICR :

L'**OCR** (Optical Character Recognition, ou reconnaissance optique de caractères) permet de reconnaître une suite de caractères sur un document, par reconnaissance de forme - autrement dit par comparaison de la forme lue avec une base de données.

L'**ICR** (Intelligent Character Recognition, ou reconnaissance intelligente de caractères), est un système d'OCR avancé intégrant des technologies d'intelligence artificielle. Un logiciel d'ICR peut compléter sa base de connaissance au fur et à mesure de la reconnaissance et donc étendre sa capacité de reconnaissance. Ces techniques sont utilisées notamment pour la reconnaissance de caractères manuscrits.

Grâce à ces technologies, il est donc possible de convertir l'image d'un document textuel (texte scanné par exemple) en un texte ASCII²⁰. Les avantages de ce processus sont nombreux et importants.

Tout d'abord, il permet d'intervenir directement sur le texte, autrement dit d'exploiter le contenu des documents textuels – qui viennent d'être numérisés ou photographiés par exemple. L'information peut être ainsi retraitée dans un outil bureautique – de type traitement de texte par exemple -, intégrée dans une base de données – dans un système de GED par exemple -, faire l'objet d'une édition électronique, etc.

²⁰ American Standard Code for Information Interchange.

De plus, les traitements OCR et ICR présentent un autre atout majeur, ils donnent la possibilité d'indexer automatiquement – ils ne nécessitent aucune intervention humaine - l'information contenue dans le texte – indexation en texte intégral par exemple.

Autre avantage, ces processus permettent de réduire considérablement le poids des fichiers numériques, qui deviennent ainsi plus facilement stockable et diffusable. Par exemple, une page de texte au format A4 et scannée à 300 dpi pèse en général quelques centaines de Ko (kilooctets), mais le fichier ASCII qui est produit après le traitement OCR ne pèse plus lui qu'une dizaine de Ko.

Enfin, globalement, la reconnaissance optique de caractères optimise la visualisation à l'écran et la qualité d'impression, en termes de rapidité mais aussi de qualité.

Les premiers algorithmes de reconnaissance de caractères ont été développés dès les années 50, mais ils se sont considérablement développés et perfectionnés depuis. Cette technologie est aujourd'hui disponible à un prix abordable et est intégrée dans toutes les solutions de GED du marché.

Actuellement, les applications OCR et ICR combinent deux formes d'algorithmes de reconnaissance, qui sont les plus courants : la comparaison de matrice, l'extraction et la reconnaissance de formes. Le premier compare chaque caractère du document numérisé à des bases de caractères connus, tandis que le second identifie chaque caractère par ses formes et la façon dont elles sont disposées.

Pour obtenir des résultats satisfaisants avec cette technologie, autrement dit pour bénéficier d'un taux de reconnaissance de caractères élevé, il est nécessaire que la résolution de l'image du document textuel soit, au minimum, de 200 dpi – la résolution la plus souvent employée étant 300 dpi. Et pour des textes très petits, une résolution de 300 dpi est exigée.

Lorsque des conditions satisfaisantes de traitement ont été réunies – résolution adéquate, bonne qualité d'impression, de contraste, typographie standard, etc -, le degré de fiabilité de la reconnaissance est généralement très élevé, avec un taux de reconnaissance pouvant aller jusqu'à 99%.

En revanche, ce taux chute rapidement lorsque les documents comportent des écritures manuscrites, ou des typographies très particulières – des caractères gothiques par exemple.

Et d'une manière générale, la reconnaissance de caractères reste très dépendant de la qualité des documents à traiter, et particulièrement des éléments suivants :

- un mauvais contraste ;
- un mauvais alignement des lignes de texte ;

- des caractères tordus ou qui se touchent (cas des écritures manuscrites) ;
- l'existence d'un fond de page coloré, ou de tâches ;

Comme il a été expliqué plus haut, des corrections et des retouches – parfois prises en charge automatiquement par le module d'acquisition et le scanner – peuvent rectifier ces défauts, afin d'élever la fiabilité de la reconnaissance.

Certains logiciels, en plus de la reconnaissance optique de caractères, sont également capables de garder la mise en forme originale du document : distinction entre les zones de « texte » et les zones « image » par exemple, conservation des couleurs exactes du document et de la typographie, etc.

Tout document textuel papier comporte en effet :

- une structure logique (un titre, des parties, etc) ;
- une structure physique (mise en page, typographie, etc) ;
- une présentation (fond de page coloré, illustration avec des images, etc) ;

Ces éléments figurent naturellement dans le mode image, mais pour les conserver après le traitement de reconnaissance optique de caractères, le logiciel doit être capable de les identifier et de les ajouter au texte qui vient d'être reconnu et extrait du fichier image. Le résultat final sera enregistré dans un format particulier dit de « description de page ». Le plus connu de ces formats, et actuellement préconisé dans la majorité des projets de dématérialisation de dossiers documentaires, est le PDF.

Enfin, concernant les innovations les plus attendues en matière de reconnaissance de caractères dans les années à venir, il apparaît qu'un des grands enjeux technologiques sera la reconnaissance optique de l'écriture manuscrite, notamment avec de gros progrès potentiels en termes d'intelligence artificielle (capacité d'apprentissage, de raisonnement, d'auto-amélioration et d'auto-correction).

2.3.4 L'indexation

L'indexation se définit comme un processus destiné à représenter, au moyen des termes ou indices d'un langage documentaire ou au moyen des éléments d'un langage libre, les notions caractéristiques du contenu d'un document - ressource, collection - ou d'une question, en vue d'en faciliter la recherche, après les avoir identifiés par l'analyse.

Mais l'indexation est aussi le cœur de la GED puisque c'est l'opération qui consiste à décrire et caractériser le document afin de permettre une exploitation sans nécessairement recourir à la consultation du document lui-même.

Elle doit être une représentation fidèle et la plus exhaustive possible du document et de son contenu, afin de permettre une recherche facile et pertinente. C'est une opération complexe, qui implique souvent des investissements matériels et humains importants.

La difficulté de cette opération réside donc dans la nécessité de produire une représentation formalisée et réduite d'un document et de son contenu, tout en retenant l'ensemble des éléments essentiels de ce dernier. La qualité et la pertinence de l'indexation deviennent ainsi absolument essentielles et cruciales lorsqu'il s'agit d'exploiter un système de GED comportant plusieurs dizaines ou centaines de milliers de documents.

Dans un système de GED, l'indexation du contenu d'un document peut être manuelle ou automatique.

• **L'indexation manuelle**

La méthode dite d'indexation manuelle consiste à créer une fiche descriptive dans l'application GED qui gère les dossiers documentaires, ce qui passe par le remplissage d'une notice associée au document. C'est donc une opération totalement prise en charge par les documentalistes, et qui fait appel aux langages d'indexation – ou langages combinatoires.

Ces derniers sont répartis en deux catégories :

- les langages libres, matérialisés sous forme de mots-clés et de descripteurs libres ;
- les langages contrôlés matérialisés sous forme de listes d'autorité et de thésaurus ;

Avec ce type d'indexation, forcément empreinte d'une certaine subjectivité, le rôle de la machine se limite à indiquer la validité des descripteurs entrés dans la GED en les comparant aux listes d'autorité ou aux thésaurus. L'indexation manuelle sur des documents textuels aux contenus intellectuels complexes est toujours plus pertinente et plus fine que l'indexation automatique, et permet ainsi une recherche plus fiable. Elle est donc plus intéressante pour les documentalistes. En revanche, elle est beaucoup plus coûteuse en temps puisqu'elle mobilise beaucoup plus ces documentalistes.

• **L'indexation automatique**

L'indexation automatique s'est imposée avec l'arrivée de la technologie de reconnaissance de caractères (OCR et ICR). Elle permet d'indexer tous les mots du document. La recherche n'est plus alors limitée aux descripteurs mais porte sur l'intégralité du texte : on parle de recherche en texte intégral (ou « full text » en anglais).

Le principal avantage de l'indexation automatique, qui est totalement prise en charge par l'ordinateur, est qu'elle est bien plus rapide à réaliser, et par conséquent bien moins coûteuse.

Son principal inconvénient est qu'elle est moins pertinente et fiable – en termes de corrélation entre la requête effectuée et les documents obtenus après identification – lorsqu'elle porte sur des documents complexes, et qu'elle apporte généralement du bruit. En effet, l'intelligence de la machine est encore trop limitée lorsqu'elle se confronte à des problèmes de sens et à des notions équivoques.

En revanche, l'indexation automatique convient et se montre relativement efficace lorsqu'elle porte sur des documents structurés et ne comportant pas de termes ambigus.

Ainsi, il existe deux types principaux de traitements automatisés, qui, lorsqu'ils sont utilisés, apportent des gains de productivité absolument énormes : la reconnaissance automatique de documents (RAD), et la lecture automatique de documents (LAD).

La **LAD** est en fait une variante de l'OCR qui se limite au traitement des formulaires dont le cadre géométrique bien défini permet de restituer une information structurée au lieu d'un texte brut comme c'est le cas pour l'océrisation standard. La LAD permet ainsi de réaliser une saisie automatique qui peut servir pour l'indexation des documents. Par exemple, dans des questionnaires papier, ils sont capables de repérer où sont les croix, de lire des chiffres ou des lettres manuscrites dans des cases, d'effectuer un contrôle et/ou de remplir une base de données.

La **RAD** est une technique plus récente que la LAD, et permet quant à elle l'identification automatique des documents à structure fixe ou variable. Elle consiste à distinguer un type de document d'un autre à partir d'une image du document, en la comparant à une base de connaissance des formes de documents (par exemple une base de formulaires types). Un logiciel de GED doté de cette technologie pourra ainsi classer les documents électroniques selon leur type et appliquer le traitement LAD prévu pour ce type de document.

C'est donc bien le couplage d'un module de RAD et d'un module de LAD qui assure une indexation automatique différentielle des documents selon leur type.

Ces traitements connaissent actuellement un succès et un développement toujours croissants, mais ne peuvent s'appliquer qu'à des documents fortement structurés et plutôt figés - de type formulaires, feuilles de paie, factures, questionnaires, corpus administratifs, etc.

Par exemple, si le logiciel de reconnaissance automatique détecte :

- un courrier manuscrit, il applique une reconnaissance de texte et de caractères ;

- une enquête, il lit les cases à cocher et le texte pré-casé ;
- un RIB, il lit le numéro de compte ;
- une facture, il réalise notamment des contrôles sur le numéro de facture et les montants ;

Les traitements sur ces documents structurés peuvent se répartir en trois phases successives :

- il faut tout d'abord déterminer avec précision la structure du document, afin d'obtenir des métadonnées les plus structurées possibles ;
- il s'agit ensuite, avec ces métadonnées, de produire des contenus au format ASCII qui seront indexables par un moteur de recherche ;
- enfin, on relie ces métadonnées descriptives au fichier capturé ;

Des indexations complémentaires

Même si un des objectifs principaux de la dématérialisation est un gain de temps pour les documentalistes, il est impossible de supprimer la tâche d'indexation manuelle des documents via le logiciel de GED. En effet, la recherche en texte intégral ne peut suffire car elle se révèle dans certains cas peu précise et pas assez pertinente, en générant notamment du bruit. De plus, l'indexation automatique n'est pas exploitable pour beaucoup de documents – notamment ceux non ou pas assez structurés. La documentaliste doit donc assumer ce rôle d'ajout d'information et de classement du document, effectués lors de la production de ce dernier et de son intégration dans la base de données.

Mais d'une manière générale, on peut néanmoins dire que, à l'inverse de l'indexation manuelle qui intéresse et concerne directement les utilisateurs spécialistes comme les documentalistes, plus le public est hétérogène et occasionnel et plus l'indexation automatique – avec la recherche intégrale – se révèle être adaptée.

Dans les centres de documentation, la diversification des pratiques documentaires actuelles et l'accroissement continu des corpus documentaires font que les modes d'indexation manuelle et automatique se complètent, c'est pourquoi ils sont souvent utilisés de manière combinée.

Sur le plan du rapport qualité/coûts, « aux dires des concepteurs de systèmes d'indexation automatique, l'indexation manuelle, malgré sa subjectivité, reste la meilleure en termes de pertinence. Meilleure, mais aussi plus chère. Et ingérable pour les systèmes traitant de très gros volumes d'information »²¹.

²¹ TOSCA CONSULTANTS, LENART, Michèle (étude réalisée par). La gestion documentaire : évolutions fonctionnelles et description de dix logiciels. Paris, ADBS Editions, 2004.

2.3.5 Le stockage

En termes de stockage informatique, la plupart du temps, une solution de GED doit utiliser et intégrer plusieurs types de support, afin de pouvoir répondre à plusieurs problématiques. Un système de GED doit être capable de concilier une vitesse rapide d'accès et de consultation aux documents avec une bonne condition de conservation et une restitution fidèle de ces derniers. De plus, la sauvegarde et l'archivage des documents au sein du système doivent se concevoir comme deux travaux complémentaires et non antinomiques.

Les différents critères de choix pour les supports de stockage sont donc :

- la capacité de stockage ;
- le temps d'accès aux données ;
- les standards pour la diffusion ;
- la pérennité du contenu ;
- la réinscriptibilité ou non-réinscriptibilité ;
- la sécurité d'accès ;
- le coût ;

Pour satisfaire à tous ces critères, et répondre à tous les impératifs, les solutions de GED combinent presque toujours les supports magnétiques – pour la consultation et la sauvegarde des informations - et les supports optiques – essentiellement pour l'archivage électronique. Un tableau comparatif des différents supports de stockage est consultable en annexe 2, à la page 167 de ce document.

• Les supports magnétiques

Actuellement, les supports magnétiques capables de stocker des corpus numériques très lourds sont les disques magnétiques et les bandes magnétiques.

Le principe de l'écriture sur un support magnétique consiste à orienter une couche d'oxyde métallique présente sur une surface de mylar et à l'orienter par un courant électrique dans un sens – codage du 0 –ou dans l'autre – codage du 1.

Ces supports magnétiques ont des vocations et des rôles bien distincts au sein d'une GED.

- Les disques sont destinés à assurer la consultation on-line d'une base de données documentaire, étant donné leur vitesse de transmission très rapide et la grande taille de leur espace mémoire.
- Les bandes ont un tout autre usage que celui des disques durs, puisque, comparées à ces derniers, elles sont relativement lentes et avec des capacités de stockage limitées. Elles sont

donc utilisées essentiellement pour la sauvegarde systématique des données à moyen terme, pour des opérations dites de « back up ».

Les disques magnétiques sont rigides mécaniquement et conditionnés dans des emballages étanches – du fait de leur sensibilité aux poussières – comprenant le support d'enregistrement et les dispositifs d'enregistrement et de lecture. Ces disques sont des dispositifs fixes, intégrés dans les appareils qui les utilisent. Communément dénommés « disques durs », ils sont aujourd'hui connus de tous puisqu'intégrés dans tous les ordinateurs, et constituent le moyen le plus simple et le plus efficace de sauvegarder de l'information. Ils offrent souplesse et rapidité, et les capacités de stockage sont de plus en plus conséquentes grâce à l'évolution des technologies. Mais avec un disque dur, la suppression d'un fichier peut prendre une forme définitive, ce qui est contraire à une logique d'archivage. Reproche qui concerne également les bandes magnétiques – même si une opération de suppression demeure beaucoup plus simple sur un disque dur.

Dans le cadre des processus de consultation on-line ou de sauvegarde des informations – et non d'archivage -, et dans une perspective de durabilité, on pourra faire appel à la mise en œuvre de la technologie RAID²² - signifiant en français Ensemble redondant de disques indépendant - pour assurer la pérennité du stockage et garantir l'accès aux informations même en cas de panne d'un disque dur. Le RAID permet en fait de constituer une unité de stockage à partir de plusieurs disques durs, et cette unité, ainsi créée, possède une grande tolérance aux pannes. De plus, la répartition des données sur plusieurs disques durs permet d'augmenter la sécurité et de fiabiliser les services associés. Le mode Raid 5 est notamment à considérer avec un intérêt tout particulier.

Les bandes magnétiques sont conditionnées dans des cartouches amovibles, du fait de la souplesse de leur support. Un appareil d'enregistrement et de lecture comporte un dispositif de chargement de la bande pour faire défiler le support devant les têtes d'enregistrement et/ou de lecture. Le choix d'un format de bande se fait en fonction de la capacité, mais également du taux de transfert des informations. Les divers formats courants de bandes magnétiques ont été regroupés dans un tableau consultable en annexe 2, à la page 167 de ce document. La durée de vie des bandes est en général limitée à 100000 heures, ce qui impose une échéance à l'administrateur informatique. Il est indispensable de procéder à un contrôle régulier de la lisibilité et de la recopie des informations contenues sur les cartouches de bande magnétique.

²² Redundant array of independent/inexpensive disks.

- **Les supports optiques**

C'est une famille de supports basée sur la technologie des diodes lasers et apparue à la fin des années 80. Le principe de l'écriture numérique sur un disque optique consiste à modifier microscopiquement un état de surface par un faisceau laser ou un champ magnétique en créant un état binaire par des brûlures ou en formant des bulles. Le principe de la lecture consiste à identifier les changements d'état de surface par la mesure de la variation d'intensité de la lumière réfléchie.

Ces supports sont adaptés pour le stockage off-line des données numériques. Ils se divisent en deux catégories : les disques WORM inscriptibles une seule fois, et les disques WMRA²³ qui sont réinscriptibles.

Les disques WORM sont particulièrement adaptés pour l'archivage des documents numériques sur le long terme, car ils sont inscriptibles une seule fois et garantissent l'intégrité des données. Leur principe est le suivant : la chaleur qui se dégage du laser utilisé pour graver les fichiers sur les disques optiques déforme en effet irrémédiablement leur surface de polyuréthane et interdit par conséquent toute réinscription. Parmi ces supports, on trouve le CD-ROM, le CD-R (ou CD-WORM), le DVD, le DON²⁴ WORM. Ce dernier est particulièrement recommandé pour les systèmes de GED - notamment grâce à sa capacité de stockage supérieure-, et constitue la technologie phare de l'archivage. Les fabricants de DON WORM garantissent leurs disques 30 ans, et le format le plus courant est actuellement le disque optique de 35cm offrant jusqu'à 25 Go par média. Mais le DON est aujourd'hui sérieusement concurrencé par le DVD qui est moins onéreux. Si le premier est encore et toujours le support d'archivage le plus usité, le second est en passe de lui succéder. On peut également, d'ores et déjà, évoquer la solution que représente le disque Blu-Ray – utilisant un laser bleu à plus haute fréquence -, dont l'arrivée se fait très progressivement. C'est une technologie qui s'imposera très probablement dans quelques années puisqu'il offre déjà une capacité de 25 Go pour le disque simple couche, et 50 Go pour le disque double couche.

- **Les librairies (ou Juke Box)**

Afin d'augmenter les capacités directement accessibles en ligne mais surtout afin de simplifier les manipulations, il existe pour les bandes magnétiques et aussi pour les disques optiques des librairies (ou Juke Box) pouvant contenir une multitude de cartouches de bandes magnétiques ou de disque optiques, accessibles via des systèmes robotisés. Ces librairies autorisent actuellement des capacités de stockage extrêmement importantes, de

²³ Write Many Read Always.

²⁴ Disque Optique Numérique.

l'ordre de plusieurs centaines de téraoctets. Les temps de changement de support varient de 5 à 15 secondes.

2.3.6 La recherche

La recherche dépend directement des possibilités d'indexation des documents offertes par le système de GED. Et, naturellement, la pertinence des résultats de la recherche dans le système de GED est aussi directement dépendante de la qualité, de la justesse et de la précision fournies par la documentaliste lors de la phase d'indexation des documents numérisés.

On a vu que les documents peuvent être identifiés par des mots-clés – qui caractérisent un document de manière plus ou moins contrôlée – et des descripteurs libres.

La recherche pourra être aidée grâce à l'accès à des listes d'autorités et des thésaurus – regroupant l'ensemble des mots-clés conseillés pour l'indexation. Ces derniers sont particulièrement intéressants puisqu'ils permettent d'homogénéiser le vocabulaire utilisé pour l'indexation.

La recherche pourra aussi se faire à l'aide d'un plan de classement. Ce dernier permet de ranger un document selon des thèmes et des sous-thèmes - il est parfois possible d'intégrer plusieurs plans de classement. Certains outils de GED offrent la possibilité de réaliser la recherche par feuilletage du plan de classement.

Il peut exister plusieurs modes de recherche dans un système de GED.

La recherche en texte intégral se fait sur l'ensemble des mots contenus dans les documents, en excluant les mots vides de sens tels que les prépositions, les articles, etc.

La recherche en langage naturel repose sur une analyse linguistique de la requête de l'utilisateur et conduit à l'élaboration d'une requête interne au système à partir de laquelle la recherche est réalisée.

La recherche par logique floue permet de formaliser l'usage de termes vagues ou imprécis, dans le but de les rendre compréhensibles par le système. Elle permet entre autres de s'affranchir des erreurs d'indexation.

La recherche multi-critères, autrement dit une interrogation sur tous les champs de la notice du document avec la possibilité de combiner les critères à l'aide d'opérateurs booléens et/ou d'opérateurs de comparaison.

Les opérateurs booléens sont des fonctions permettant de lier différents mots ou groupe de mots caractérisant ou non un document : ET, OU et SAUF en règle générale.

Les opérateurs de comparaison sont des opérateurs arithmétiques : <, >, =. Ils sont utiles par exemple dans le cas où l'on souhaite restreindre la recherche par dates, si celles-ci caractérisent un document.

Des outils de GED offrent aussi des possibilités supplémentaires pour la recherche, comme les fonctions de proximité et la troncature.

Les fonctions de proximité permettent de rechercher deux mots adjacents ou se trouvant à une certaine distance l'un de l'autre – par exemple 2 mots espacés de n caractères ou situés dans la même phrase, ou encore dans le même paragraphe.

La troncature permet de caractériser un ensemble de mots possédant des lettres communes. Elle est souvent formalisée par une étoile. Ainsi, la troncature à droite va par exemple caractériser un ensemble de mots commençant par les mêmes lettres.

Enfin, trois autres options de recherche présentent également un réel intérêt lorsqu'elles sont disponibles dans les logiciels de GED : la présence d'un historique de recherche, la possibilité de croiser des requêtes et la veille.

L'historique de recherche permet de voir la démarche de recherche adoptée, et de revenir le cas échéant à une étape précédente.

Le croisement de requêtes permet de réaliser des requêtes plus complexes et donc forcément plus restrictives.

La veille passe tout d'abord par l'enregistrement d'une requête. Cette dernière sera ensuite lancée automatiquement et de manière planifiée. Les résultats pourront alors être soit visualisés par la personne intéressée de manière active - autrement dit celle-ci ira consulter les résultats lorsqu'elle en ressentira le besoin, fonction « pull » - soit envoyés de manière automatique par mail à chaque lancement de la requête – fonction « push ».

2.3.7 La consultation

La consultation et la modification d'un document dans une application de GED se fait à partir d'un écran informatique, et à l'aide d'un programme dit de visualisation (ou « visualiseur »), qui comporte en général un certain nombre d'options dont celle bien connue du zoom - pour agrandir ou réduire l'affichage du document. Au sein des centres de documentation, la GED doit s'intégrer dans l'environnement matériel bureautique existant, tout en induisant une intensification de l'utilisation de l'affichage écran. Par conséquent, le moniteur devient un élément très important dans la configuration, et les centres de documentation devront y porter une attention particulière dans le cas d'une éventuelle

acquisition d'un nouveau matériel informatique. Les critères de choix pour un moniteur informatique dans le cadre d'une utilisation de la GED sont la résolution, la taille, la fréquence de rafraîchissement et le nombre de couleurs.

La résolution indique le nombre de points (ou pixels) affichés par pouce. Actuellement, les moniteurs informatiques sur le marché sont presque tous de type LCD (écrans plats) et proposent des résolutions autour de 96 dpi. Les écrans un peu plus anciens offrent une résolution souvent inférieure, de l'ordre de 72 dpi. Les moniteurs actuels sont très performants dans l'affichage de documents textuels enregistrés dans un format de fichier image de type vectoriel (cf. la partie 2.3.2 sur le formatage). En effet, ces formats offrent de la précision dans les tracés de courbe, et surtout ils autorisent des changements d'échelle (autrement dit des zooms) sans entraîner de phénomènes de crénelage (ou « aliasing »). C'est une des raisons pour lesquelles le format PDF est aujourd'hui le plus préconisé pour la consultation de documents textuels. En revanche, pour la visualisation de documents textuels dans un format de type bitmap, la résolution des moniteurs actuels vient dégrader la qualité de l'affichage et provoque de l'aliasing. En effet, dès que la résolution de numérisation du document est bien supérieure à la résolution d'affichage de l'écran – par exemple un texte scanné à 300 dpi, alors que l'écran n'offre que 96 dpi -, le crénelage apparaît si on souhaite par exemple visualiser le document en taille 100%. La seule solution consiste alors à zoomer – dans l'exemple cité il faudrait zoomer deux fois puisque 3×96 est proche de 300 - pour diminuer cette dégradation, mais alors la lisibilité est réduite à la fenêtre d'affichage du visualiseur.

La taille de l'écran s'exprime en général à l'aide d'une valeur en pouces, cette dernière correspondant en fait à la longueur de la diagonale de l'écran. La taille moyenne des moniteurs sur le marché ne cesse d'augmenter, et les standards actuels se situent entre 19 et 21 pouces. Une diagonale de 21 pouces présentant l'avantage de pouvoir afficher une double page A4 en taille 100%. Avec une telle taille et à la résolution de 96 dpi, cela signifie que l'écran affiche 1600 pixels en largeur et 1200 pixels en hauteur. Pour gérer autant de points les ordinateurs doivent être équipés de cartes graphiques spécialisées – notamment dotées de processeurs de calculs capables de prendre en charge le mode vectoriel -, et ce type de carte est désormais intégré comme un standard dans presque toutes les configurations d'ordinateurs actuellement sur le marché.

La fréquence de rafraîchissement des écrans LCD actuels est supérieure à la fréquence de balayage – on parle de balayage à cause du faisceau à électrons - de la plupart des anciens écrans à tubes cathodiques - ou écrans CRT. Les moniteurs LCD sont capables aujourd'hui

de rafraîchir à une fréquence de 75 Hz (autrement dit 75 fois par seconde) alors que les écrans à tubes rafraîchissaient à 60 Hz. Or une fréquence plus élevée donne une meilleure stabilité à l'image et préserve mieux de la fatigue oculaire.

La plupart des cartes graphiques et des écrans actuels permettent d'afficher 16 millions de couleurs, ce qui suffit largement dans le cadre d'une gestion de dossiers documentaires.

D'une manière générale, sur le plan de la consultation des documents numériques, on constate donc les évolutions technologiques et la baisse importante du coût du matériel informatique qui ont eu lieu ces dernières années, et qui se prolongent, permettent aujourd'hui d'exploiter pleinement la puissance et le potentiel des logiciels de GED. Néanmoins, la majorité des utilisateurs de GED ont souvent l'impression que leurs yeux fatiguent plus vite lors d'une lecture à l'écran que lors d'une consultation de documents papier. Très souvent ils sélectionnent donc à l'écran les documents qu'ils souhaitent consulter, puis les impriment pour les lire. L'amélioration du confort de consultation, et ce changement progressif de comportement, passe donc par un investissement dans des moniteurs adaptés et performants de la part des centres de documentations.

2.3.8 La diffusion

C'est la dernière étape de la chaîne GED, qui consiste à mettre en ligne les documents numériques sur le réseau Internet ou sur un intranet via des serveurs web. Ces documents intégrés dans l'application de GED sont ainsi accessibles quasi immédiatement, depuis n'importe quel poste connecté au réseau de diffusion, et ce simultanément par plusieurs utilisateurs. Néanmoins, ces accès peuvent être bien entendu limités et contrôlés, puisque tous les logiciels de GED actuels intègrent des options de gestion des droits de diffusion et d'utilisation – par personne ou par groupe d'utilisateurs.

2.4 L'offre logicielle de GED

Le marché des technologies et des prestataires de la dématérialisation s'est fortement enrichi et développé ces dernières années, il est en constante évolution et propose sans cesse de nouvelles technologies. Cela donne aux centres de documentation une grande marge de décision sur le périmètre de leur projet. Et il existe aujourd'hui sur le marché mondial une offre très large et très fournie de logiciels de GED, répondant à la plupart des besoins et compatibles avec la majeure partie des plates-formes informatiques et des systèmes d'exploitation actuels.

• Les différentes catégories de logiciels de GED

Cette offre peut se diviser en trois catégories.

- Les logiciels GED d'entrée de gamme :

Il ne représentent actuellement qu'une très infime partie du marché, mais touche un très large public du fait de leur prix attractifs. Il correspondent en général à des solutions mono-postes destinées à des entreprises artisanales ou individuelles. La majeure partie de ces applications ne fonctionnent qu'avec des scanners de bureau et ne permettent de saisir qu'un nombre limité de documents par jour.

- Les logiciels GED à composants :

Cette catégorie recouvre des modules logiciels intégrables à des applications au travers d'outils de développement (Visual C++) ou d'API²⁵. Ces modules – numérisation, reconnaissance optique de caractères, indexation automatique, lecture de codes à barres, etc - sont destinés aux programmeurs ou aux sociétés de services informatiques désirant ajouter des fonctions de GED à leurs logiciels.

- Les logiciels GED génériques :

Ce sont des solutions s'intégrant en complément des applications informatiques – bureautiques, documentaires, transactionnelles, etc – des organisations. Ils sont disponibles pour de nombreux systèmes d'exploitation et s'intègrent généralement dans une architecture client/serveur, en connexion ou non avec l'informatique centrale. Ils jouent le rôle de fédérateur au sein du système de gestion d'informations de l'organisation.

Cette dernière catégorie constitue la majeure partie de l'offre actuelle, et c'est elle qui concerne directement notre étude.

²⁵ Application programming interface, une « interface de programmation applicative ».

Enfin, des logiciels ou des modules de gestion de processus, regroupés sous le terme de « workflow », sont souvent associés aux programmes de GED. Ces outils permettent d'automatiser la circulation de documents ou d'informations selon des procédures précises. En général, ils vont plus loin que la simple distribution et sont capables d'attribuer des tâches de suivi de documents, puis de suivre et de surveiller les traitements réalisés.

- **Le conseil en dématérialisation**

Sur le plan du « conseil », l'étude du marché actuel révèle que, même si le prix des licences logicielles tend à baisser, les prestations de service en conseil ne sont en général pas formatées pour les organismes non munis de gros budgets. Ainsi le simple recours à un consultant spécialisé en dématérialisation, ne serait-ce que pour une ou deux journées, ne paraît pas être à la portée de tous les budgets des centres de documentations.

- **Les tendances fortes du marché**

Deux tendances fortes semblent émerger sur ce marché de la dématérialisation : le SAAS²⁶ et la « dématérialisation en boîte ». Ce sont deux technologies d'avenir qui vont être probablement de plus en plus prééminentes dans le futur.

Le SAAS est une offre de dématérialisation en ligne et qui cible plutôt les petits et moyens budgets. Les services SAAS sont souvent facturés à l'utilisation ou en mode locatif, incluant les frais de maintenance, sans aucun investissement de départ pour l'acquisition et le déploiement de logiciels.

La « dématérialisation mise en boîte » consiste en le principe suivant : des scanners et des copieurs multifonction peuvent être associés à des logiciels de dématérialisation prêts à l'emploi et conçus par des partenaires spécialisés. C'est une tendance qui consiste à faire converger les matériels de copie et d'impression et les logiciels de GED. Elle a surtout l'avantage d'offrir une simplicité d'utilisation et d'intégration, puisque, directement sur le copieur, on dispose d'un module d'accès à la GED via une interface intuitive et souvent simplifiée à l'extrême.

- **Les innovations technologiques en matière de compression de fichiers**

Sur un plan purement technologique, on peut citer des exemples très intéressants de la part d'éditeurs qui améliorent et font de gros efforts pour les algorithmes de traitement de la couleur dans les fichiers PDF. Leur objectif premier étant d'optimiser la restitution, le stockage et le traitement des fichiers.

Ainsi la société Iris, dans la nouvelle version de son serveur d'OCR et de compression de

²⁶ Software As A Service.

documents, propose de compresser une image couleur, en JPEG, en l'encapsulant dans un fichier PDF. Ainsi les éléments « textuels » sont distingués des « photos », et la technologie de compression peut ensuite être adaptée à chacun de ces éléments sans nuire à la qualité du texte sur lequel une technologie de reconnaissance peut être appliquée.

Autres innovations intéressantes, celles de l'américain Nuance et du français Pixelion, qui s'appuient eux sur la technologie de compression de page PDF MRC (Mixed Raster Content). Ce qui permet de segmenter chaque page en plusieurs parties : texte, image, couleur, texte OCR. Pixelion revendique ainsi une prouesse notable, celle de réussir à compresser un fichier PDF en couleur dans une taille équivalente à celle d'un fichier en noir et blanc.

Troisième partie

Les dossiers documentaires au centre de documentation des musées des Arts Décoratifs

3.1 Présentation des Arts Décoratifs

3.1.1 L'institution

Organisme privé, association loi 1901 reconnue d'utilité publique, Les Arts Décoratifs²⁷ furent créés en 1882 dans le sillage des Expositions universelles par des collectionneurs soucieux de valoriser les beaux-arts appliqués et de tisser des liens entre industrie et culture, création et production.

Longtemps connus sous la dénomination Union Centrale des Arts Décoratifs (Ucad), Les Arts Décoratifs ont modernisé leur image en décembre 2004 en restant fidèles à leur vocation d'origine : exercer des missions de conservation des collections et de diffusion culturelle, et de soutien à la création.

Les différentes composantes des Arts Décoratifs sont réparties sur trois sites à Paris :

- au 107 rue de Rivoli, les ailes de Rohan et de Marsan du palais du Louvre abritent le musée des Arts Décoratifs, le musée de la Mode et du Textile, le musée de la Publicité, la bibliothèque des Arts Décoratifs ;
- au 63 rue de Monceau, l'hôtel Camondo accueille le musée Nissim de Camondo ;
- au 266 boulevard Raspail est installée depuis 1988 l'école Camondo (appellation née d'une implantation antérieure dans les communs de l'hôtel Camondo), spécialisée dans le design et l'architecture intérieure ;

Les Arts Décoratifs, dont les collections appartiennent à l'état, sont liés à ce dernier par une convention renouvelée périodiquement. Son financement est mixte, près de la moitié de son budget provient de ses ressources propres et finance la politique culturelle et patrimoniale de l'établissement. Les Arts Décoratifs continuent à se développer essentiellement grâce au soutien de donateurs et d'entreprises partenaires.

3.1.2 Le musée de la Publicité

En 1978, l'Ucad ouvra le musée de l'Affiche - rue de Paradis - dans l'intention d'explorer toutes les voies des arts décoratifs, y compris les plus quotidiennes, les plus banalisées et les plus largement partagées. En 1990, le musée de l'Affiche est devenu le musée de la Publicité et s'est installé rue de Rivoli. Le musée de la Publicité a ainsi pour mission de collecter et de mettre en valeur un fonds dédié à la publicité et à l'affichage. Outre l'exceptionnel fonds

²⁷ Site internet : www.lesartsdecoratifs.fr ; Un organigramme du musée des Arts Décoratifs est consultable en annexe 5 à la page 179.

d'affiches (50.000 affiches anciennes du XVIII^e siècle à la Seconde Guerre mondiale et 50.000 affiches contemporaines de 1950 à aujourd'hui), les collections du musée se sont enrichies de films publicitaires (plus de 20. 000 films, français ou étrangers, des années 30 à aujourd'hui), d'annonces presse (plus de 30.000), de spots radio, d'objets promotionnels.

A travers ces collections, le musée transmet un héritage culturel important, et oeuvre à le faire connaître et apprécier du public. Outre le travail de collecte et de conservation, le musée organise ainsi la diffusion de ce fonds à travers des expositions, des colloques ou des ateliers, destinés notamment au jeune public. Le musée participe également à la promotion de la culture par le biais de nombreuses publications et de collaborations à divers ouvrages.

La gestion administrative étant commune à tous les départements des Arts Décoratifs, le personnel dédié exclusivement au musée de la publicité se compose de deux conservatrices et d'une documentaliste – chargée d'études documentaires. La documentation du musée de la Publicité dépend conjointement de la conservation du musée et du centre de documentation des Arts Décoratifs.

3.1.3 Le centre de documentation des musées des Arts Décoratifs

Les Arts Décoratifs proposent aux chercheurs, universitaires, créateurs, historiens et critiques d'art, un centre de documentation en liaison avec les quatre musées, accessible sur rendez-vous. Créé au fil des décennies en fonction des accroissements de certains fonds ou de la création de départements spécialisés, il documente les œuvres des collections.

Les fonds documentaires sont répartis au sein des documentations suivantes :

- documentation des arts décoratifs du 17^{ème} et 18^{ème} siècle
- documentation sur l'Art nouveau Art Déco
- documentation du 20^{ème} siècle (le design, les créateurs, le bijou, la céramique)
- documentation du verre
- documentation du jouet
- documentation de la mode et du textile
- documentation de la publicité

Ces documentations sont réparties par bureaux, mais elles partagent une salle de consultation et d'accueil du public, où se trouvent notamment 6 postes informatiques sous Windows XP. Les utilisateurs – que ce soient des consultants internes comme les conservateurs, ou externes avec le public accueilli – accèdent via ces postes aux bases de données des collections et du fonds documentaire. Ces bases sont hébergées sur des logiciels de la société Mobydoc : Micromusée pour la gestion des collections, et Mobytext pour la gestion du fonds documentaire.

Depuis quelques années, le centre de documentation propose également des ressources en ligne sur Internet dans les domaines suivants : le verre, le design, le bijou, la céramique contemporaine, le jouet, la mode et le textile, la publicité.

3.1.4 La documentation de la publicité

Créé en 1980, ce fonds documentaire est spécialisé dans l'histoire, la création, la technique de l'affiche et de la publicité, françaises et étrangères, anciennes et contemporaines. Il a principalement vocation à documenter les collections du musée, à l'intention des conservatrices du musée mais aussi du public. Il est également destiné à apporter des informations complémentaires relatives à la publicité d'une manière plus générale. Ce qui passe par la collecte de toutes les informations pouvant être utiles aux utilisateurs, qu'elles se rapportent ou non aux collections.

Les missions de la documentaliste responsable de ce fonds documentaire sont la gestion des abonnements, l'acquisition et le traitement des documents, la veille documentaire, et l'accueil du public. Un travail important est notamment effectué sur les dossiers documentaires, qui représentent une plus-value très importante de cette documentation de la publicité. La documentaliste procède donc à un dépouillement régulier de la presse spécialisée, ainsi que de la presse généraliste mais dans une moindre mesure, afin d'alimenter ces dossiers en fonction des tendances et de l'actualité. Et de nouveaux dossiers peuvent être créés en fonction des demandes du public, ou de l'apparition de nouveaux thèmes sur le marché de la publicité. La documentaliste est aussi en charge de la gestion de la base de données des collections - notamment l'insertion, la correction et l'harmonisation des entrées dans cette base, et la préparation de la numérisation.

Le fonds documentaire de la publicité comprend notamment :

- environ 4500 ouvrages, traitant de la publicité et de ce qui s'y rapporte d'une manière plus générale ;
- environ 70 revues françaises et étrangères, anciennes et contemporaines ;
- environ 6800 dossiers monographiques d'artistes ;
- plusieurs centaines de dossiers documentaires thématiques (répartis dans près de 200 boîtes d'archives) ;

Actuellement, on compte une dizaine d'abonnements courants, uniquement des revues spécialisées dans la publicité et la communication (dont Stratégies et CB news). Les numéros récents sont disponibles en salle de consultation.

Une partie de la documentation – notamment le stock des anciens numéros des abonnements courants, archivés dans la majorité des cas sous forme de reliures – est conservée dans une salle d'archives commune à toutes les documentations – dénommée salle des compactus.

Tous les documents du fonds documentaire sont consultables uniquement sur place. En 2008, 40 personnes ont été reçues par la documentaliste, sur rendez-vous et après avoir défini au préalable avec elle le sujet de leur recherche. Ce public accueilli est composé au trois quarts d'étudiants issus de l'enseignement supérieur - dans des domaines allant généralement de la communication aux arts appliqués -, et leurs demandes sont la plupart du temps ciblées et précises. Le reste du public se compose de chercheurs, d'universitaires, et de professionnels de la publicité.

3.2 L'analyse de l'existant

3.2.1 Les dossiers thématiques de la publicité

Par définition, au sens le plus large, un dossier rassemble des documents et des informations sur un sujet déterminé. Cette définition englobe donc nombre de produits et nombre de métiers.

Ce mémoire se base sur une étude de faisabilité réalisée dans le cadre d'un projet de dématérialisation, et traite essentiellement de la numérisation des dossiers documentaires illustrant les collections de la publicité.

Dans le vocabulaire de la documentation, la définition d'un dossier documentaire précise que c'est « un produit documentaire constitué d'un ensemble de documents de sources diverses, choisis et réunis sur une question donnée, et organisé de façon à faciliter l'accès à l'information rassemblée. Les éléments qui le composent peuvent être de nature et de support très variés : articles de revues, brochures, etc. ... Il peut être constitué ponctuellement à la demande ou systématiquement sur des thèmes précis. Un dossier documentaire peut être sur support papier ou électronique. »

Au sein du centre de documentation des musées des Arts Décoratifs, la documentation de la publicité gère des dossiers documentaires, au format papier, qui sont en fait de deux types :

- les dossiers thématiques (classés par thème) ;
- les dossiers d'artistes (classés par ordre alphabétique) ;

L'étude de faisabilité a été centrée plus spécialement sur les dossiers thématiques. Néanmoins, elle propose aussi un début de réflexion et d'analyse sur l'opportunité d'une dématérialisation d'autres dossiers documentaires au sein des autres documentations du centre.

Les dossiers thématiques de la publicité sont actuellement alimentés par des panoramas de presse, réalisés à partir des périodiques suivants – auxquels est abonné le centre de documentation - : CB news (qui fait actuellement l'objet d'un abonnement spécial proposant la version électronique en plus de la version papier), Stratégies, Marketing magazine (également disponible en version papier et en version électronique), Etapes graphiques, La revue des marques, Le temps des médias, Market management, Histoire d'entreprises. Les panoramas peuvent aussi intégrer parfois, mais de manière beaucoup plus ponctuelle, des articles issus de journaux auxquels le centre de documentation n'est pas abonné – un quotidien national par exemple -, ou bien d'autres documents divers – brochure publicitaire, enquête, rapport, sondage, etc.

Ces dossiers thématiques - permettant de trouver des informations et des documents sur des thèmes afférents à la publicité - sont ainsi constitués très majoritairement d'articles photocopiés. Ces derniers sont conditionnés et triés selon un plan de classement thématique²⁸, dans des boîtes d'archives²⁹ au format A4 - regroupant des thèmes - qui contiennent elles-mêmes des chemises - regroupant des sous thèmes.

Ces dossiers sont principalement utilisés par la documentaliste responsable de la publicité, ainsi que par les conservateurs et assistants de ce département. Ils sont aussi signalés sur le site des Arts Décoratifs, et sont consultables sur demande par des visiteurs.

La constitution de ce fonds relève des fonctions et des compétences de la seule documentaliste de la publicité, qui prend ainsi en charge toutes les étapes de la chaîne de traitement documentaire.

3.2.2 La chaîne documentaire

La chaîne de traitement documentaire s'organise actuellement en les phases suivantes :

- réception des périodiques (au format papier par voie postale) ;
- bulletinage (mise à jour des abonnements en vue de garder une traçabilité) ;
- dépouillement (travail intellectuel de sélection des articles les plus intéressants et les plus pertinents) ;
- photocopie des articles sélectionnés (à partir des périodiques originaux) ;
- indexation manuelle consistant à écrire à la main les références des articles sur les photocopies (titre du périodique, numéro, date de parution, parfois quelques mots en commentaire) ;
- indexation électronique dans le logiciel Mobytext pour les revues reçues en version électronique (actuellement CB news et Marketing magazine) ;
- classement définitif dans les boîtes d'archives suivant un plan de classement thématique (tournant autour des domaines de la publicité, du marketing et de la communication) ;

Dans cette chaîne, les opérations de photocopie occupent une part importante, puisque tous les articles sélectionnés lors du dépouillement sont systématiquement photocopiés. Dans certains cas seulement, lorsque l'article est issu d'un périodique auquel n'est pas abonné le musée – par exemple un journal quotidien comme Le Monde -, les articles sont découpés aux ciseaux directement à partir de l'original. Ils sont ensuite très souvent collés sur des feuilles A4, qui sont alors elles-mêmes photocopiées puis classées – ou classées directement. Ceci afin d'obtenir une meilleure présentation de l'article. Cette tâche manuelle est donc

²⁸ Un extrait de ce plan de classement thématique est consultable en annexe 7.

²⁹ Des photos illustrant la structure de ces boîtes d'archives sont consultables en annexe 8.

relativement lourde, d'autant que, dans certains cas, l'article concerne plusieurs dossiers ou thèmes, ce qui multiplie alors d'autant le nombre d'exemplaires à produire lors de l'acte de photocopie.

L'opération d'indexation numérique concerne uniquement les périodiques reçus par voie électronique – actuellement il s'agit de CB news et Marketing magazine -, et se fait à l'aide d'un thésaurus. Cette opération est actuellement effectuée, manque de mieux, via le logiciel Mobytext qui se trouve être inadapté à la gestion de dossiers documentaires dématérialisés.

Il est livré ci-dessous des informations et des éléments clés permettant de faire une première évaluation de la situation, et d'engager une réflexion sur la faisabilité et l'opportunité de la dématérialisation.

3.2.3 Recensement de l'existant et typologie des dossiers thématiques de la publicité

La première étape d'une étude de faisabilité d'un projet de dématérialisation consiste à :

- faire un recensement de l'existant, notamment la volumétrie et les caractéristiques pouvant influencer sur la méthode ou le résultat ;
- établir une typologie du fonds selon les caractéristiques les plus contraignantes ;
- prendre en compte toute une série de facteurs qui concernent plus particulièrement le texte des documents ;

Tous ces points ont été regroupés dans un tableau consultable en annexe 1, à la page 165 de ce document.

En fonction de tous les éléments et de toutes les informations qu'il a été possible de recueillir au centre de documentation, voici ce que l'on peut dire sur chacun des points évoqués dans ce tableau.

- **Le type de document**

Ce sont exclusivement des articles de presse ou de revue.

- **La forme :**

Il s'agit de photocopies – et, beaucoup plus rarement, d'originaux - d'articles de presse ou de revues, qui comportent du texte et des images, en couleur ou noir et blanc. Ces documents contiennent parfois de courtes annotations manuscrites, la plupart du temps les références de l'article, qui ont été écrites à la main soit directement sur l'original, soit sur la photocopie.

- Le contenu :

Il s'agit de textes relativement riches qui n'ont donc pas une structure forte (à l'inverse par exemple de documents de type formulaires).

- L'usage futur :

Ces documents ont vocation à être lus sur des postes informatiques, en accès libre ou contrôlé. Ils n'ont pas de valeur probante, ne constituent pas des preuves, ne font pas l'objet d'une durée de conservation légale, ne seront pas achetés. Une fois ces documents dématérialisés et intégrés dans la GED, seuls pourront être modifiés leur indexation et leur classement électroniques. Autrement dit, aucune édition électronique du contenu de ces documents n'est envisagée après leur stockage dans la base. De même, aucune action du type « réimpression-modification sur papier-numérisation » n'est programmée.

• Le format du document

- La taille :

Les documents sont très majoritairement au format A4. Quelques-uns sont au format A3. Ces formats sont pris en charge par la quasi-totalité des scanners semi-professionnels actuellement sur le marché, y compris les moins performants.

- La couleur :

Une partie non négligeable des documents sont en couleur, ce qui ne pose aucun problème sur le plan de l'efficacité de l'OCR.

- Le papier :

Concernant les dossiers thématiques le grammage du papier est en général standard (papier pour photocopieuse). La majorité des documents sont en bon état.

Seuls quelques documents sont jaunis avec un papier dégradé, mais, excepté pour une petite poignée de documents très dégradés, leur état ne paraît pas incompatible avec une numérisation. De plus, aucun document n'est véritablement rare au point que l'on doive prendre en considération un risque potentiel de dégradation lors de la scannerisation.

Concernant les revues et journaux originaux papier qui seront numérisés dès leur réception, la qualité de leur papier est également compatible avec une numérisation. Précisons qu'une option très intéressante et efficace, appelée « filtre anti-moiré », est disponible sur beaucoup de scanners et permet d'atténuer la présence et l'apparition de « trames d'impression » caractéristiques des journaux et revues commercialisés.

- La pagination :

Dans ces dossiers documentaires, on trouve majoritairement des feuilles photocopiées seulement au recto. Une petite proportion des documents – autour de 20 % - sont toutefois

imprimées recto-verso. Certains articles s'étalent sur plusieurs pages, on trouve donc un grand nombre de feuilles agrafées. Ce problème sera abordé plus loin dans l'étude.

- **Le volume du fonds documentaire à traiter**

Concernant le stock des dossiers thématiques de la publicité, le volume des pages à numériser est assez important. Actuellement, le nombre de boîtes d'archives s'élève à environ 200. Toutes les boîtes sont au format A4. Un comptage manuel a permis d'évaluer qu'une boîte comporte environ 400 feuilles lorsqu'elle est remplie au 2/3 – ce qui correspond à peu près au taux de remplissage moyen des boîtes. On peut donc en déduire que le volume de ce fonds tourne autour de 80000 feuilles. La très grande majorité, autour de 80%, ne sont imprimées que sur le recto. Ainsi le nombre total de pages A4 à dématérialiser pour ce projet tourne autour de 100000.

- **L'homogénéité du fonds**

La nature des documents est plutôt homogène, puisque la quasi-totalité des documents sont des photocopies d'articles – plus rarement des originaux. Il paraît donc pertinent et judicieux de les traiter d'un seul bloc et au cours d'un seul et même projet.

En termes de format, le fonds des dossiers thématiques est plutôt homogène puisque la quasi-totalité des documents sont des photocopies au format A4.

En revanche, sur le plan de la qualité de reprographie, le fonds n'est pas vraiment homogène et l'on trouve des qualités d'impression assez variées selon les documents.

- **L'état et la fragilité du document initial**

Les documents présents dans les dossiers thématiques de la documentation publicité ne sont pas spécialement fragiles ou précieux. Il n'y a donc pas, sur ce point, une obligation à faire un choix technologique très précis ou à se doter d'un modèle de scanner en particulier.

En revanche, on trouve dans d'autres documentations du musée - comme celle du 17^{ème} et 18^{ème} siècle par exemple - des revues précieuses et fragiles à la valeur historique et patrimoniale importantes. La dématérialisation de ces documents exigera une grande précaution, et il est préconisé dans ces cas-là de faire appel à un prestataire spécialisé en numérisation, possédant des scanners particuliers capables de numériser sans abîmer les reliures ni le papier.

- **La juridiction liée au document**

Sur le plan juridique, parmi les dossiers documentaires du centre de documentation, il n'existe aucun document qui soit de nature « non reproductible », aucun document à caractère « probant », aucun document n'ayant le statut de « preuve ». Il n'y a donc aucune

obligation pour le centre de documentation de conserver un quelconque document original, et il est donc possible en théorie de détruire les documents après leur numérisation sans que cela ne mette potentiellement les documentations dans une situation de risque ou de mise en défaut. Il n'y a également aucune obligation particulière pour ce projet – mais en revanche des recommandations que cette étude est censée apporter - en termes de pérennité du support, de format de fichier, de signature électronique, de datage, ou d'identifiant unique.

Les aspects juridiques du projet qui concernent le droit d'auteur et le droit de reproduction seront traités plus loin dans la partie 3.9 de l'étude.

3.2.4 Recensement de l'existant et typologie des dossiers dans les autres documentations

- **Documentation de la mode et du textile**

Le fonds documentaire de la documentation de la mode et du textile comprend environ 1300 dossiers, pour la plupart consacrés aux couturiers et aux créateurs.

Les documents archivés dans les dossiers de cette documentation sont de toute nature : articles photocopiés de revues ou de journaux, dossiers de presse, catalogues de collections, documents originaux, photos, etc. Beaucoup de ces documents ont une valeur patrimoniale, puisque ce sont des originaux que seul le musée possède. Les dossiers peuvent aussi contenir des objets, on peut citer comme exemple cette reproduction d'appareil photo en plastique faisant office de carton d'invitation pour un défilé de haute couture. Ainsi se trouve dans ces boîtes des documents « non dématérialisables ».

L'état des documents est très variable. Certains sont sous forme de photocopies de très bonne facture, d'autres sont au contraire dans un très mauvais état et non numérisables (papiers originaux complètement noircis ou jaunis, photocopies de très mauvaise qualité).

- **Documentation du 20^{ème} siècle**

Le fonds est déjà très important, malgré qu'il n'ait été constitué que depuis les années 90. Il comprend aujourd'hui plus de 1500 dossiers, soit plusieurs centaines de milliers de pages de documents.

Les documents sont de nature très variée : articles photocopiés de revues ou de journaux, dossiers de presse, catalogues d'artistes, documents originaux extraits de dossiers personnels d'artistes, etc

Tout comme la documentation de la mode du textile, l'état des documents est assez variable - photocopies de bonne qualité, papiers originaux noircis ou jaunis, etc.

- **Documentation Art nouveau Art déco**

Le fonds des dossiers documentaires de la documentation Art nouveau Art Déco est assez important, il porte sur les oeuvres du 19^{ème} siècle et aussi beaucoup sur l'exposition universelle des Arts décoratifs à Paris en 1925. Les documents sont majoritairement des articles - originaux ou photocopies -, plutôt dans un bon état de conservation, et d'une qualité d'impression assez satisfaisante. Aucun document dans ces dossiers n'est à caractère patrimonial ou d'une très grande valeur de rareté.

- **Documentation du 17^{ème} et du 18^{ème}**

Le fonds documentaire de cette documentation n'est pas très volumineux et assez jeune, puisqu'il a commencé à être constitué seulement en 2002. Les documents sont majoritairement des articles - originaux ou photocopies. Le fonds ne comporte pas de documents à caractère patrimonial ou d'une très grande valeur de rareté. L'état de conservation des documents est globalement satisfaisant, et tous sont compatibles avec une numérisation physique.

3.2.5 L'initiative d'un dépouillement électronique de certains périodiques à l'aide du logiciel Mobytext

Soulignons une initiative importante et intéressante qui a été menée depuis quelque temps par la documentaliste de la publicité : celle d'un dépouillement électronique des revues CB News et Marketing magazine. Des numéros de ces revues ont ainsi pu être téléchargés dans leur intégralité en version électronique au format PDF. La documentaliste a instauré, via le logiciel documentaire Mobytext, un système qui permet de faire référence et d'indexer, de façon très sommaire, un article d'un numéro. Un lien permet ainsi d'ouvrir le fichier PDF du numéro concerné, mais pas d'accéder directement à la page du numéro contenant l'article référencé. En effet, Mobytext est un logiciel spécialisé dans la gestion des collections patrimoniales, dont la conception autorise seulement le catalogage d'articles de périodiques, et non pas leur gestion – autrement dit l'acquisition, la modification, la publication, etc. Ce point sera exposé plus précisément dans la partie 3.3 de cette étude.

Aujourd'hui, on trouve donc plus de 3000 références d'articles sur la publicité dans la base de Mobytext. Néanmoins, cette solution technique est forcément provisoire, car elle ne bénéficie pas de la puissance d'un véritable logiciel de GED. Mais elle est surtout symbolique d'une volonté de la part de la documentaliste de disposer d'un fonds complètement numérisé.

A propos de ce dépouillement électronique de périodiques, qui est de plus en plus pratiqué par les documentalistes des centres de documentation, il semble que certains éditeurs de revue aient réagis et soient maintenant opposés au téléchargement de leurs numéros en version intégrale. C'est justement le cas pour CB news et Marketing magazine depuis quelques mois. L'explication repose dans la problématique actuelle de l'exploitation des panoramas de presse électronique, et dans ses aspects juridiques particuliers. Ce point important sera largement traité dans la partie 3.9.4 de ce mémoire.

3.3 Analyse de contenus de boîtes d'archives

Pour cette étude, il a été décidé de procéder à une analyse assez fine, sous forme de statistiques, du contenu de boîtes d'archives renfermant des dossiers thématiques de la publicité.

Ceci afin de :

- sonder précisément la variété des ressources présentes dans ces dossiers ;
- évaluer la prédominance de certains périodiques ;
- évaluer l'état des documents contenus dans ces boîtes ;

Nous avons réalisé le dépouillement d'une première boîte - c'est-à-dire d'un thème - dont le titre principal inscrit dessus est « Femme et publicité ».

A l'intérieur de cette boîte, les ressources sont classées en sous thèmes, ces derniers étant inscrits en titres secondaires sur la boîte. Les sous thèmes pour cette boîte étaient : protection de la femme, typologie de la femme, la pin-up.

Cette boîte contient 271 documents - très majoritairement des articles -, sachant qu'un document peut bien évidemment comporter parfois plusieurs pages.

A partir de l'analyse de ces documents, on a souhaité isoler et identifier les périodiques - revues, journaux, etc - les plus représentés, afin d'essayer de déterminer s'il pourrait s'avérer intéressant de procéder à une numérisation rétrospective à partir des originaux de ces périodiques - qui sont conservés par le musée -, plutôt que de numériser à partir des photocopies déjà présentes et classées dans les dossiers thématiques.

Voici les résultats du dépouillement de cette boîte sous forme de statistiques (où ne figurent, volontairement, que les titres de périodiques les plus représentés) :

Cb news	28 documents (10,3 %)
Stratégies	26 documents (9,5 %)
Le figaro	14 documents (5,1 %)
Libération	13 documents (4,8 %)
Le monde	10 documents (3,7 %)
Le journal du textile	6 documents (1,84 %)
Médias	5 documents (1,82 %)
Le point	4 documents (1,47 %)
Vogue	4 documents (1,47 %)
Com.in	4 documents (1,47 %)
Le nouvel observateur	3 documents (1,1 %)
L'express	3 documents (1,1 %)
Télé star	3 documents (1,1 %)
Femme actuelle	2 documents (0,73 %)
Télérama	2 documents (0,73 %)

Marianne 2 documents (0,73 %)
Fashion daily news 2 documents (0,73 %)

Il ressort de ces chiffres que deux périodiques se détachent et sont vraiment très présents dans cette boîte thématique. Il s'agit de CB news et de Stratégies (à hauteur de 10 % chacune).

Les originaux de ces deux revues étant archivés et disponibles au centre de documentation – ce qui n'est pas le cas des trois périodiques qui suivent dans le classement, c'est-à-dire Le figaro, Libération et Le monde - une numérisation à partir de ces originaux serait donc possible. Le désavantage de cette solution étant que ces périodiques originaux sont actuellement conservées sous forme de reliures. Une numérisation massive avec utilisation d'un chargeur automatique impliquerait donc au préalable de détruire ces reliures, puis de les reconstituer une fois l'opération terminée.

En plus de ces articles, on a trouvé aussi :

- 1 rapport d'experts de plusieurs dizaines de pages, relié avec des agrafes ;
- 1 carton d'invitation pour un vernissage, assez petit et plié ;
- 1 enquête pour la télévision, sous forme d'une pochette contenant des pages agrafées ;
- 5 textes de lois, comprenant chacun plusieurs pages agrafées ;

Concernant l'état et la forme de conservation de ces documents, on a recensé parmi ces derniers ceux qui pouvaient présenter des caractéristiques susceptibles de nuire, gêner ou retarder leur dématérialisation. De la même manière que pour la diversité des revues exposée plus haut, les résultats sont présentés ci-dessous sous forme de statistiques :

Documents au format A3 :	25 (9,22 %)
Documents composés de moins de 5 pages agrafées	36 (13,28 %)
Documents composés de plus de 5 pages agrafées	23 (8,48 %)
Documents reliés	2 (0,73 %)
Documents comprenant des références manuscrites	96 (35,4 %)
Documents comprenant des lignes de textes inclinées	1 (0,37 %)
Documents cornés	1 (0,37 %)
Documents froissés	2 (0,73 %)
Documents pliés	1 (0,37 %)
Documents d'origine inconnue	10 (3,7 %)
Documents comprenant des lignes illisibles	2 (0,73 %)
Documents insérés dans des pochettes plastiques	11 (4,05 %)

Parmi ces résultats, les deux chiffres importants et qui sont probablement les plus déterminants dans l'optique d'une numérisation sont :

- le nombre de documents comprenant des références manuscrites ;
- le nombre de documents comprenant des pages agrafées ;

On voit que plus d'un document sur cinq est agrafé, ce qui nécessiterait par conséquent l'intervention d'un opérateur avant de pouvoir placer les feuilles dans le chargeur automatique. C'est un paramètre très important et fortement pénalisant dans le cas où l'on souhaiterait une dématérialisation rapide et massive.

Si on additionne tous les autres pourcentages de documents qui ne pourront pas être placés directement dans le chargeur sans un minimum de préparation au préalable (c'est-à-dire les documents pliés, cornés, froissés, insérés dans des pochettes plastiques), on avoisine un total de 17 %, ce qui représente là encore près de 1 document sur 5.

Sur le plan de la performance du traitement OCR, et des potentialités d'indexation ou de classement automatiques qu'il serait susceptible d'induire, on voit que, là aussi, les choses ne se présentent pas favorablement.

En effet, plus d'un document sur 3 comporte ainsi des références manuscrites fondamentales et non reconnaissables de façon fiable avec l'OCR - cela est clairement expliqué dans la partie 3.4.4 de ce mémoire. Dans la très grande majorité des cas, ces inscriptions manuscrites sur les photocopies sont : le titre, le numéro et la date de parution du périodique. Des informations qu'il faudra donc forcément re-saisir manuellement lors de la numérisation. Une tâche assez coûteuse en temps et en attention si on la compare au reste des opérations comme le chargement des feuilles dans le chargeur, le temps de balayage de la surface du document par le laser, le traitement OCR pour le texte intégral, l'enregistrement du fichier au format de sortie souhaité, etc.

En fait, à ce taux de 1 sur 3, il faut ajouter en plus le pourcentage de documents d'origine inconnue (3,7 %). Ces devront en effet obligatoirement faire l'objet eux aussi d'une saisie manuelle lors de la numérisation.

Le fonds des dossiers thématiques de la publicité étant très riche et très varié, nous avons considéré qu'il était indispensable de dépouiller et d'analyser le contenu d'une autre boîte d'archivage. Le thème sélectionné fut celui des « Boissons non alcoolisées », qui se décline selon 5 sous thèmes dans le plan de classement : Sodas, Jus de fruits, Sirops, Apéritifs sans alcool, Boissons énergétiques.

La boîte comprend 374 documents, chacun pouvant comporter plusieurs pages.

Voici les périodiques les plus représentés dans ce dossier thématique :

Stratégies	106 documents (28,3 %)
CB news	87 documents (23,2 %)
Le figaro	31 documents (8,2 %)
Marketing magazine	9 documents (2,4 %)
Médias	9 documents (2,4 %)

Cash marketing	8 documents (2,1 %)
Télé 7 jours	7 documents (1,8 %)
Le point	5 documents (1,3%)
Le monde	4 documents (1 %)
Le nouvel observateur	4 documents (1 %)

De la même façon que pour le thème « Femme et publicité », il ressort là encore de ces statistiques que ce sont les deux mêmes périodiques qui se détachent : Stratégies et CB news. Mais on constate cette fois qu'ils sont encore plus majoritaires dans ce thème des boissons non alcoolisées, en représentant près de 1 document sur 4. L'opportunité et l'intérêt de numériser les articles de ces deux revues à partir des originaux apparaissent ici de façon encore plus évidente.

Comme cela a été expliqué plus haut, les originaux de ces revues sont déjà reliés et il apparaît donc nécessaire de débattre et réfléchir sur la stratégie qui serait la plus intéressante en termes de coût et de temps.

Vaudrait-il mieux numériser les revues page par page sur un scanner « à plat », et sans procéder à la destruction des reliures, ou bien serait-il plus rentable de détruire les reliures afin de pouvoir placer les pages dans le chargeur automatique ?

Une interrogation persiste aussi concernant le fait de savoir si oui ou non il est possible de détruire la reliure sans abîmer les revues ? Et est-il possible de procéder ensuite à une nouvelle reliure ? Sachant qu'il est souhaité et prévu de continuer à conserver les originaux de ces revues après la dématérialisation.

Dans le cas d'une dématérialisation qui ferait appel à un prestataire de numérisation externe, une autre alternative est aussi à étudier : celle de l'utilisation par ce prestataire d'un scanner professionnel capable de tourner les pages automatiquement.

Toujours dans l'hypothèse d'une numérisation à partir des originaux, il faut souligner le fait que l'OCR pourrait peut-être avoir la capacité (cela reste tout de même à tester et à vérifier, peut être avec le paramétrage de positions et de zones) de reconnaître le titre, le numéro et la date de parution du périodique en cours de numérisation. D'où une indexation et un classement qui pourraient être alors partiellement automatisés, ce qui ferait gagner du temps.

Voici les statistiques sur l'état de conservation des documents dans ce dossier :

Documents au format A3 :	29 (7,7 %)
Documents composés de moins de 5 pages agrafées	33 (8,8 %)
Documents composés de plus de 5 pages agrafées	4 (1 %)
Documents comprenant des références manuscrites	149 (39,8 %)
Documents comprenant des lignes de textes inclinées	2 (0,5 %)
Documents cornés	2 (0,5 %)
Documents froissés	2 (0,73 %)
Documents pliés	41 (11 %)

Documents d'origine inconnue	5 (1,3 %)
Documents insérés dans des pochettes plastiques	19 (5 %)
Documents trop petits pour être placés dans le chargeur	26 (6,9 %)

On constate que près d'1 document sur 10 est agrafé et nécessitera une intervention manuelle pour préparer le document avant la numérisation (d'où un retard et un surcoût dans le traitement).

Et puis, là aussi, beaucoup de références manuscrites sont présentes sur les photocopies, avec près de 40 % des documents qui sont annotés avec des éléments indispensables à l'indexation.

Les documents pliés (1 document sur 10), ainsi que les petits morceaux d'articles découpés directement à partir des revues originales (et ne pouvant pas être entraînés par le chargeur), devront aussi faire l'objet d'un travail de préparation qui coûtera du temps.

Les documents au format A3 ne poseront à priori pas de problème, sauf qu'étant stockés dans des boîtes au format A4 ils ont forcément été pliés. Certaines des pliures sont donc susceptibles de gêner l'entraînement par un chargeur.

De manière générale il est donc recommandé de mener des tests avec des échantillons sur ces points précis, afin d'évaluer les difficultés potentielles pour un traitement de masse.

3.4 L'inadéquation des logiciels Mobydoc pour la gestion de dossiers thématiques dématérialisés

La dématérialisation représente une intéressante opportunité pour le centre de documentation de mettre en place un système de Gestion Electronique de Documents, qui viendrait en complément de l'outil déjà existant de gestion informatique documentaire spécifique aux collections. En effet, les seuls logiciels documentaires disponibles actuellement sur les postes du centre de documentation sont tous distribués par la société Mobydoc³⁰, et peuvent être considérés comme des outils relevant de la simple gestion informatique documentaire. Et une rapide analyse des spécifications fonctionnelles et techniques des deux logiciels Micromusée et Mobytext implantés au musée révèlent qu'ils ne permettent pas de réaliser une dématérialisation et une gestion électronique de dossiers thématiques. Ils sont même incapables de procéder à la première, et la plus élémentaire des phases de la dématérialisation : la phase de capture numérique.

En effet :

Micromusée permet seulement la gestion des collections des musées d'Art, d'Archéologie, d'Ethnographie ou de musées spécialisés dans tous ses aspects :

- l'inventaire ;
- la description physique des objets ;
- la documentation scientifique et littéraire : photographie, bibliographie... ;
- l'analyse iconographique et historique ;

et :

La vocation de Mobytext est exclusivement de cataloguer dans le respect de l'ISBD³¹ :

- des documents de tout type de support (livre, rapport, périodique...)
- des articles de périodiques ;
- des notices de dépouillement ;

Avant de rentrer plus en détails dans les spécifications de ces deux logiciels, et d'exposer leurs insuffisances en vue de la dématérialisation des dossiers thématiques de la documentation publicitaire, il n'est pas inutile de faire un petit rappel sur la société Mobydoc.

L'éditeur Mobydoc est spécialisé depuis l'origine dans l'informatique appliquée à la **gestion des collections patrimoniales** : musées, Collections d'entreprises ou privées, Archives,

³⁰ www.mobydoc.fr/index.htm

³¹ International Standard Bibliographic Description.

Centres de Documentation, Photothèques, Cinémathèques, Architecture. On peut même dire que c'est une société spécialisée **exclusivement dans la gestion de collections**, qui propose des produits aux professionnels du patrimoine.

Elle offre :

- des solutions logicielles pour les **collections patrimoniales** ;
- un système intégré multi-métier (MUM) permettant **la gestion centralisée de collections diverses** (un système actuellement unique sur le marché) ;
- l'OPAC Web pour la **mise en ligne des collections** : un outil intégré à la mise en œuvre assez simple, qui ne nécessite pas de compétence informatique préalable ;

Dans les spécifications de **Mobytext**, il est indiqué que le catalogage d'articles de périodiques est possible. Néanmoins, après une analyse plus détaillée, il s'est avéré que ce logiciel n'est pas, de par son origine, sa nature et sa conception, un outil qui permettrait de numériser et de gérer massivement et électroniquement les dossiers documentaires du centre de documentation.

En effet :

- il ne dispose d'aucun module de capture numérique ni de pilotage de scanner ;
- il ne permet pas le traitement de très grands volumes d'enregistrements de données ;
- ses fonctionnalités concernant l'indexation, le classement et la recherche électroniques des documents sont trop limitées pour le traitement de nombreux et volumineux dossiers documentaires ou thématiques ;
- les contraintes pour son paramétrage (création ou modification de champs par exemple) sont lourdes et imposent une maintenance que seul l'éditeur ou un service informatique peut réaliser, diminuant ainsi l'autonomie du centre de documentation à faire évoluer le système ;
- il est muni d'une ergonomie et d'un visualiseur peu adaptés à la consultation d'articles de périodiques numérisés ;
- il manque globalement de souplesse et de flexibilité ;
- il ne dispose d'aucun module de workflow autorisant un véritable travail collaboratif ;

Enfin, voici une série de **limites** que révèle l'analyse de Mobytext et Micromusée, et qui les rendent inaptes à la gestion et à la consultation de dossiers documentaires numériques.

• **Les limites de l'aide à la saisie**

- Elle est facilitée seulement par des fichiers d'autorité, exploités sous la forme de listes hiérarchisés ou de thésaurus ;
- Aucune aide n'est disponible sous la forme d'index ouvert automatiquement alimentés et

portant sur toutes les valeurs déjà saisies. Il faut alimenter manuellement l'index, ou importer d'autres thésaurus ;

- **Les limites de la recherche**

Le gros inconvénient de la recherche dans ces applications est qu'elle ne porte que sur des termes exacts, avec une gestion pénalisante de tables et de listes de référence. Et la possibilité d'importer des listes d'autorité ou des thésaurus provenant d'autres utilisateurs des logiciels Mobydoc n'apporterait rien dans le cadre du projet de dématérialisation d'un fonds documentaire spécialisé dans la publicité.

La seule possibilité de la recherche étendue passe donc par des listes organisées (termes associés et hiérarchie). Il n'existe aucune possibilité de faire une recherche libre multicritère.

En résumé, les limites de ces logiciels sont :

- pas de recherche libre, obligation de passer par des listes ;
- pas de recherche en texte intégral ;
- pas de recherche avec troncatures ;

3.5 L'analyse des besoins et les bénéfices attendus

Les logiques et les circonstances qui amènent à envisager une dématérialisation des dossiers documentaires au centre de documentation des musées des Arts Décoratifs peuvent se décrire en quatre points principaux : un besoin de place, une évolution de l'usage, un meilleur accès aux archives, un moyen de préservation et de conservation.

Il y a un évident **besoin de place**, puisque les rayonnages des documentations qui reçoivent les dossiers documentaires arrivent à saturation.

Une **évolution de l'usage** du fonds documentaire est souhaitée, à la fois par les documentalistes mais aussi par les consultants. Les documentalistes souhaitent fortement pouvoir disposer d'une indexation et d'un classement des documents plus fins, ne plus avoir à manipuler physiquement les documents papier via les boîtes de rangement, et pouvoir éventuellement dépouiller électroniquement les revues accessibles en version électronique.

Globalement, la dématérialisation est entrevue ici comme un processus permettant **un meilleur accès aux archives**. Concernant les consultants internes du centre de documentation – les conservateurs et leurs assistants - et ceux externes – le public invité à visiter le centre -, il est ainsi souhaité qu'ils puissent consulter le fonds électroniquement, via des postes de consultation, de façon plus autonome et en disposant d'outils de recherche modernes et performants. Ces évolutions ayant principalement pour objectif d'apporter des informations plus riches et plus détaillées aux utilisateurs.

Les dossiers documentaires ont actuellement pour fonction principale d'illustrer les collections du musée – et si l'on prend l'exemple des dossiers thématiques de la publicité autour desquels a été focalisée cette étude, on voit que ce fonds a plutôt un caractère actuel et contemporain -, mais ils pourraient dans le futur et avec le temps revêtir un caractère plus patrimonial. La dématérialisation est donc aussi envisagée pour ce projet comme un excellent **moyen de préservation et de conservation**.

Après avoir énoncé dans les grandes lignes les logiques et les circonstances qui motivent la dématérialisation, essayons maintenant d'identifier plus précisément tous les besoins du centre de documentation, et de cerner tous les bénéfices que l'on peut attendre d'un tel projet.

3.5.1 Amélioration du produit documentaire et du service aux utilisateurs

La dématérialisation, partout où elle se met en place, quels que soient la nature et la typologie des documents, quelle que soit l'organisation concernée, apporte une plus value et un progrès extrêmement intéressants. Le numérique offre en effet une très grande souplesse de traitement comparé à l'ère du papier, et les combinaisons possibles dans les traitements sont quasiment illimitées.

Dans le cas d'un centre de documentation au sein d'un musée, il n'est peut-être pas opportun de parler de gain de productivité, ni approprié d'essayer de quantifier ou chiffrer le bénéfice - même si, comme on le verra plus loin, il est toujours possible de définir des indicateurs assez précis et significatifs.

En revanche, sur un plan qualitatif, il est certain que la dématérialisation représenterait un pas en avant décisif au niveau :

- des conditions et du confort de travail des documentalistes ;
- de la qualité du produit documentaire ;
- du service proposé aux utilisateurs ;

Et on peut notamment citer les améliorations suivantes qui accompagnent la dématérialisation :

- des articles mieux classés et plus visibles ;
- des articles mieux décrits et mieux indexés ;
- une recherche beaucoup plus souple et performante ;
- aucune détérioration ou dégradation du support puisque le produit n'est plus manipulé mais seulement consulté numériquement ;
- aucun risque de vol ou de perte de documents (un document mal rangé est un document perdu) ;

3.5.2 Réduction de la consommation de papier

La documentation de la publicité ne tient aucune statistique sur le nombre de photocopies d'articles qui sont faites par mois dans le cadre de l'alimentation des dossiers thématiques. Aucun chiffre n'a donc été disponible sur ce point pour cette étude.

Le nombre de photocopies de documents qui sont faites par les consultants internes à partir des dossiers thématiques et dans le cadre de leurs travaux est difficilement quantifiable, lui aussi, puisqu'il peut être très variable d'un mois à l'autre.

Dans le cadre de cette étude, un petit sondage a été effectué auprès des différentes documentations du centre, sous forme de rencontres individuelles avec chacun des documentalistes responsables de ces documentations (cf. la partie 3.5). Ceci dans l'optique de faire une première évaluation et reconnaissance de la situation, et de jauger de l'opportunité d'une extension éventuelle de la dématérialisation à d'autres documentations du centre.

Et sur ce plan de la consommation du papier, une première constatation est apparue au fil de ces entretiens, et qui était commune à tous les départements : les conservateurs – premiers destinataires du travail des documentalistes - sont fortement attachés au papier. La manipulation physique des documents fait partie de leurs habitudes de travail. Et ils sont très accoutumés à la recherche parmi les boîtes et les chemises, puisqu'ils connaissent très bien le fonds et son organisation. De plus, beaucoup préfèrent lire les documents sur support papier plutôt que devant un écran. Ainsi, même s'il est tout à fait envisageable que les conservateurs adhèrent, à terme, à une recherche électronique dans une base de données, parce que convaincus qu'ils y gagnent en termes de vitesse d'accès et en termes de pertinence de résultats, il est en revanche certain que nombre d'entre eux souhaiteront posséder une impression de tous les résultats obtenus. Il est donc très probable que, si la dématérialisation avait lieu, se mettrait naturellement en place un système du style : « recherche sur ordinateur / lecture sur papier ». Il n'y aurait donc aucune réelle économie de papier ou d'encre à envisager comme bénéfice de la dématérialisation.

3.5.3 Gain de place

Concernant le stock des dossiers thématiques papier et l'éventualité d'une dématérialisation rétrospective, dans un premier temps il n'est pas prévu de détruire ces documents papier après leur numérisation. Une des solutions envisageables serait de descendre au fur et à mesure aux archives du musée les dossiers qui auraient été dématérialisés, afin de gagner de la place sur les étagères du centre de documentation. Dans l'hypothèse où tous les dossiers thématiques de la documentation de la publicité seraient intégrés dans la GED – autrement dit une dématérialisation rétrospective totale -, le gain de place s'élèverait à environ 6 m² au sol. Néanmoins, certains documents se trouvant dans un état de conservation – ou ayant une forme - non compatible avec une dématérialisation, il est envisagé de les laisser dans les dossiers papier sur les étagères, mais en procédant tout de même à l'intégration de leurs références dans la GED - indexation et classement électroniques. Le gain de place pourrait donc être minoré par rapport à l'évaluation faite plus haut, et il est aujourd'hui, en amont du processus, difficile à quantifier exactement.

Concernant la dématérialisation du flux entrant – autrement dit l'alimentation des dossiers thématiques au fil de l'eau après la mise en place du système de GED -, il est prévu naturellement de ne plus alimenter les dossiers papier. Seuls continueront à être stockés au format papier, de manière assez exceptionnelle, les documents non numérisables à cause de leur état ou de leur forme.

3.5.4 Gain de temps

Du côté du documentaliste, il faut essayer de quantifier le gain de temps éventuellement induit par l'utilisation d'un système de GED.

On peut le décomposer sur trois niveaux :

- lors de l'alimentation des dossiers ;
- lors du renseignement des consultants internes ;
- lors du renseignement du public extérieur ;

3.5.4.1 Alimentation des dossiers : lecture et dépouillement du périodique, numérisation physique des pages, découpage numérique de l'article, indexation des articles, métadonnées, classement et rangement des documents

Concernant l'alimentation des dossiers papier avec le système actuel, on dénombre quatre étapes :

- parcourir la revue et sélectionner les articles ;
- faire autant de photocopies que nécessaire (en plusieurs exemplaires si l'article recoupe plusieurs thèmes et qu'il doit être en conséquence classé simultanément dans plusieurs boîtes thématiques) ;
- inscrire manuellement les références de l'article sur la photocopie (la plupart du temps le titre du périodique, le numéro et la date de parution) ;
- accéder aux boîtes d'archives, puis aux chemises qu'elles contiennent, afin d'y classer les photocopies à l'endroit le plus pertinent (en suivant un ordre alphabétique, ou chronologique, s'il existe un classement de ce type dans les chemises, ce qui n'est pas toujours le cas) ;

Si on observe ces étapes du système papier actuel, et si on s'attache à comparer ce dernier avec le système électronique qui pourrait être érigé, on en déduit que :

- c'est au niveau de la quatrième étape (classement et rangement physique du document) que le système numérique pourrait permettre un gain de temps assez substantiel ;
- la troisième étape (indexation) pourrait être plus lourde et plus longue que dans le système papier ;

- **La lecture et le dépouillement du périodique**

Le parcours et le dépouillement d'un périodique sont des étapes intellectuelles inévitables et qui suivent la même logique que ce soit pour un système papier ou pour un système numérique.

- **La numérisation physique des pages**

Concernant la deuxième étape, celle de la numérisation physique du document, il n'est pas aisé d'essayer de faire une comparaison du temps passé, en termes de manipulation, entre une photocopie papier et une scannerisation de document. Pour ce qui est de la photocopie papier, le processus technique est généralement très simplifié et quasi-automatique (il se fait dans une qualité par défaut et sans aucun paramétrage précis, excepté celui du choix du format et de la couleur), et relativement rapide (quelques secondes au total). Concernant l'acte de numérisation d'un document par un scanner, le processus est un peu plus compliqué et lent. En effet, en fonction du document à dématérialiser, on peut décider d'appliquer certains traitements spécifiques (suppression des tâches, optimisation du contraste, résolution dpi élevée pour le document en sortie, traitement OCR, etc).

Néanmoins, sur ce plan de la numérisation physique, un élément important - et déjà relevé plus haut - est tout de même à prendre en considération : dans un futur plus ou moins proche, les documents natifs papier vont lentement disparaître au profit des documents électroniques - natifs numériques. L'économie de l'édition numérique est actuellement en pleine évolution et en pleine mutation, et la généralisation des périodiques de presse en version électronique paraît à moyen terme presque certaine - surtout dans les domaines de la communication et du marketing.

Cette disparition de l'édition papier sera donc synonyme de disparition du processus d'acquisition numérique avec le scanner - scannerisation. Le temps gagné par les documentalistes grâce à la suppression de cette étape - lourde en termes de temps et de manipulation - pourra éventuellement être consacré à l'optimisation de l'indexation - étape décisive et très importante pour valoriser et faciliter l'accès au document. Cela constitue un argument qui peut dès à présent convaincre d'adopter un système de GED au sein d'un centre de documentation.

On a vu que la documentation de la publicité procède déjà à un dépouillement électronique de deux revues disponibles à la vente en version électronique.

- **Le découpage numérique de l'article**

Dans le système actuellement en place, avec le format papier, lorsqu'un article est intéressant et sélectionné pour intégrer les dossiers thématiques, mais qu'il n'occupe pas l'intégralité

d'une page, on n'a pas d'autre choix que de photocopier la page entière, puisque les photocopieurs en service au centre de documentation ne sont pas dotés d'option de prénumérisation ni d'outil de découpage numérique. Comme tous les photocopieurs standards, ils gèrent seulement les formats classiques (A4, A3, etc) et le taux d'agrandissement / réduction.

Après l'opération de reprographie, il est donc nécessaire de désigner l'article choisi sur la photocopie – autrement dit le distinguer des autres -, ce qui est la plupart du temps effectué en inscrivant manuellement au crayon papier une petite croix, et plus rarement en surlignant le titre de l'article à l'aide d'un feutre de couleur.

Parfois, une page de revue peut contenir plusieurs articles intéressants et destinés à être stockés dans des dossiers thématiques différents. Dans ces cas-là, il faut donc faire autant de photocopies que de thèmes concernés, et identifier les articles sélectionnés sur chacune des photocopies.

Ainsi c'est la technique actuellement mise en place pour dépouiller tous les périodiques auxquels est abonnée la documentation publicitaire.

Malgré tout, le fonds comporte aussi une proportion non négligeable d'articles ayant été découpés aux ciseaux directement à partir d'autres revues ou journaux originales – par exemple, un article peut être découpé dans un journal quotidien comme Le Monde, de manière ponctuelle. Parfois, ces articles découpés aux ciseaux ont ensuite été photocopiés au format A4, mais d'autres fois ils ont été directement classés et rangés tels quel dans les boîtes – ce qui, on l'a vu, pourrait empêcher leur placement dans le chargeur d'un scanner lors de la dématérialisation.

Sur cet aspect de l'alimentation des dossiers, les possibilités d'une gestion électronique de documents seront donc très intéressantes. En effet, un logiciel de GED permet d'isoler l'article dès le processus de numérisation, à l'aide d'outils classiques de prénumérisation, de découpage et de recadrage. En définitive, contrairement au système papier, ce n'est donc pas la page entière qui sera stockée, mais seulement l'article sélectionné. L'utilisateur n'aura plus à chercher le signe distinctif d'un article sur une photocopie, mais ouvrira directement un fichier informatique contenant exclusivement l'article choisi. Cette option de découpage numérique est d'autant plus recommandée pour la dématérialisation qu'elle permettra de faire une économie importante de mémoire de stockage informatique – si on compare au cas où la page entière numérisée serait stockée sur les serveurs.

Sur le plan du gain potentiel pour la documentaliste, cette solution de découpage numérique est moins coûteuse en temps qu'un découpage physique du papier aux ciseaux, puisque

quelques clics suffisent. En revanche, cela nécessitera un petit temps d'apprentissage informatique de la part du documentaliste lors de la mise en place de la solution GED.

Et d'une manière générale, pour tous les utilisateurs, le fait d'isoler systématiquement l'article apportera un gain significatif sur le plan du confort et de l'ergonomie.

• **L'indexation des articles, les métadonnées**

La numérisation physique du document ne permet pas sa gestion et son traitement, c'est-à-dire : son analyse, son classement, son indexation. Des opérations annexes sont donc nécessaires, qui ont pour objectif de produire des identifiants et des métadonnées.

Pour ce projet de dématérialisation, il est prévu d'exploiter les deux grandes familles de techniques existantes pour décrire et classer les documents : l'indexation structurée, et l'indexation en texte intégral.

Avec le système d'indexation structurée de type base de données, la documentaliste renseignera les index et les mots clés de chaque article de périodique dans l'outil de GED.

Avec le système de GED en texte intégral, il sera possible d'auto-indexer les articles par tous les mots du texte qu'ils contiennent. La reconnaissance optique de caractères (OCR) fournissant le moyen d'extraire l'information textuelle de l'article scanné. Il est donc important de souligner encore une fois que tout le texte résultant de l'OCR est une métadonnée. Dans certains cas où la reconnaissance aura parfaitement fonctionné, l'intégralité du texte du document, c'est-à-dire tous ses mots, seront indexés.

Au final, on trouvera ainsi dans la GED des métadonnées qui seront :

- insérées manuellement et simultanément au processus de numérisation, grâce à un travail intellectuel d'analyse et de description du document (thesaurus, mots clés, descripteurs libres, listes d'autorité, etc) ;
- issues du document numérique lui-même par OCR (indexation en plein texte) ;
- insérées automatiquement (avec ou sans paramétrages préalables) au moment du processus de capture ou de reconnaissance de caractères (nom du fichier image, chemin de stockage, taille et type du fichier, date de création, etc) ;

Il est nécessaire aussi de s'interroger ici sur l'opportunité d'un recours à la technique de reconnaissance intelligente de caractères (ICR).

Tout d'abord, on rappelle que la technologie OCR permet à l'ordinateur de « lire » un texte, c'est-à-dire de transformer l'image d'un texte en un texte ASCII qui est ensuite indexable et réutilisable.

Mais il existe en fait un autre type de reconnaissance de caractères : l'ICR. Cette technique apporte des avantages supplémentaires dans la reconnaissance des formes non-textuelles d'une page : colonnes, illustrations ou tableaux. Certains systèmes peuvent automatiquement séparer le texte des graphiques et les stocker séparément. Pour le projet de dématérialisation des dossiers thématiques de la publicité, le recours à l'ICR ne semble pas véritablement indispensable. En effet, les articles de périodiques sont essentiellement constitués de textes, et de plus il n'est pas envisagé pour le traitement documentaire de ces articles de séparer les illustrations des textes.

Autre remarque concernant l'opportunité ou non d'un usage de l'ICR : dans les définitions qui en sont parfois données, il est dit que l'ICR peut être utilisé pour la reconnaissance de caractères manuscrits. Mais il est nécessaire de bien préciser dans quelles conditions exactes l'ICR peut aider à ce genre de reconnaissance.

Les systèmes actuels savent reconnaître l'écriture manuelle moyennant certaines contraintes :

- caractères séparés ;
- souvent majuscules et chiffres ;
- parfois caractères ou signes dans des cases ;

Or, dans le cadre du projet, ces contraintes ne sont pas respectées puisque des annotations manuscrites sont souvent présentes sur les photocopies des articles des dossiers thématiques. Ces annotations ont été rédigées par la documentaliste pour indiquer le titre, le numéro et la date de parution de la revue dont est extrait l'article, ainsi que parfois des informations importantes et complémentaires de celui-ci. Donc, sur ce point également, le recours à l'ICR dans le cadre de notre projet ne paraît pas adapté.

En précisant que l'OCR ne sera pas plus efficace à reconnaître ces caractères écrits à la main qui ne sont ni en majuscules, ni séparés.

• **Le classement et le rangement des documents**

Abordons maintenant la dernière phase de l'alimentation des dossiers : le classement, le tri, et le rangement physique des photocopies dans les boîtes. L'organisation actuelle des dossiers thématiques de la publicité est basée sur une logique de plan de classement thématique.

Le concept proposé par la majorité des éditeurs de logiciels de dématérialisation est une organisation similaire au classement papier (armoire, dossier, sous dossier, document) mais sous forme électronique. Et c'est véritablement dans cette dernière étape de l'alimentation des dossiers que la dématérialisation peut apporter un gain très significatif, en termes de puissance et de souplesse.

Premier avantage du numérique : contrairement au classement et au rangement d'une photocopie dans une boîte d'archive, le classement et le stockage d'un fichier dans la GED ne prend que quelques secondes, puisque l'accès aux répertoires et aux dossiers numériques se fait instantanément via un simple clic.

A cela il faut ajouter qu'un classement alphabétique ou chronologique à l'intérieur d'un dossier informatique peut être réalisé automatiquement, à l'aide d'un simple et unique paramétrage et sans une intervention systématique du documentaliste.

Ce premier avantage est d'autant plus significatif qu'un article se rattache parfois intellectuellement à plusieurs thèmes ou sous thèmes, c'est-à-dire à plusieurs dossiers – ainsi, le système papier actuel oblige parfois à faire plusieurs photocopies d'un même article et à les ranger dans toutes les boîtes d'archives concernées. Dans ces cas-là, le gain avec le numérique sera par conséquent multiplié par le nombre de dossiers dans lesquels on doit classer l'article.

Un autre avantage du système GED est la possibilité de pouvoir modifier ou de faire évoluer très facilement et rapidement un plan de classement électronique. Alors que la mise à jour et les corrections d'un plan de classement sont très pénalisantes au format papier, puisqu'elles demandent la consultation et la manipulation physique des boîtes, chemises, photocopies papiers. Ce qui peut s'avérer très délicat et très long lorsque les boîtes sont assez chargées.

Les avantages qui viennent d'être exposés amènent à la conclusion que, affranchi de la contrainte de la photocopie et du manque de place, le plan de classement pourrait être sans cesse optimisé, et encore plus approfondi. Et la documentaliste étant seule en charge de ce fonds et de ce plan, il est par conséquent évident qu'une dématérialisation l'aiderait considérablement pour ce travail.

• **Conclusion**

Le passage d'un système de dossiers papier à un système GED induit des opérations supplémentaires lors de l'indexation, qui viennent alourdir le travail du documentaliste lors de l'alimentation des dossiers documentaires. Actuellement, les articles sont organisés avec un simple plan de classement thématique, mais l'apport de l'informatique autorise une souplesse et une puissance supérieures dans l'indexation des documents, comme par exemple l'usage de mots clés. La numérisation s'accompagne donc d'un travail intellectuel supplémentaire, mais qui, en contre-partie, permettra ultérieurement un gros gain de temps et de performance lors de la recherche de documents.

Concernant l'éventualité d'un traitement rétrospectif du stock des dossiers thématiques - autrement dit la dématérialisation du fonds accumulé - la problématique de l'indexation est

similaire à celle d'une numérisation et intégration des documents dans la GED au fil de l'eau. La numérisation rétrospective des dossiers documentaires aurait vocation à les rendre plus visibles, à les valoriser et à faciliter leur exploitation, mais à la condition d'effectuer un complet travail de réindexation (thésaurus, mots clés, etc) afin d'exploiter au maximum le potentiel de la GED. Cette tâche intellectuelle serait sans aucun doute extrêmement lourde en termes d'investissement et de temps, d'autant qu'elle ne pourrait être réalisée que par un documentaliste spécialiste de ce fonds, sans possibilité d'externaliser ni de faire appel à des stagiaires.

3.5.4.2 Impact sur le temps passé à renseigner les consultants internes

Il n'est pas possible de quantifier précisément la fréquentation de la documentation de la publicité du centre de documentation par les consultants internes – essentiellement les conservateurs -, puisque aucune statistique sur ce point n'est réalisée. Ces consultants internes sont très habitués et familiers des dossiers, et ils ne demandent donc que très exceptionnellement de l'aide à la documentaliste. Cependant, même s'ils sont déjà très autonomes dans leurs recherches, on peut imaginer que l'accès à une recherche électronique performante, approfondie et avancée (sous forme de mots clé, de thésaurus, de recherche en texte intégral, etc) pourrait les rendre encore plus autonomes.

3.5.4.3 Impact sur le temps passé à renseigner le public extérieur

Concernant la fréquentation du centre de documentation par le public externe, principalement composé d'étudiants et de chercheurs, en 2008 on a relevé 478 visites. Il est intéressant d'essayer de savoir si la dématérialisation pourrait permettre de recevoir plus de visiteurs extérieurs, et sans que cela n'affecte la charge de travail des documentalistes. Actuellement, à l'inverse des conservateurs, le public qui consulte le centre de documentation a plus ou moins besoin d'une aide à la recherche de la part de la documentaliste. Cette aide ne prend pas seulement la forme d'indications sur l'organisation du plan de classement, mais passe par un véritable conseil et une analyse des besoins des visiteurs. Ce travail intellectuel ne pourrait bien sûr pas être pris en charge par des solutions informatiques, puisqu'il est inhérent à une véritable connaissance de la publicité et de la communication. En revanche, dans certains cas, notamment lorsque le visiteur a une bonne idée de ce qu'il recherche et qu'il n'a pas un réel besoin de conseil, il est certain qu'une consultation électronique - assistée du moteur de recherche intégré à la GED - pourrait éviter qu'il ne sollicite l'aide ou l'intervention de la documentaliste.

Par exemple, avec le système papier actuel, lorsqu'un visiteur ne trouve pas un document dans une boîte d'archives, il est conscient que le temps manque pour chercher massivement

dans d'autres boîtes, et il est donc amené à demander à la documentaliste s'il a une chance de trouver le document souhaité dans une autre boîte (et si oui laquelle ?). Alors que la possibilité d'une recherche globale dans tout le fonds numérisé, notamment avec la fonctionnalité de recherche en texte intégral, pourrait satisfaire le visiteur dans nombre de situations et sans aucune intervention de la documentaliste. Sans négliger l'aspect humain de la relation de conseil et de service, il apparaît donc que la dématérialisation offrirait de nouvelles perspectives pour accueillir un plus grand nombre de visiteurs, tout en élargissant le profil de ce public.

Néanmoins, des obstacles pourraient venir se dresser et compliquer une telle démarche d'ouverture de la part du centre de documentation.

Tout d'abord, les postes de consultation informatique qui sont à disposition du public se trouvent dans une salle spécifique – appelée salle de consultation. A l'intérieur de cette salle se trouve également des ouvrages et des revues, qui doivent donc faire l'objet d'une surveillance régulière lorsque des visiteurs consultent les ordinateurs. Compte tenu des effectifs réduits du personnel du centre de documentation, et justement parce que la mise en place de la GED a vocation à libérer du temps aux documentalistes, il paraît difficile de demander à ces derniers d'organiser une surveillance régulière et alternée de cette salle de consultation, dans l'hypothèse où sa fréquentation connaîtrait une nette hausse.

Ensuite se pose un autre problème : celui des impressions éventuellement demandées par les visiteurs. Pour des raisons juridiques, liées au droit de reproduction et au droit d'auteur, il est exclu d'autoriser le visiteur à faire des copies électroniques des documents, ni même de lui transmettre par mail le moindre document numérisé. Les aspects juridiques de ce projet de dématérialisation, et notamment la question du droit d'impression à partir d'un panorama de presse électronique – panoramas qui composent majoritairement les dossiers thématiques des dossiers de la publicité – seront traités en détail dans la partie 3.9 de cette étude.

Dans l'hypothèse où les clauses juridiques du projet donneraient la possibilité aux visiteurs d'obtenir des impressions des documents stockés dans la GED – on verra plus loin que cela reste très improbable – deux stratégies sont possibles :

- Le visiteur serait autorisé à déclencher les impressions de façon autonome à partir du poste de consultation. Dans ce cas, faudrait-il mettre en place un outil logiciel pour comptabiliser le nombre d'impressions effectuées par le visiteur, ou plutôt continuer à fonctionner sur un mode de confiance comme c'est le cas actuellement ?
- Le visiteur ne serait pas autorisé à déclencher les impressions de façon autonome, système qui paraît le plus sûr et le plus probable. Dans ce cas, la documentaliste devra prendre sur

son temps pour réaliser les impressions, et la dématérialisation ne lui apportera sur ce point précis aucun bénéfice.

En conclusion, sur le plan strictement de l'accueil du public externe, il est très difficile de se prononcer sur l'opportunité et la pertinence de dématérialiser en termes de réduction de charge de travail pour les documentalistes. En fait, la réponse est aussi conditionnée par la volonté et la politique que le centre souhaite mettre en place, étant entendu que le fonds documentaire accumulé est prestigieux, rare, et peut être un peu sous exploité au regard de sa valeur.

3.5.5 La pérennité du fonds numérisé, et la question du stockage physique des données

Tout d'abord, avant de traiter de la stratégie et des choix éventuels à opérer en termes de stockage informatique, il faut bien entendu souligner le fait que les données qui auront été numérisées par le musée ne devront pas pouvoir par la suite être modifiées ou détruites accidentellement par une opération logicielle. La solution de GED adoptée devra donc être capable de mettre en place des identifications d'accès et la création des profils utilisateurs.

Ensuite, dans toute étude préalable à un projet de dématérialisation, il faut s'attacher à bien analyser la question de la « pérennité des supports de stockage ». Cet aspect du projet est sensible et relève de compétences techniques. Il devra par conséquent être traité et géré par la direction des systèmes informatiques. Il est également nécessaire de souligner que cette question de la pérennité des supports n'est pas directement liée à celle du choix de la solution de GED, puisque les applications actuellement proposées sur le marché de la dématérialisation sont compatibles avec les principaux supports de stockage existant.

Le choix d'un support est important dans la mesure où il conditionne la vitesse à laquelle on peut accéder aux documents, ainsi que leur condition de conservation. Pour concilier ces deux impératifs, il est en général indispensable de combiner les supports optiques et magnétiques.

- **Stockage sur support magnétique**

Compte tenu de l'espace mémoire très important que pourrait représenter le corpus numérique des dossiers thématiques, les supports magnétiques envisagés pour ce projet sont les disques durs et les bandes.

Les **disques durs** assureront la **consultation on-line** de la base de données documentaire du système GED – assurant une vitesse élevée de transmission et un grand espace mémoire. Ils

seront montés sur des **systèmes RAID**, eux-mêmes implantés sur des serveurs Web. Ces derniers pourront alors offrir de très grands espaces mémoires, de l'ordre de plusieurs téraoctets. Actuellement, c'est ce type de système qui est adopté par le service informatique des Arts Décoratifs pour l'exploitation du fonds numérisé des catalogues et des collections – consultables, pour le moment, uniquement à partir de Micromusée ou de Mobytext. Le musée est en effet doté d'un système RAID 5 qui contribue à assurer, en partie du moins, la pérennité de ce fonds numérisé. Mais soulignons que cette garantie de sécurité et de fiabilité apportée par le système RAID ne concerne que les serveurs dits de « production », c'est-à-dire ceux destinés à l'exploitation régulière du fonds (requêtes, interrogations, modifications, consultations, etc). Ce sont ces machines qui sont sollicitées lors de toute utilisation de la base de données à partir d'un poste de consultation. Or, la problématique des serveurs de production du musée doit être séparée de celle de l'archivage sur le long terme - et éventuellement de la mise en place d'un véritable SAE (système d'archivage électronique) répondant à la norme NF Z42013 sur l'archivage numérique.

Les **bandes magnétiques** sont, quant à elles, préconisées ici essentiellement pour programmer des **sauvegardes systématiques et périodiques des données à moyen terme**, - autrement dit des « back up ». Mais ces opérations régulières de sauvegarde appartiennent, là encore, à une problématique différente de celle de l'archivage numérique sur le long terme et déjà abordée plus haut. Actuellement, le dispositif informatique du musée n'est pas doté de cette technologie de bandes magnétiques. Néanmoins, des opérations de sauvegarde sont régulièrement menées sur des serveurs - donc sur la base de disques durs - de réplication - les Arts Décoratifs disposent précisément de deux serveurs indépendants du serveur de production et dédiés à la sauvegarde. Mais, conscient que les serveurs à disques durs n'ont pas vocation à la sauvegarde des données à moyen terme, il est prévu que les Arts Décoratifs se dotent très prochainement de bandes magnétiques de type LTO. Il faut souligner que les bandes magnétiques ont en outre le grand avantage d'être amovibles, elles peuvent donc être stockées par précaution dans un lieu physique extérieur au musée, ou être déplacées en urgence en cas d'incendie par exemple.

En conclusion, il est très important de rappeler encore une fois que ni les disques durs ni les bandes ne sont en principe employés pour archiver des corpus numériques sur le long terme, compte tenu de leur fragilité. Ces deux supports sont par nature réinscriptibles et n'offrent pas de garanties suffisantes pour la conservation des données. Cette dernière est en général plutôt assurée par des supports optiques.

- **Conservation sur support optique**

Pour ce projet, il est préconisé d'investir dans l'achat de DON WORM pour archiver. C'est également ce que dit la norme NF Z42013 sur l'archivage numérique à propos des processus et des matériels qui doivent être utilisés. Cette norme précise, dans son paragraphe 3.7, que le support de stockage adopté doit être un « support optique pour lequel l'écriture des bits codant les données se fait par transformation irréversible d'un ou plusieurs constituants de ce support ». Elle désigne donc clairement les supports optiques de type WORM.

Dans le cadre de la réalisation de cette étude, des entretiens avec le directeur informatique ont eu lieu. Et il m'a paru très important d'y aborder avec lui cette question de l'archivage du fonds numérisé « sur le long terme ». A ce propos, il ne voit pas réellement l'utilité, disons plutôt l'obligation, de procéder à un stockage sur support optique de type WORM. En effet, il considère que le stockage des données sur les serveurs - avec tout de même une double réplication - des Arts Décoratifs, et bientôt sur des bandes magnétiques extractibles, serait suffisant. Cette position est naturellement défendable, néanmoins il me semble que, si les serveurs ou les bandes présentent des garanties à peu près suffisantes pour stocker et consulter des documents numériques à moyen terme, les disques optiques gardent eux toute leur pertinence pour l'archivage des corpus figés à très long terme. Dans cette perspective, cette étude préconise donc un archivage périodique des dossiers documentaires dématérialisés sur des supports de type DON WORM. Il restera à déterminer la durée entre deux opérations d'archivage sur support optique, en fonction de la vitesse d'alimentation des dossiers mais surtout des moyens - en termes de temps et d'argent - informatiques disponibles.

Enfin, sur le plan de la pérennité des supports numériques, on peut citer comme exemple la stratégie qui a été adoptée par l'INA³². Ce dernier a prévu de recopier ses stocks de données sur de nouveaux supports tous les 10-15 ans, afin d'anticiper et de prévenir tout début de dégradation physique des supports, et qui pourraient amener à la perte de documents archivés.

3.5.6 La gestion des processus, ou « workflow »

La gestion des processus, ou « workflow », est un aspect stratégique de la GED.

Mais qu'en est-il réellement des besoins du centre de documentation sur ce point ?

³² Institut National de l'Audiovisuel, www.ina.fr

Pour répondre à cette question, il convient :

- d'analyser le flux des tâches ou des actions successives à réaliser pour dématérialiser les documents, et d'identifier les différentes personnes intervenants dans ces actions ;
- de définir, si elles existent, les grandes étapes de la vie d'un document numérisé (validation, modification, publication, etc) ;

Concernant les dossiers thématiques de la documentation publicitaire, et pour le traitement au fil de l'eau, on a vu que toutes les étapes de la dématérialisation d'un document seront prises en charge par le seul documentaliste responsable du fonds : acquisition numérique, indexation électronique, classement électronique, stockage numérique et publication.

Par exemple, les traitements éventuellement nécessaires et préalables à son intégration dans la GED, comme l'amélioration colorimétrique, le recadrage, etc, seraient tous effectués en une seule fois et de manière définitive par la documentaliste. Sans que cela ne nécessite l'intervention ultérieure d'une autre personne.

A priori, ni les conservateurs, ni les assistants conservateurs, ne seront amenés à numériser, archiver, modifier ou publier eux-mêmes un document. Leur utilisation de la GED se limitera donc aux recherches et aux consultations.

Une fois le document intégré et stocké dans la base de la GED, il n'est prévu ou planifié aucune modification sur ce dernier – en tout cas si l'on se calque sur le système papier actuel où les photocopies sont seulement consultées, et jamais modifiées ni reclassées.

Ainsi, concernant la numérisation et la gestion des dossiers documentaires, et au regard des besoins et des méthodes de travail actuels, il apparaît que l'intérêt de la GED pour le centre de documentation ne résiderait donc pas fondamentalement dans la nécessité de mettre en place un outil de gestion électronique des processus qui « orchestrerait la vie d'un document numérisé », c'est-à-dire un enchaînement de tâches initié par un élément déclencheur. Dans l'organisation actuelle du travail à la documentation publicitaire, il n'existe pas de véritables « étapes de décision », de temporisation ou de conditions qui nécessitent et entraînent la circulation organisée et orchestrée - de façon transverse aux structures et fonctions du centre de documentation - des documents stockés dans les dossiers documentaires.

Par exemple, aucune des fonctions suivantes :

- spécifications des procédures ;
- filtres informationnels ;
- gestion de la durée des échéances ;
- retour d'informations ou alertes ;

n'est véritablement requise pour le projet.

Ceci étant posé, de telles fonctions sont pourtant disponibles dans la majorité des logiciels de GED actuellement sur le marché, y compris les moins puissantes.

Cette étude, même si elle reste centrée autour des dossiers documentaires de la documentation publicitaire, a aussi pour objectif d'essayer d'anticiper un futur élargissement de la dématérialisation à tout le centre de documentation.

Dans cette perspective, les possibilités d'un module de workflow dans l'outil de GED pourraient faire naître dans le futur de nouvelles formes de collaboration entre les documentations, et permettre quelques évolutions dans les méthodes de travail.

Par exemple, on peut imaginer une possible forme de mutualisation des abonnements à toutes les documentations du centre. La réception d'un numéro de périodique au format PDF par une documentation pourrait faire l'objet d'une copie / renvoi automatique, ou bien d'un simple partage en ligne, aux autres documentations concernées par ce périodique.

Autre piste pour l'exploitation d'un module de workflow : celui d'un partage automatique des documents qui ont trait à des études ou à des recherches communes à plusieurs documentations. Les filtres informationnels, et les alertes, pourraient ainsi permettre l'envoi automatique d'une copie ou d'un signal à tous les documentalistes concernés, chaque fois qu'un document injecté dans la GED est susceptible de les intéresser. Cette programmation de l'automatisation pourrait se faire à l'aide de listes de mots clés établis par chacun des documentalistes.

Enfin, on peut aussi soumettre l'idée que, dès la mise en activité d'un système de GED, la fonction de retour d'informations d'un module de workflow pourrait permettre de tenir des statistiques sur les taux de consultation des documents. Ceci permettrait notamment de connaître quels sont les documents les plus consultés – par les conservateurs ou par le public – et quels sont les départements et les documentations les plus utilisateurs de la GED.

3.5.7 L'abolition des contraintes spatiales, et de nouvelles formes de collaborations et d'échanges entre les conservateurs et les documentalistes

Un des trois concepts fondateurs de la dématérialisation est : améliorer la qualité et la rapidité des échanges entre personnes par l'utilisation de moyens de communication, tels que le courriel ou les espaces de travail partagés. Pour ce projet, l'exploitation d'une GED pourrait en effet faciliter les échanges d'informations et apporter de nouvelles possibilités de collaborations entre les documentalistes et les conservateurs. Les documents étant dématérialisés et en ligne, on peut par exemple imaginer que les conservateurs puissent les consulter depuis leur poste informatique, et sans se déplacer physiquement dans les documentations concernées - sous réserve que les Arts Décoratifs aient financièrement les

moyens de déployer un outil de GED sur tous les postes des conservateurs. Les conservateurs pourraient ainsi formuler leurs demandes d'informations par mail ou par téléphone, et les réponses des documentalistes pourraient prendre la forme de références électroniques dans la base de la GED - par exemple des hyperliens si la GED possède une interface web. Bénéficiant d'une nouvelle forme de réactivité induite par l'outil informatique, les échanges et les collaborations entre la documentation et la conservation seraient potentiellement plus riches et plus développés. D'une manière générale, dans toutes les organisations où ont été menés des projets de dématérialisation, de nouvelles méthodes de travail ont pu éclore à court ou à moyen terme.

3.5.8 L'opportunité d'un désherbage et d'une redécouverte du fonds documentaire

La dématérialisation est aussi à envisager comme une période pouvant être propice et adéquate non seulement au désherbage mais aussi à la redécouverte du fonds.

En effet, le processus engagé entraînera forcément :

- L'ouverture systématique de toutes les boîtes d'archives, alors que certaines ne le sont finalement que très exceptionnellement. Cette exploitation relativement inégale des dossiers du fonds ne donne pas en fait l'occasion ni la possibilité à la documentaliste d'évaluer le fonds dans sa globalité, ni de désherber.
- La nécessité de parcourir et redécouvrir rapidement chaque article archivé dans les boîtes, condition indispensable pour procéder à l'indexation électronique. En effet, on rappelle que le système papier actuel ne propose qu'une organisation documentaire sur la base d'un plan de classement thématique, sans aucune indexation de type de mots-clés ou thésaurus, exceptés pour les articles issus du dépouillement électronique des revues disponibles en version électronique et indexés dans Mobytext.

Il est important de souligner, dans le cadre d'une organisation précise et d'une planification de la dématérialisation, que ce désherbage éventuel devrait forcément être réalisé par la documentaliste responsable de ce fonds spécifique à la publicité, ce dernier pouvant difficilement être appréhendé par un non-spécialiste.

3.5.9 Garantir la sécurité de la partie du fonds documentaire ayant une très grande valeur de rareté

Lors des entretiens, certains documentalistes ont clairement émis l'avis qu'il serait intéressant et judicieux de numériser la partie du fonds – pas seulement des dossiers documentaires mais aussi certaines revues - qui possède une valeur patrimoniale et

historique particulière et très importante. Par exemple, pour la documentation de la publicité, la numérisation de la revue *Vendre* est fortement souhaitée. Les documentalistes voient donc aussi dans la numérisation une opportunité de pérenniser la partie du fonds ayant une très grande valeur de rareté, en parant ainsi à tout accident ou toute dégradation naturelle susceptible de le faire disparaître.

3.6 Comptes rendus des entretiens avec les documentalistes des autres documentations

Des entretiens ont eu lieu avec les documentalistes responsables des différentes documentations du centre, au sujet de la faisabilité et de l'opportunité d'une dématérialisation d'une partie du fonds de ces documentations. Ces entretiens, en plus d'apporter des éléments concrets sur la nature, le volume, la forme et l'état de conservation des fonds archivés dans le centre, furent très enrichissants puisqu'ils ont notamment permis l'expression de certaines interrogations, résistances ou réticences vis-à-vis d'une numérisation et d'une mise en place d'un outil de GED. En effet, les discussions avec certaines documentalistes ont dévoilé les différentes approches, appréhensions et attentes face à l'arrivée du document électronique au sein d'un organisme culturellement et historiquement attaché au papier, et, plus généralement, à l'aspect des documents. Leurs témoignages ont également permis d'obtenir certains échos sur les mêmes appréhensions vis-à-vis de l'outil informatique de la part de certains conservateurs ou assistants. Il est essentiel de recueillir de tels échos puisque ces personnels du musée seraient, en cas de mise en place d'une GED comme outil de consultation pour les dossiers du centre, les premiers utilisateurs et donc les premiers concernés.

Voici donc une synthèse des éléments les plus significatifs qui sont ressortis de ces entretiens.

3.6.1 Compte rendu de l'entretien avec la documentaliste responsable de la documentation de la mode et du textile

Le fonds documentaire de la documentation de la mode et du textile comprend environ 1300 dossiers, pour la plupart consacrés aux couturiers et aux créateurs.

- **Principes d'usage, objectifs et évolutions souhaitées**

L'idée et le souhait principaux de la documentaliste, en vue de la numérisation, est de rassembler tous les documents relatifs à un artiste ou à une collection dans un seul dossier, et ce afin que tout soit disponible d'un seul coup, avec une seule et unique recherche. A noter que les dossiers de cette documentation ne comprennent pas de sommaires d'articles renvoyant à des revues ou à d'autres dossiers, contrairement à la publicité où l'on trouve des feuilles manuscrites qui sont des sommaires ou des listes de références d'articles renvoyant à des revues disponibles dans le centre de documentation.

Le plan de classement des dossiers été élaboré au fur et à mesure par des personnes

différentes, puisque sont intervenus sur ce fonds de nombreux stagiaires qui ont été amenés à prendre des initiatives. Et même si ce plan semble plutôt pertinent et efficace aux dires de la documentaliste, il n'en reste pas moins qu'il manque forcément d'un peu de cohésion. Le vocabulaire employé par les stagiaires était parfois personnel, et il reste donc parfois un peu opaque aux documentalistes, qui ne savent pas toujours eux-mêmes ce que contiennent les dossiers. Remarque et réflexion qui ont été réitérées par d'autres documentations au cours des entretiens. Certains dossiers ont quant à eux été carrément désorganisés, et le plan de classement d'origine bouleversé, suite à des travaux effectués par des stagiaires. La refonte et l'amélioration du plan de classement sont donc des objectifs de la documentaliste. Ces derniers pourraient être réalisés uniquement avec l'aide de la GED, vu la charge de travail énorme que demanderait un reclassement papier, et vu l'emploi du temps extrêmement serré de la documentaliste.

- **Consultants, visiteurs, conditions d'accès**

La documentaliste est moins restrictive que ces autres collègues concernant le public autorisé à consulter ce fonds. Ainsi, il arrive même que certaines personnes non professionnelles consultent le fonds. Néanmoins, le critère de sélection consiste tout de même globalement à ne laisser consulter que les personnes bloquées dans leurs recherches, dans l'impasse, et ne pouvant satisfaire leurs recherches que dans le centre de documentation des musées des Arts décoratifs. Le fonds sur la mode et le textile étant unique en Europe, si ce n'est dans le monde, le centre de documentation est souvent la dernière alternative pour trouver ce qu'ils cherchent. La documentation de la mode et textile reçoit ainsi plusieurs centaines de personnes par an. Concernant le public en interne, il se compose essentiellement des 3 conservateurs pour la mode, et des 3 assistants de conservation. En moyenne, ils consultent le centre de documentation une fois par semaine.

- **Problématique, enjeu et gain éventuel d'une dématérialisation**

Après avoir découvert et évalué, notamment au travers de voyages à l'étranger, d'autres fonds portant sur la mode et le textile, la documentaliste a pris conscience du caractère unique et exceptionnel de sa documentation aux Arts Décoratifs. Elle voit donc là une évidente opportunité d'exploiter commercialement ce fonds. Reste à savoir sous quelle forme et à quelles conditions ? Selon elle, de nombreuses personnes dans le monde souhaiteraient consulter leurs dossiers documentaires, et éventuellement en dupliquer une partie. Ils seraient prêts pour cela à déboursier de fortes sommes d'argent, qui pourraient venir amortir le processus de dématérialisation.

Cela étant posé, il convient d'aborder la question épineuse et pourtant essentielle du droit

d'auteur et de la propriété intellectuelle. Question qui se répète pour toutes les autres documentations du centre qui seraient intéressés et volontaires pour mettre une partie de leurs dossiers documentaires en ligne sur internet. Sur ces questions de l'opportunité d'une exploitation commerciale de leur fonds, qui passe là aussi forcément par une numérisation et une mise en ligne des documents, les documentations de la publicité et de la mode sont à l'unisson. L'explication étant que ce sont les deux documentations les plus largement sollicités et contactés par le public extérieur.

Néanmoins, malgré tout cet intérêt et cette motivation, vu la quantité de documents concernés et le potentiel d'exploitation, ce travail d'évaluation et de négociation juridiques nécessiterait l'emploi à temps plein d'une personne qualifiée et compétente dans ce domaine. De même, une étude préalable et à titre purement expérimentale, portant par exemple sur un échantillon de documents représentatifs - un catalogue de collections, une photo, etc -, devrait être menée afin de sonder et d'interroger les éventuels détenteurs des droits. La documentaliste responsable du fonds de la mode et du textile, en place depuis 5 ans, manque de temps et n'a jamais pu véritablement commencer ce travail de prospection et de sondage, ni faire l'ébauche d'un cahier des charges. Elle est submergée par les tâches quotidiennes, ce qui l'amène à sans cesse repousser cette étude qu'elle souhaiterait ensuite transmettre à sa direction.

Lors de la préparation des expositions, les assistants consultent beaucoup le centre de documentation, et réalisent de véritables « dossiers de préparation » constitués principalement de photocopies d'originaux - revues, catalogues, livres, look book. Ces photocopies sont souvent en couleurs, et les assistants ont l'habitude de travailler avec des dossiers papiers parfois très volumineux. La préparation d'une exposition peut mener à la constitution de dossiers de plus de 1000 pages. Ces dossiers sont ensuite de véritables mines d'or pour les consultants extérieurs qui voient ainsi toute la genèse de l'exposition et tout le travail important de documentation effectué par les assistants. D'après la documentaliste, la numérisation de ces dossiers et leur exploitation commerciale pourraient également faire l'objet d'un projet très séduisant.

Enfin, contrairement à la documentation de la publicité, les dossiers thématiques sont moins consultés que les dossiers dédiés aux créateurs ou aux couturiers – dits « dossiers d'artistes ». C'est un exemple qui vient rappeler que les priorités et les stratégies, en termes de planning de dématérialisation, pourraient donc sensiblement varier selon les documentations.

Toujours en termes de priorité, une initiative serait également à considérer, qui plus est commune celle-ci à toutes les documentations : la numérisation et la mise en ligne d'une

partie des dossiers d'œuvres. Ces dossiers sont la propriété du musée et pourraient donc être librement exploités, du moins une partie des documents compris dans ces dossiers - certains étant tout de même d'ordre strictement confidentiel. Cette mise en ligne pourrait évidemment intéresser beaucoup d'acheteurs en France ou à l'étranger. Ainsi, dans le cas de ventes d'objets similaires ou proches de ces œuvres, un acheteur pourrait saisir l'opportunité de consulter en ligne un dossier d'œuvre afin de mieux estimer son futur achat.

Une suggestion a également été faite – elle aussi relayée par d'autres documentalistes -, à propos des revues très anciennes, et qui sont donc libres de droit - comme les revues du 19^{ème}.

Autre proposition, la salle des compactus du centre de documentation contient aussi des ressources d'informations secondaires, entièrement élaborées par des personnels des Arts Décoratifs. Ces documents pourraient ainsi faire l'objet d'une dématérialisation libre de droits. Par exemple, une armoire contient des fiches bibliographiques, classées par ordre alphabétique, portant sur des marques ou des créateurs, et renvoyant à des revues de mode. Ces fiches manuscrites pourraient faire l'objet d'une retranscription sur ordinateur et être mises en ligne, afin de renseigner un public désireux de retrouver la trace de certains articles mais ne sachant pas comment s'y prendre. On trouve aussi un très intéressant « dictionnaire de la mode », lui aussi entièrement élaboré par les personnels du musée, et qui pourrait faire l'objet d'une mise en ligne.

• Conclusion

Exceptée la documentation de la publicité – et sur laquelle porte plus spécifiquement cette étude -, il est ressorti des entretiens que la documentaliste la plus optimiste et la plus motivée pour ce projet de dématérialisation était celle de la documentation de la mode et du textile.

Cependant, malgré un très vif enthousiasme, elle reste toutefois interrogative et dans l'expectative quant aux moyens et à l'engagement qu'engagerait le musée dans l'hypothèse du lancement du projet. Sachant qu'elle est toute seule pour gérer toute la documentation, et qu'il n'existe même pas actuellement de référencement complet des livres de la documentation mode et textile dans Mobytext. Elle pense donc que le principal obstacle à la dématérialisation est le manque de moyens humains. Selon elle, l'indexation et le classement électroniques ne peuvent clairement pas être pris en charge par une seule personne, vu le volume du fonds.

Le premier avantage qu'elle voit dans la dématérialisation est celui d'une voie possible à la rentabilité de sa documentation. Comme deuxième avantage, elle y voit aussi un gain de place qui pourrait être crucial dans le futur.

Enfin, sur le plan de l'usage, elle semble en complet accord avec les autres documentalistes sur le fait que les consultants n'auraient pas une réelle préférence à consulter les documents en version électronique plutôt qu'en version papier – ce serait même plutôt l'inverse d'après elle.

3.6.2 Compte rendu de l'entretien avec la documentaliste responsable de la documentation du 20^{ème} siècle (le design, les créateurs, le bijou, la céramique)

Le fonds de cette documentation est déjà très important, malgré qu'il n'ait été constitué que depuis les années 90. Il comprend aujourd'hui plus de 1500 dossiers, soit plusieurs centaines de milliers de pages de documents.

- **Principes d'usage**

Ils sont assez similaires à ceux de la documentation de la mode et du textile. Le souhait de cette documentation est aussi de pouvoir retrouver tous les documents numérisés relatifs à un artiste ou à un objet, d'un seul coup, après une seule et même recherche.

Les dossiers de cette documentation ne comprennent pas de sommaires d'articles renvoyant à des revues ou à d'autres dossiers. Mais, par exemple, lorsque les articles concernant un artiste sont trop longs - plus d'une dizaine de pages -, ils ne sont pas photocopiés et stockés, mais seulement référencés dans Mobytext - via une indexation assez fine. La dématérialisation permettrait ainsi de contourner ce genre de contrainte liée au volume, et l'on voit bien, avec cet exemple, à la fois les limites de l'archivage papier mais aussi celles d'un outil documentaire comme Mobytext - destiné à un simple catalogage et non à la GED.

- **Conditions d'accès et publics**

La documentaliste est très restrictive concernant le public autorisé à consulter ce fonds. Seulement une soixantaine de personnes externes au musée environ consultent le fonds chaque année. Cette documentation est, elle aussi, gérée par une seule documentaliste, qui ne peut donc pas élargir son public. D'une manière générale, les demandes de consultation sont renvoyées vers la bibliothèque qui dispose d'un fonds d'ouvrages important et d'une photocopieuse. Les consultants internes sont essentiellement les personnels du service des publics, les conservateurs du département, les personnels du service de presse ainsi que la directrice scientifique.

- **Problématique, enjeu et gain éventuel d'une dématérialisation**

Cet entretien fut réellement très intéressant et très instructif. En effet, son approche et son

avis, à la fois concernant l'opportunité et l'intérêt d'une numérisation au centre de documentation mais aussi la faisabilité de celle-ci, sont très clairs et bien tranchés. Plus particulièrement, sur la question des moyens humains – le manque d'effectifs -, et à propos des solutions envisagées, elle a semblé très lucide et s'est montrée d'une grande franchise.

Tout d'abord, comme critère le plus fondamental, elle met en avant et souligne le travail très lourd et vraiment imposant qui devrait accompagner une dématérialisation rétrospective du stock des dossiers. Et ce même si cette dernière se faisait « à petits pas » et de manière partielle. Les moyens humains faisant selon elle cruellement défaut, et étant seule en charge de cette documentation, elle considère comme impossible et irréaliste le fait de devoir consacrer une ou deux journées par semaine à la dématérialisation rétrospective. Et à propos des solutions alternatives, sur le plan du recrutement, qui pourraient être envisagées par le centre de documentation pour dématérialiser massivement, elle s'est montrée plutôt réservée et sceptique. Ainsi, le recrutement de personnels en travail temporaire et extérieurs au musée ne semble pas avoir sa faveur ni son approbation.

Dans le cadre d'une dématérialisation rétrospective et massive, les deux solutions pour recruter des personnels extérieurs au musée sont les suivantes :

- une première solution résiderait dans l'engagement d'un ou plusieurs stagiaires pour des périodes de plusieurs mois. Ces stagiaires ne seraient pas indemnisés pour ces travaux pourtant répétitifs et fastidieux. La documentaliste a clairement exprimé le fait que, d'une manière générale, cette option consistant à faire appel de façon répétée à des stagiaires non indemnisés ne lui convenait pas sur un plan déontologique ;
- la deuxième solution consisterait à faire appel à des personnels employés dans le cadre de CDD, ce qui lui paraît plus qu'improbable vu le très faible budget alloué au centre de documentation ;

L'avis de la documentaliste sur la stratégie à employer pour ce projet est le suivant : si on optait pour une numérisation partielle des dossiers documentaires, elle considère que la meilleure solution serait de numériser à partir des revues originales et non des photocopies déjà stockées. Ce qui permettrait, selon elle, une numérisation de meilleure qualité avec notamment une indexation en texte intégral plus fiable. Concernant sa documentation, les revues prédominantes et fondamentales dans ses dossiers seraient « Intramuros » et « Crafts », qu'il faudrait donc numériser en priorité.

Sur le plan du gain potentiel pour les utilisateurs, la documentaliste n'est pas persuadée que la dématérialisation apporterait un grand bénéfice et un meilleur confort aux consultants internes. Elle considère que les usages et les habitudes sont déjà en place depuis très

longtemps, et que les consultants internes connaissent assez bien le fonds et la disposition des dossiers. Et l'abolition des contraintes spatiales ne serait pas quelque chose susceptible de franchement les séduire. Elle estime aussi que l'accès à une version numérique des documents ne serait pas forcément synonyme d'une meilleure visibilité ou d'une meilleure circulation de l'information – notamment entre les conservateurs et les documentalistes.

Elle semble en revanche convaincue de l'intérêt de dématérialiser pour libérer de la place sur les étagères.

- **Conclusion**

Globalement, la documentaliste de cette documentation ne considère pas la dématérialisation comme étant une étape actuellement indispensable pour le centre de documentation. Et surtout, cette opération lui paraît irréalisable au regard des moyens humains et budgétaires aujourd'hui à sa disposition.

3.6.3 Compte rendu de l'entretien avec la documentaliste responsable de la documentation Art nouveau Art déco

Le fonds des dossiers documentaires de la documentation Art nouveau Art Déco porte sur les oeuvres du 19^{ème} siècle et aussi beaucoup sur l'exposition universelle des Arts décoratifs à Paris en 1925.

Les seuls consultants internes sont le conservateur du département et son assistant, ainsi que des conférenciers mais de façon plus sporadique. Quelques visiteurs externes sont accueillis, mais cela reste assez exceptionnel. Ils sont pour la plupart redirigés vers la bibliothèque. En effet, la documentaliste s'attache à renseigner les chercheurs sur les œuvres des collections, alors que la bibliothèque peut répondre à toutes les recherches généralistes sur les Arts décoratifs.

La documentaliste ne semble pas très enthousiaste au sujet d'une éventuelle numérisation. Elle est très attachée au papier et ne semble pas véritablement prête pour un passage à la documentation électronique. Elle imprime par exemple systématiquement pour lire ou consulter, et estime que les écrans - plus particulièrement ceux qui équipent le musée - ne sont pas vraiment adaptés à la consultation mais plutôt à la saisie en base et pour le référencement. Ainsi, alors que le musée est abonné aux versions électroniques et papier de La Gazette de l'hôtel Drouot, elle consulte toujours systématiquement la version papier. Elle reste ainsi convaincue que l'ordinateur n'est pas approprié pour la visualisation des photos, et que rien ne peut remplacer les tirages sur papier.

Elle explique également que, même si elle disposait de tous les périodiques en version électronique, elle préférerait continuer à fonctionner avec le système papier, en imprimant les PDF et en rangeant ces impressions dans les boîtes d'archives. Elle ne souhaite donc pas initier un dépouillement électronique comme cela est déjà un peu pratiqué par la documentaliste de la publicité.

En revanche, elle serait intéressée par l'abolition de certaines contraintes spatiales grâce à la dématérialisation. Elle aimerait notamment ne plus avoir à se déplacer à la bibliothèque, dans l'hypothèse où cette dernière numériserait les périodiques qui l'intéressent. Sans toutefois que cela ne vienne remettre en question le système papier actuel d'une manière générale. Même si la dématérialisation ne lui semble pas être un projet vraiment prioritaire et stratégique pour sa documentation, elle conçoit néanmoins Internet et le monde numérique comme un outil essentiel pour son travail. Par exemple, elle considère les bases de données spécialisées comme une de ses sources d'informations fondamentales, et les utilise très souvent.

Très peu de visiteurs externes au musée sont invités à consulter le fonds de cette documentation, seulement une vingtaine par an. La documentaliste n'a pas vraiment d'avis, d'opinion ou d'intuition à propos de l'impact sur le fonds que pourrait avoir la numérisation en termes de visibilité ou de popularité. Elle n'a pas non plus vraiment réfléchi aux possibilités que pourrait offrir la dématérialisation pour recevoir du public sans qu'elle ne doive interrompre ses tâches ou être mobilisée.

En conclusion, malgré un intérêt très modéré envers la dématérialisation, elle serait très intéressée par la numérisation des périodiques anciens de la bibliothèque, qu'elle consulte fréquemment. Elle estime ainsi que c'est plutôt le rôle de la bibliothèque des Arts Décoratifs, et non celui du centre de documentation, de prendre en charge cette dématérialisation.

Les périodiques stockés dans la documentation qu'elle utilise le plus sont La gazette de l'hôtel Drouot, Le journal des Arts, et les catalogues des ventes.

3.6.4 Compte rendu de l'entretien avec la documentaliste responsable de la documentation du 17^{ème} et du 18^{ème}

Le fonds documentaire de cette documentation n'est pas très volumineux et assez jeune, puisqu'il a commencé à être constitué seulement en 2002.

Les seuls consultants en interne sont les conservateurs, mais ils utilisent en fait très peu cette partie du centre de documentation. Quelques visiteurs externes au musée sont accueillis à la

documentation, mais cela reste assez exceptionnel. La plupart des demandes sont redirigées vers la bibliothèque, qui répond aux demandes n'ayant pas directement de liens avec les œuvres des collections du musée.

Les revues les plus consultées sont La Gazette de Drouot (qui fait également l'objet d'un abonnement en version numérique) et L'estampille - L'objet d'art. La documentaliste consulte également beaucoup les bases de données spécialisées, qui contiennent selon elle beaucoup d'informations.

Globalement, l'appréhension et la position de cette documentation vis-à-vis d'un projet de dématérialisation sont assez proches de celles de la documentation Art nouveau Art déco. La documentaliste n'est pas très enthousiaste à l'idée d'une numérisation massive. Elle ne semble pas y voir une nécessité absolue ni un intérêt substantiel pour le centre de documentation. C'est une documentation qui reste donc, elle aussi, fortement attachée au système papier actuel.

Bilan à propos de ces entretiens

Il apparaît que le centre de documentation, disposant de moyens humains et financiers assez limités, pare au plus pressé et au plus urgent, gérant le quotidien sans pouvoir réellement anticiper ni entamer un réel travail de réflexion sur des pistes futures d'évolution ou de mutation en matière de numérique.

De plus, les métiers et les professions propres aux musées des Arts Décoratifs sont, de manière générale, culturellement et historiquement liés et attachés au papier. Un facteur qui est assez peu favorable à une dématérialisation rapide et bien vécue par tous.

Concernant les documentalistes, le gain de place potentiel dans leur documentation semble être un critère en faveur de la dématérialisation.

En revanche, concernant les conservateurs, l'abolition des contraintes spatiales ne semble pas être un critère déterminant pour engager une dématérialisation. En effet, une grande majorité d'entre eux ne voit aucun inconvénient à se déplacer au centre de documentation et à rechercher les documents parmi les étagères.

3.7 Préconisations techniques complémentaires

Sur le plan technique et technologique, il est nécessaire d'effectuer une vraie démarche qualité pour ce projet. Ce qui suppose d'être capable de faire une distinction nette entre le besoin à satisfaire et la solution choisie pour satisfaire ce besoin. Il faut donc identifier le « juste besoin » sans aller au deçà ni au-delà.

Dans cette optique, cette étude livre ici quelques préconisations techniques complémentaires qui aideront dans le choix des futures solutions.

3.7.1 En matière de scanner

La capture numérique est la première phase du projet de dématérialisation, et elle requiert l'utilisation d'un numériseur.

Actuellement, le centre de documentation est doté de photocopieurs. Ces machines offrent des possibilités de numérisation des documents, et peuvent être en théorie connectées à presque tous les logiciels de GED sur le marché. Ainsi, certains organismes procédant à une dématérialisation régulière mais très modérée - c'est-à-dire sans traitement d'un gros volume de documents - utilisent les fonctionnalités de leurs photocopieurs, sans investir dans des scanners dédiés et spécialisés.

Mais cette solution paraît peu recommandable dans le cadre de notre, pour les raisons suivantes :

- Tout d'abord, les capacités de numérisation, de traitements, et de paramétrages des photocopieurs du centre de documentation sont moins évoluées que celles des scanners vendus actuellement et conçus pour des projets de dématérialisation. Il serait donc regrettable de disposer d'un puissant outil de GED sans pouvoir exploiter toute la puissance des scanners actuels.
- Ensuite, il a été évoqué plus haut la possibilité pour le centre de documentation de procéder à des opérations rétrospectives de numérisation du stock des dossiers documentaires - en quelque sorte des périodes intenses et massives de dématérialisation -, qui pourraient être notamment mises en place avec le recours à des stagiaires. Dans cette hypothèse, il serait nécessaire de pouvoir disposer à plein temps d'un scanner spécifique à ces opérations, et sans devoir mobiliser un ou plusieurs photocopieurs utilisés quotidiennement par les différents personnels des musées.

L'étude de faisabilité a vocation à fixer une première sélection de scanners – assez large - qui pourraient potentiellement répondre aux besoins en termes d'acquisition numérique. Mais il existe actuellement sur le marché des scanners une multitude de produits, et l'offre n'est donc pas facile à appréhender.

Il faut aussi souligner également un point important : dans nombre de projets de dématérialisation, le choix final du scanner est très fortement conditionné par une proposition faite par l'éditeur de la solution d'OCR ou de GED retenue.

3.7.1.1 Le choix du meilleur segment de scanners

Parmi la multitude de marques existant sur le marché, présentant chacune de très nombreux modèles, il a été décidé dans le cadre de cette étude de faisabilité d'exclure les imprimantes multifonctions, qui sont plus ciblées pour le grand public.

Malgré tout, on se rend compte que la frontière est parfois floue (surtout pour la gamme des scanners de bureau) entre les produits destinés aux professionnels et ceux pour le grand public (qui la plupart du temps font les mêmes tâches, mais plus lentement).

L'offre des scanners peut se découper actuellement en neuf segments, et il faut dans un premier temps bien cibler le segment le plus adapté au projet du centre de documentation.

Voici les neuf segments :

- les scanners de bureau (jusqu'à 29 pages par minute) ;
- les scanners départementaux (de 30 à 79 pages par minute) ;
- les scanners de production (à partir de 80 pages par minute) ;
- les scanners réseau ;
- les scanners de chèques ;
- les scanners de livres ;
- les scanners de photos ;
- les scanners micrographiques ;

Concernant la numérisation au fil de l'eau des périodiques originaux, on a vu que la vitesse d'alimentation des dossiers thématiques était modérée.

Et dans l'hypothèse d'une dématérialisation rétrospective du stock des dossiers thématiques, en interne, elle se fera plutôt de manière ponctuelle, par lots de boîtes d'archives, et sans nécessiter une haute vitesse de numérisation.

Ainsi, le produit recherché pour ce projet appartiendra au segment des scanners dits de « bureau »³³.

3.7.1.2 La technologie des capteurs : généralisation des CCD en matière de GED

Les deux technologies de capteurs existant actuellement, à savoir le CCD et le CMOS, ont été présentées dans la partie 2.3.1 consacrée à l'acquisition numérique.

Actuellement, tous les scanners dédiés à la GED documentaire – notamment ceux appartenant au segment des scanners de bureau- utilisent des capteurs CCD dernière génération. C'est donc cette technologie qui est préconisée pour notre projet.

Néanmoins, il n'est pas inutile de donner ici quelques petites précisions sur ce qui différencie ces deux systèmes optoélectroniques, notamment sur la question de l'entretien et de la maintenance :

Avantage du CCD : en cas de panne les éléments défectueux peuvent être changés indépendamment les uns des autres.

Inconvénient du CCD : la nécessité d'un petit temps de préchauffage pour être opérationnel.

Avantage du CMOS : le système est moins consommateur d'énergie que la technologie CCD, et ne nécessite aucun temps de préchauffage pour être opérationnel.

Inconvénient du CMOS : en cas de panne, l'ensemble formant un tout doit être changé dans son intégralité.

3.7.1.3 Les caractéristiques fondamentales

Afin de mieux cerner l'offre matérielle et de faciliter la sélection, on s'est ensuite attaché à définir et analyser le plus précisément possible toutes les caractéristiques fondamentales qui sont préconisées pour le scanner.

- Le format d'alimentation

Le scanner devra pouvoir numériser aux formats A4 et A3.

- Glace d'exposition et chargeur ?

Le scanner devra disposer absolument d'une glace d'exposition au format A3, pour la numérisation à plat des périodiques originaux reliés.

³³ Cf. Scanners de bureau. Archimag, 2008, La dématérialisation : guide pratique, p. 40. ISSN 0769-0975

Mais, si la documentation de la publicité envisage aussi de dématérialiser, en interne et sur une période relativement courte, tout le stock des dossiers thématiques au format papier, alors le scanner devra aussi autoriser la numérisation par défilement et disposer pour cela d'un chargeur.

Or actuellement très peu de scanners cumulent les deux technologies citées plus haut. En effet, les scanners disposant d'une vitre d'exposition appartiennent à la catégorie dite des « scanners à plat » alors que les scanners disposant d'un chargeur appartiennent eux à une autre catégorie dite des « scanners à défilement ». Ces deux catégories étant bien distinctes sur le marché actuel, très peu de scanners intègrent et disposent des deux technologies en même temps. Le prix de ces produits est par conséquent toujours très élevé.

Le choix du scanner sur ce point sera donc en fait complètement dicté et conditionné par la stratégie qu'aura fixé le centre de documentation en matière de dématérialisation :

- Si on décide ne pas procéder à une dématérialisation rétrospective du stock des dossiers thématiques, mais plutôt de continuer à exploiter le fonds papier déjà existant, alors un scanner à plat suffira. L'inconvénient d'une exploitation simultanée d'un fonds papier et d'un fonds électronique semble tout de même très pénalisant, et cette solution apparaît donc étant seulement provisoire. Ce point sera d'ailleurs abordé plus largement dans la partie 3.7 consacrée aux contraintes et aux stratégies de la dématérialisation.

- A l'inverse, si la documentation publicité décide de mener en parallèle, et ce dès le début du projet, une dématérialisation au fil de l'eau du flux entrant et une dématérialisation rétrospective du stock des dossiers, alors il sera très certainement indispensable d'acquérir un produit muni aussi d'un chargeur et permettant la numérisation par défilement. Plus précisément, si la dématérialisation rétrospective est prévue pour être massive et intense, alors le chargeur sera obligatoire. En revanche, si cette dématérialisation rétrospective plutôt envisagée de manière très ponctuelle et très légère – par exemple, numériser un dossier thématique qui serait particulièrement demandé et consulté par les utilisateurs -, alors l'opportunité et l'utilité d'un chargeur et d'un défilement automatique devront être discutés et évalués avant de se prononcer pour un choix final.

Dans le doute, et sans stratégie arrêtée, le centre de documentation devra se doter d'un produit cumulant la technologie du scanner à plat et celle du scanner à défilement.

- La résolution de numérisation

On a vu qu'il s'agit à proprement parler de la précision du scanner, et qu'elle s'exprime en points par pouce (dpi).

Pour la numérisation des dossiers documentaires, le scanner devra pouvoir numériser avec des résolutions comprises entre 300 et 400 dpi.

- Le mode de numérisation

Cela correspond à l'ensemble des fonctionnalités que propose le scanner, et qui sont accessibles depuis le logiciel de numérisation pour le paramétrage du scanner.

Le scanner devra pouvoir numériser en mode noir et blanc, en mode niveau de gris et en mode couleur.

Parmi les multiples fonctionnalités qui peuvent être théoriquement associées à ces modes, voici celles qui sont fortement recommandées pour le choix du numériseur :

- pouvoir numériser le recto et le verso ;
- détection automatique du format de page ;
- correction des désalignements ;
- diminution du moiré ;
- détection des pages blanches ;
- prénumérisation ;
- correction des gammas ;
- suppression de la couleur de fonds ;

Cette dernière fonctionnalité est particulièrement importante dans le cadre du traitement OCR. En effet, une des principales causes d'échec de ce dernier est la présence d'un fond de page, ou « background », sur le document scanné. Il faudra donc rechercher dans les fiches descriptives des produits si ce genre de système « intelligent et automatisé » est intégré ou non.

- La vitesse de numérisation

Ce point a déjà été abordé plus haut. La vitesse à laquelle le scanner travaille s'exprime en pages/minute.

Une vitesse entre 15 et 30 pages par minute suffirait largement pour ce projet, dans l'hypothèse où les éventuelles opérations de dématérialisation rétrospectives seraient assez peu fréquentes (puisque sans caractère d'obligation ni d'urgence), pas très massives ni très intenses – qui plus est sans possibilité de mettre une place une indexation totalement automatisée, excepté celle plein texte offerte par l'OCR.

- La volumétrie

Cela correspond au volume de documents traités par jour.

Le scanner devra pouvoir numériser plusieurs centaines de pages par jour - une valeur d'environ 1000 pages/jour serait convenable.

- La connectique

Cela concerne les connexions entre l'ordinateur et le scanner. Actuellement il existe deux types de connexions, le USB 2.0 et le Firewire. Il est préconisé que l'ordinateur pilotant le scanner soit doté d'un port Firewire.

- Les caractéristiques physiques

Elles sont liées au fonctionnement du scanner, comme le niveau sonore en fonctionnement, la consommation électrique, et l'encombrement. Aujourd'hui les différences sur ce point entre les modèles sur le marché sont relativement minimes, et seront peu décisives pour le choix final. A ce stade de l'étude, aucune préconisation particulière n'est faite sur ces caractéristiques physiques, et elles seront éventuellement considérées si, au final, un choix devait s'opérer entre deux ou trois scanners très proches au niveau de toutes leurs autres caractéristiques.

3.7.1.4 Logiciel d'acquisition fourni et pilotes standardisés

Chaque scanner est livré avec son propre logiciel d'acquisition, qui permet d'utiliser l'ensemble des fonctionnalités disponibles pour produire des images. Concernant les pilotes (ou drivers) des scanners, il en existe actuellement de deux types : ISIS et TWAIN. Ces derniers sont également toujours fournis avec le scanner et sont pilotables par tous les logiciels construits autour de ces normes.

D'une manière générale, tous les scanners aujourd'hui proposés sur le marché sont facilement intégrables dans n'importe quel système d'information existant, et notamment avec les logiciels de GED.

3.7.2 En matière de logiciel de GED³⁴

Après numérisation et reconnaissance des documents et des caractères, il est forcément nécessaire de pouvoir gérer et manipuler les fichiers résultants de ces processus. C'est donc bien grâce à un outil de GED que les informations dématérialisées deviendront du contenu et seront utilisables par les utilisateurs.

Les caractéristiques techniques qui figurent dans les fiches commerciales des logiciels doivent être étudiées et décryptées par le responsable informatique des Arts Décoratifs, qui est également le responsable du système d'information. L'architecture informatique actuelle du musée se base notamment sur un OS Windows et, depuis peu, sur une base Oracle. A priori, il n'existe donc aucun problème particulier ni d'obstacle majeur pour y intégrer un

³⁴ A titre d'exemple, ce mémoire propose une présentation de l'offre logicielle EverSuite - de l'éditeur Ever Team -, consultable en annexe 3 à la page 168.

logiciel de GED. Néanmoins, lors de la rédaction du cahier des charges, certains points importants devront être analysés par la direction informatique, comme par exemple les capacités de stockage des serveurs, ou la bande passante disponible sur le réseau Ethernet.

A ce stade de l'étude, et en amont de la rédaction de ce cahier des charges, voici les préconisations techniques qui sont faites pour le choix³⁵ du logiciel de GED.

3.7.2.1 L'architecture de la solution

En termes de spécifications logicielles, l'architecture de la solution sera de type client / serveur – le poste client pouvant être soit lourd, soit léger. Dans le cas d'un client léger, l'ensemble des traitements applicatifs s'effectue au niveau du serveur. Le principal intérêt de cette architecture est la rapidité du déploiement et des mises à jour qui s'effectuent uniquement au niveau du serveur.

3.7.2.2 L'administration de la solution

Cet aspect du logiciel, et les besoins dans ce domaine, doivent faire l'objet d'une concertation entre le responsable du projet - du côté des documentalistes du centre de documentation - et le responsable informatique des Arts Décoratifs.

Pour assurer la sécurité et l'intégrité des documents, le logiciel de GED devra donc comporter une véritable interface d'administration, qui permettra notamment de configurer précisément des droits d'accès personnalisés aux utilisateurs et/ou aux groupes d'utilisateurs. Cette option de « gestion des rôles et identités » des utilisateurs est donc préconisée pour la solution, puisqu'elle permettra non seulement de contrôler et fiabiliser l'utilisation du fonds documentaire dématérialisé, mais aussi de coordonner et surveiller précisément la production de documents numérisés.

Par exemple, on sait déjà à ce stade de l'étude qu'il est prévu que le logiciel de GED soit installé sur les postes informatiques de la salle de consultation, afin que les conservateurs et le public visiteur puissent consulter le fonds dématérialisé. Ces postes devront donc être configurés de telle sorte que les documents soient accessibles uniquement en lecture, afin que ces derniers ne puissent pas être modifiés ou détruits.

Toujours sur le plan de l'administration, une option de type « tableau de bord » dans le logiciel est fortement préconisée, puisque cela constituera une assurance pour le suivi du projet et les bilans, en offrant un retour intéressant pour les documentalistes et

³⁵ Archimag a réalisé fin 2008 un tableau récapitulant les caractéristiques des principaux logiciels et modules de GED. Cf. Logiciels et modules de GED. Archimag, 2008, La dématérialisation : guide pratique, p. 72-77. ISSN 0769-0975

l'administrateur du système. Un tableau de bord permettra notamment de délivrer des statistiques - sur le type de recherches effectuées, la fréquence de recherches infructueuses, sur la fréquence par utilisateur, etc - et de mettre en place des indicateurs qualitatifs et quantitatifs. Et sur le plan de la gestion du rythme de travail, par exemple, si la stratégie adoptée consiste à découpler le dépouillement et la numérisation physique de l'indexation électronique, en procédant à des phases respectives intenses, alors cet outil pourra aider à équilibrer et contrôler l'état d'avancement de ces tâches – interrompre par exemple momentanément le dépouillement pour se consacrer exclusivement à l'indexation.

Enfin, la question du besoin ou non d'une option de « gestion des quotas d'espace disque » est surtout du ressort du directeur informatique, et non des documentalistes.

3.7.2.3 Le gestionnaire de base et les capacités de stockage

Sur le plan du stockage de l'information, il sera nécessaire que le logiciel retenu puisse s'adosser à un gestionnaire de base de données Oracle, configuration qui a été récemment implantée sur les serveurs du musée. Concernant la capacité de stockage du serveur SQL, elle devra être déterminée par le service informatique lors de la rédaction des spécifications fonctionnelles et techniques du projet.

3.7.2.4 La déclinaison de l'interface utilisateur

Pour la solution de GED, il est fortement préconisé de pouvoir disposer de deux types d'interfaces utilisateurs, en prévision d'une future politique qui consisterait à rendre le fonds visible sur un parc informatique de plus en plus large et important, et pour des utilisateurs de plus en plus nombreux et variés.

La solution retenue devra donc proposer la possibilité :

- d'implanter une interface utilisateur par téléchargement sur un poste de travail ;
- d'utiliser directement Internet Explorer ;

Elle devra donc être une application basée sur l'utilisation d'un serveur Internet en mode ASP.Net³⁶, permettant la récupération de documents via l'utilisation d'un explorateur web.

La différenciation entre ces deux interfaces permettra une meilleure adéquation avec les différentes catégories d'utilisateurs amenés à consulter les documents numériques.

³⁶ ASP.Net est un ensemble de technologies de programmation web créé par Microsoft, qui succède à la technologie Active Server Pages (ASP).

La première catégorie d'utilisateurs sera composée des documentalistes, qui lanceront l'interface « client » depuis leur poste personnel afin de travailler sur le fonds numérisé, et qui seront amenés notamment à numériser ou modifier des documents.

La deuxième catégorie d'utilisateurs sera composée des conservateurs qui ne seront pas amenés à numériser ou à modifier un document. Leur utilisation se limitera à une simple consultation des documents, et l'utilisation avec Internet Explorer sera donc satisfaisante. Cette solution aura l'avantage de permettre une connexion directe au serveur Web de la GED sans demander aucune installation au préalable sur le poste client. Cette possibilité pourrait permettre une généralisation de l'utilisation de la GED par tous les conservateurs et leurs assistants, depuis leur poste de travail et sans induire de coûts matériels.

Enfin, la troisième catégorie d'utilisateurs sera représentée par le public externe invité à consulter sur place, qui consultera également la base GED via le serveur Web et le navigateur Internet Explorer. Tous les postes de la salle de consultation pourraient être ainsi opérationnels sans aucune maintenance informatique au préalable.

La richesse de l'interface utilisateur sera un facteur d'efficacité mais aussi d'adoption du logiciel de la part des conservateurs et des documentalistes des musées. Il est bon de rappeler encore une fois que la recherche, via cette interface, devra évidemment avoir plusieurs dimensions (sur métadonnées, sur plan de classement, booléenne, sur texte intégral, etc).

3.7.2.5 Une configuration multi-utilisateurs

Le logiciel devra être un système de gestion de documents multi-utilisateurs puisque, au minimum, il faudra implanter le logiciel sur le poste de la documentaliste responsable de la publicité et sur un ou plusieurs autres postes de consultation. Dans cette perspective d'utilisation de la solution à partir de plusieurs postes, au vu de la configuration du centre de documentation et des usages prévisibles qui seront fait des dossiers thématiques sur la publicité, la souplesse d'une gestion de type « licences flottantes ou dénommées » pourrait être intéressante. C'est donc une option à étudier, sachant que ce type de licence permet par exemple d'installer un logiciel sous plusieurs postes de travail mais d'autoriser une seule utilisation simultanée du logiciel. Ce qui permet de mutualiser les licences et d'abaisser les coûts à l'achat.

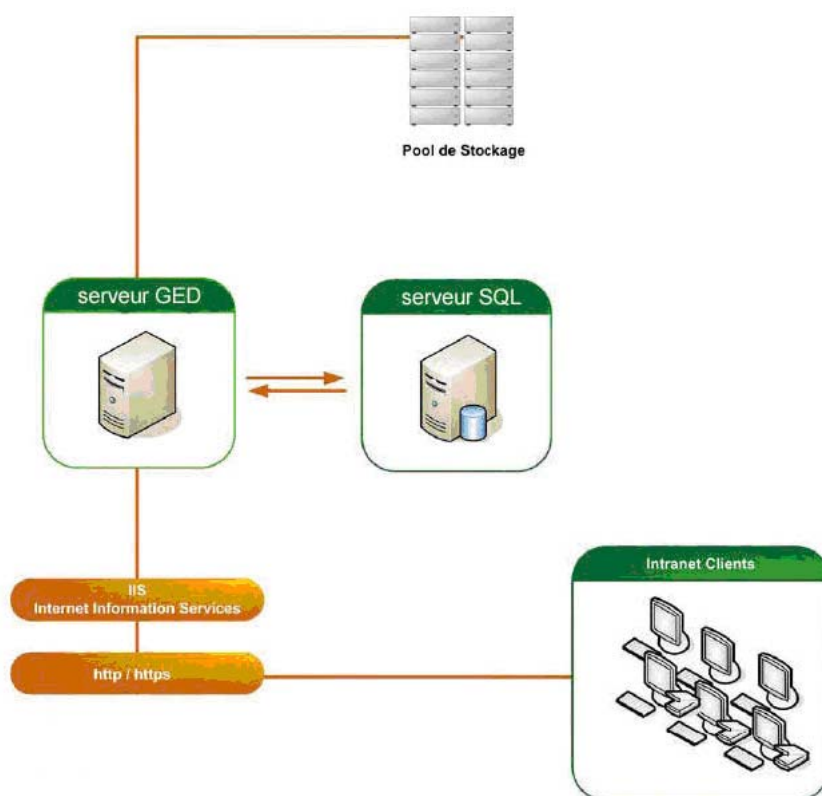
Le mécanisme de licence flottante s'appuie sur :

- un serveur de licences qui fournit en temps réels les jetons d'utilisations aux différents utilisateurs ;
- un fichier de licences qui regroupe l'ensemble des licences d'utilisations disponibles ;

Concrètement, dans l'hypothèse où seuls les dossiers thématiques de la publicité seraient dématérialisés dans un premier temps, on peut imaginer la stratégie suivante :

- Première option : Cette solution de licences flottantes permettrait d'installer le logiciel sur tous les postes informatiques de la salle de consultation pour l'achat d'une seule licence, et sans dédier un poste spécifique à la consultation des fonds des sur la publicité. Cette solution conviendrait puisque la documentation de la publicité reçoit la plupart du temps un seul visiteur à la fois, et une seule fois par semaine. La fréquence de consultation du fonds par les conservateurs du département de la publicité étant très faible – moins d'une fois par jour -, il n'y aurait ainsi quasiment aucun risque d'avoir deux demandes de consultation simultanées de la part de deux personnes – un conservateur et un visiteur externe.

- Deuxième option : Il est aussi envisageable de déployer, avec le seul achat d'une licence flottante supplémentaire, la solution de GED sur tous les postes des documentalistes et des conservateurs des autres départements et documentations du centre. Cette solution pourrait convenir puisqu'il est assez peu probable que plusieurs personnes de ces autres départements et documentations aient simultanément un besoin de consulter les dossiers thématiques de la publicité. En fait, cette solution présenterait surtout l'avantage de permettre à ces autres personnels du centre de documentation de tester et d'évaluer le système de GED. Le fonds numérisé des dossiers documentaires de la publicité serait alors une sorte de « pilote » de la dématérialisation au centre de documentation.



3.7.2.6 La question de la sécurité et le cryptage

En termes de sécurité, toutes les données transmises au travers du réseau entre les différents composants de la solution logicielle devront être sécurisées grâce à un chiffrement 128 bits, qui est le niveau de cryptage le plus répandu actuellement.

Pour la stabilité et la sécurité du système, il est également souhaitable d'instaurer une séparation entre les applications client, serveur et stockage de données. L'accès à la base de données et aux documents stockés se fera uniquement par le biais du serveur de l'application de GED choisie, sans qu'aucune autre application client ne puisse avoir accès aux index et aux documents.

De manière tout à fait classique et actuellement généralisée, le poste client se connectera au serveur de l'application via l'Intranet par l'utilisation du protocole sécurisé http(s).

3.7.2.7 Un outil de création de plan de classement

Le logiciel devra comporter un outil de création de plan de classement, et éventuellement proposer une option permettant d'utiliser un plan de classement préexistant.

3.7.2.8 Le bulletinage

Le bulletinage est une fonctionnalité avancée qui est préconisée et souhaitée pour la solution de GED. Elle consiste à enregistrer les numéros de périodiques reçus au fur et à mesure de leur entrée, pour connaître l'état de la collection grâce à une grille de pointage.

3.7.2.9 Le standard XML (eXtensible Markup Language) pour l'échange de données, et la souplesse offerte par la Déclaration de Type de Document (DTD)

Le « format standard » des systèmes d'informations, depuis déjà quelques années, est le XML, qui constitue un modèle pour la communication entre les applications. XML, orienté vers l'échange de données, est notamment une méthode universelle et standardisée de représentation textuelle des données structurées. C'est aussi un langage à balise normalisé par le World Wide Web Consortium (W3C). Avec XML, les possibilités de structuration logique du texte permettent de gérer tout à la fois le document comme un tout, pourvu de ses métadonnées de description, ainsi que les parties qui le composent. Ces dernières sont en quelque sorte des unités documentaires propres : elles disposent de leurs métadonnées spécifiques et héritent des propriétés du niveau supérieur, mais elles peuvent être utilisées indépendamment de lui. Avec XML, chaque élément significatif d'un document sera décrit

comme tel et balisé de manière à pouvoir être géré – décrit, modifié, recherché, affiché – indépendamment du reste du texte.

La notion de « document XML » permet ainsi de décrire un fonds documentaire : dossiers, collections, ouvrages, etc. Un document XML est une structure de données adressable, identifiable à une base de données : il est donc possible de lui appliquer une requête.

XML utilise des modèles- ou types - de documents appelés DTD (déclaration de type de document) qui fixent les contraintes de validité propres au document. L'utilisation de ces DTD est donc intéressante puisqu'elle permet de définir et de gérer librement plusieurs types de documents. Ces derniers présenteront chacun un choix de champs d'index et une structure d'indexation particulières et configurables, qui leurs seront propres, et sans qu'il n'existe de limite à la configuration.

En quelque sorte, le modèle « DTD XML » structure le format de la base dans laquelle sont archivés des documents, et joue d'une certaine façon le rôle d'un filtre ou d'une interface.

En matière de GED, ce langage XML est donc aujourd'hui considéré comme la norme pour l'échange de données, quelle que soit la plateforme d'origine. De par sa capacité à classer l'information et les métadonnées d'indexation sous une forme facilement visible, il est devenu actuellement un outil véritablement indispensable. Utiliser XML pour représenter un fonds d'objets numériques, textes ou images, c'est apporter à l'utilisateur une forte valeur ajoutée documentaire.

Il est donc préconisé que la solution de GED choisie supporte le format XML, et qu'elle soit parfaitement adaptée aux nouvelles spécifications en matière de gestion de documents et de plateforme web XML. Le serveur de la solution de GED devra donc être un service Web XML basé sur les dernières technologies de type client/serveur. Ce serveur gèrera l'ensemble des requêtes formulées, et les expédiera au serveur de base de données pour traitement.

3.7.2.10 La problématique de la migration des langages d'indexation documentaire et des métadonnées

La documentation de la publicité – de manière identique à toutes autres documentations du centre - utilise actuellement via Mobytext un langage d'indexation sous forme d'un thésaurus spécifique au domaine de la publicité. Ce dernier permet notamment d'indexer dans Mobytext les articles sélectionnés lors du dépouillement électronique des revues CB news et Marketing magazine. La documentaliste souhaiterait naturellement pouvoir exploiter ce même thésaurus pour la future indexation électronique, dans le système de GED, de tous les articles qui seront dématérialisés – de manière rétrospective, mais aussi au fil de l'eau. Ce thésaurus de la publicité étant de plus, dans sa structure, très proche du plan de classement actuellement en place pour les dossiers thématiques papier. Mais sera-t-il possible

techniquement d'importer un tel thésaurus dans la future application GED ? Pour répondre à cette question, il faudra au préalable contacter l'éditeur Mobydoc ainsi que les éditeurs de logiciels de GED. Après avoir recueilli toutes les informations nécessaires, et si cette opération d'importation se révèle être possible, alors cette dernière pourrait être réalisée en externe par un prestataire informatique de type SSII³⁷, ou bien en interne par la DSI³⁸ des musées des Arts décoratifs si celle-ci dispose des compétences informatiques et du temps nécessaire.

A plus ou moins long terme, un des principaux objectifs visés par les documentalistes grâce à la dématérialisation est de pouvoir centraliser, à partir de la seule application de GED et de sa seule interface, toute recherche d'informations portant sur tous les corpus du centre de documentation – bases des collections, dossiers documentaires et dossiers d'œuvres numérisés, catalogue des ouvrages et des revues, etc. Dans cette optique, il serait donc aussi très bénéfique de pouvoir organiser dans le futur la migration informatique des métadonnées à partir des logiciels documentaires Mobytext et Micromusée vers le futur logiciel de GED. Là aussi, une expertise informatique devra être menée pour évaluer la faisabilité et l'opportunité d'une telle opération. Par exemple, des ressources du centre de documentation sont décrites avec le format Unimarc³⁹ dans le logiciel Mobytext. On peut donc d'ores et déjà évoquer une piste avec la BiblioML – qui est à l'origine une initiative du ministère de la culture et de la communication. BiblioML est une DTD qui représente la notice bibliographique Unimarc. Plus précisément, il s'agit d'une spécification pour représenter les données bibliographiques et d'autorités. Et des outils sont aujourd'hui disponibles pour convertir les notices Unimarc en BiblioML. Créer un document XML avec cette DTD revient à créer un catalogue dont les documents seront décrits par une notice Unimarc. A noter que sur le site internet de Mobydoc, il est précisé que le logiciel Mobytext dispose, à titre d'option, d'un module d'échange en format Unimarc.

D'une manière générale, les éditeurs de solutions de GED sont naturellement très concernés par cette problématique des formats de métadonnées pour la description des documents, et restent donc très attentifs. Actuellement, les enjeux portent, comme d'habitude, sur la normalisation, et il est assez difficile de prédire quels seront les formats qui s'imposeront à terme. Les pistes de réflexion sont donc nombreuses – et souvent menées par les bibliothèques nationales -, et, dans un tel contexte, les éditeurs adoptent naturellement une politique plutôt attentiste. Néanmoins, tous les logiciels qui gèrent des flux XML peuvent

³⁷ Société de services en ingénierie informatique.

³⁸ Direction des systèmes informatiques.

³⁹ Universal MACHine-Readable Cataloging. C'est le format officiel d'échange de données en France.

déjà importer et exporter des métadonnées, mais hélas avec assez peu de contrôles et uniquement pour des formats de saisie classiques – propriétaires donc non échangeables, ou normalisés comme les formats MARC.

3.7.2.11 L'utilisation des codes à barres

D'une manière générale, l'utilisation d'un système de « détection de codes à barres » tend à se répandre dans les projets de dématérialisation. Cette technologie permet l'indexation automatique de documents, mais elle engage tout de même à faire un tri au préalable des documents, à préparer et à imprimer physiquement les codes, et enfin à paramétrer le logiciel avant de débiter la phase de numérisation.

Ainsi, il pourrait être judicieux de choisir une solution logicielle qui soit dotée de cette nouvelle technologie. Par exemple, pour l'indexation automatique de certains champs, comme le titre, le numéro et la date de parution d'un numéro de revue. Tous les articles sélectionnés et numérisés à partir de la revue pourraient être ainsi indexés partiellement de manière automatisée – on a vu que, sur ce point, l'OCR sera inefficace pour la reconnaissance des références écrites à la main sur les photocopies des articles.

3.7.2.12 La formation aux nouveaux outils

Pour optimiser au maximum l'exploitabilité des documents dématérialisés, il sera indispensable de maîtriser toute la puissance d'indexation et de classement du système choisi, mais aussi le paramétrage du scanner. Ceci notamment afin de fiabiliser autant que possible l'OCR. Or, même si l'ergonomie des interfaces est en perpétuel progrès, le niveau de paramétrage des logiciels de GED est aujourd'hui tel qu'il peut être délicat pour certaines personnes de se familiariser et de maîtriser l'outil sans recourir à une formation et à une démonstration, en interne ou en externe. Lors du choix de la solution, ce point devra être considéré et faire l'objet d'une attention particulière. Ceci afin de décider et d'inclure dès le début, s'il y a lieu, le budget formation dans le coût total de la solution.

3.7.2.13 L'évolutivité comme critère de choix de la solution

En termes d'évolutivité, le logiciel devra pouvoir s'adapter à une éventuelle généralisation de la dématérialisation au sein du centre de documentation.

Notamment concernant les trois aspects suivants :

- une augmentation des volumes de documents papier ou électronique à traiter ;
- une augmentation significative du nombre de postes utilisateurs connectés simultanément à la GED ;

- une diversité des documents manipulés en termes de format de stockage, de champs et d'indexation ;

Ces points très importants devront être abordés et traités lors des contacts avec les éditeurs de solutions qui auront été pré-sélectionnés.

3.8 Les contraintes et les stratégies de la dématérialisation

3.8.1 La numérisation rétrospective du stock des dossiers thématiques de la publicité

Après la mise en place de la GED, la documentation de la publicité n'alimentera plus du tout les dossiers thématiques au format papier stockés dans les boîtes.

Ainsi, dans le nouveau système qui a été prévu, dès la réception des périodiques – en version papier ou en version électronique –, ces derniers seront dépouillés et les articles intéressants systématiquement intégrés – après numérisation si nécessaire – dans la GED.

La vitesse d'alimentation de ce fonds est relativement modérée, et il n'est pas prévu qu'elle augmente après la mise en place de la GED. Le nombre d'abonnements à des périodiques, ainsi que le nombre moyen d'articles sélectionnés chaque semaine resteront sensiblement les mêmes qu'avec le système des photocopies. Cette dématérialisation au fil de l'eau pourra donc être prise en charge par la seule documentaliste de la publicité, même si, globalement, l'indexation électronique et les traitements lors de la numérisation pourraient venir allonger au final le temps qu'elle consacre chaque mois pour ces panoramas de presse et l'alimentation des dossiers.

En revanche, il est dès à présent nécessaire que la documentation de la publicité réfléchisse et fasse un choix concernant le traitement futur et l'avenir du fonds des dossiers au format papier.

Quelle stratégie et quel usage souhaite-t-on mettre en place ?

Les Arts Décoratifs auront-ils éventuellement les moyens humains et financiers, ou encore la compétence technique, permettant d'appliquer cette stratégie ?

Deux possibilités s'offrent dès à présent à lui :

- Soit il décide de démarrer, en parallèle de la dématérialisation au fil de l'eau pour le traitement du flux entrant, une dématérialisation rétrospective massive du fonds papier.
- Soit il préfère, mais d'une manière qui serait seulement provisoire, laisser les deux systèmes papier et numérique exister et être exploités simultanément par les utilisateurs et la documentaliste.

A propos de ce choix, l'avis de la documentaliste de la publicité est déjà forgé et clair : elle estime que rechercher à la fois dans les dossiers papier et dans les dossiers électroniques pénaliserait trop les utilisateurs et viendrait effacer une partie des avantages apportées par la GED. D'autant que, en plus du logiciel de GED, les utilisateurs et la documentaliste continueront à consulter les bases des collections à l'aide des autres logiciels documentaires

Micromusée et Mobytext. En effet, puisque la migration des bases des collections dans le logiciel de GED ne pourra pas se faire facilement ni rapidement – cela nécessiterait au préalable du conseil de la part des éditeurs et très probablement un développement informatique spécifique pour assurer la migration des données – notamment la récupération des métadonnées qui sont au format Unimarc dans Mobytext et Micromusée.

Dans ce contexte, le souhait profond de la documentaliste-responsable de la publicité est donc de pouvoir bénéficier le plus rapidement possible – c'est-à-dire le plus tôt possible après la mise en activité de la GED – d'un fonds numérique comportant absolument tous les dossiers thématiques dématérialisés. La recherche dans tous ces dossiers pourrait alors se faire d'un seul coup à travers le puissant et souple moteur de recherche de la GED.

Mais la faisabilité d'une dématérialisation rétrospective et massive est conditionnée par les trois points importants suivants :

- Les contraintes techniques identifiées pour dématérialiser les dossiers papier constitués de photocopies d'articles (et non de documents originaux en parfait état).
- Les moyens humains actuellement disponibles au centre de documentation.
- L'évaluation préalable du budget global qui pourrait être libéré pour ce projet.

Enfin, un dernier point capital à étudier dans l'hypothèse d'une dématérialisation rétrospective du stock des dossiers : serait-il plus intéressant de s'orienter vers une dématérialisation en interne, ou au contraire vers une prestation en externe auprès d'une société spécialisée dans ce genre de traitement ?

3.8.1.1 Contraintes et spécificités techniques liées à la dématérialisation

Avant d'aborder les difficultés et les obstacles, on rappelle tout de même que les scanners actuels permettent d'automatiser une partie du processus d'acquisition numérique, et d'alléger le paramétrage et l'intervention de l'opérateur, grâce notamment aux fonctionnalités suivantes :

- détection automatique d'un document en couleur et d'un document en noir et blanc ;
- détection automatique du format ;
- détection du recto/verso (certains scanners sont dotés d'une double matrice permettant de numériser d'un coup le recto et le verso d'une feuille) ;

• Les agrafes / Papiers froissés et cornés

Le premier obstacle de taille à une numérisation massive, partiellement automatisée et très rapide – notamment avec un scanner à défilement disposant d'un chargeur – est la présence d'agrafes dans le fonds. En effet, l'analyse du contenu de boîtes d'archives a révélé qu'entre

10 % et 20 % des documents se composent de plusieurs feuilles agrafées. Une intervention manuelle pour préparer ces documents est donc obligatoire, ce qui viendra retarder le processus d'acquisition numérique.

De plus, certains documents papier sont légèrement froissés ou cornés. Là aussi, une préparation manuelle de ces documents s'impose, afin qu'ils puissent être placés dans le chargeur sans risque d'endommager le numériseur.

- **Isoler l'article de la page avec un découpage et un recadrage numérique**

Cet aspect de l'acquisition numérique a déjà été abordé dans la partie 3.4 consacrée aux bénéfices attendus. Pour optimiser le confort du lecteur, et pour économiser de la mémoire de stockage informatique, il serait souhaitable d'isoler l'article de la page – dans le cas où plusieurs articles sont présents sur une même page. Les produits actuels disposent d'outil de prénumérisation et de découpage numérique (outils de sélection de cadre, de recadrage). Le numériseur ne connaît pas à l'avance l'article que l'on souhaite isoler de la page, et c'est donc l'opérateur qui doit intervenir manuellement pour sélectionner l'article dans la page prénumérisée. Ceci retarde donc fortement le processus d'acquisition et empêche une complète automatisation.

- **La fusion des pages numérisées en un seul fichier PDF / Les codes à barres**

Dans le cas où un article s'étale sur plusieurs photocopies, il est nécessaire à la fin du processus d'acquisition de fusionner toutes les pages numérisées en un fichier PDF unique. Ceci nécessite forcément une intervention ou un paramétrage manuel, au cas par cas. En effet, le numériseur n'est pas capable d'effectuer cette tâche seul puisqu'il ne connaît ni le début ni la fin de l'article. L'opérateur en charge de la numérisation doit donc intervenir, ce qui ralentit forcément la dématérialisation.

Afin d'essayer d'accélérer le processus, des innovations ont récemment été introduites, dont l'une consiste à imprimer un code à barres sur une page A4. Cette page est placée au-dessus des photocopies composant l'article. Le code à barres est ainsi numérisé et reconnu en premier, et il indique au scanner le nombre de pages composant l'article et qu'il faut numériser. Mais ce code à barres peut aussi donner d'autres indications et informations précises à la machine, comme par exemple le titre du périodique dont est issu l'article, ou encore des paramètres pour le traitement numérique.

Cette solution, qui peut être intéressante et alléger les interventions de l'opérateur, nécessite en contrepartie et au préalable une minutieuse préparation de ces codes à barres, ainsi qu'une programmation de ces derniers dans l'ordinateur. Un choix technique devra donc être fait sur ce point, et de manière générale il faudra évaluer le plus précisément possible le gain

potentiel de temps apporté par telle ou telle technologie. Sachant que, dans l'hypothèse où le centre de documentation se lancerait dans le traitement rétroactif et massif de ce fonds comportant plusieurs dizaines de milliers de documents, les différences à l'arrivée pourraient être énormes selon les techniques adoptées.

- **L'indexation par extraction automatique de données**

Le logiciel de GED qui sera choisi pour ce projet devra permettre une indexation en texte intégral et une indexation manuelle.

Mais il existe un troisième type d'indexation électronique : l'extraction automatique de données.

Les dossiers thématiques de la publicité sont-ils propices et adaptés à ce genre d'indexation ? Existe-t-il ici des contraintes techniques insurmontables à ce genre de traitement automatisé ?

Tout d'abord, pour répondre à cette question, il faut distinguer la dématérialisation au fil de l'eau de la dématérialisation rétrospective des dossiers.

Ensuite, il faut souligner que les principales données qui seraient susceptibles d'être extraites automatiquement des documents lors de l'acquisition numérique sont en fait les références des périodiques : titre, numéro, date de publication.

- *Extraction de données à partir du stock des dossiers thématiques au format papier*

Concernant le fonds papier des dossiers thématiques de la publicité, si on étudie tous les types et toutes les formes de documents qu'il renferme, il semble qu'il soit impossible de mettre en place des traitements systématiques de LAD ou de RAD – Lecture Automatique de Documents et Reconnaissance Automatique de Documents. En effet, les documents présents dans les dossiers sont presque tous des articles de périodiques qui présentent, de par leur nature et leur publication, des caractéristiques incompatibles avec les traitements automatisés de type LAD ou RAD.

Un premier facteur, déterminant, rend quasi impossible une utilisation massive de LAD ou RAD : la qualité d'impression très inégale, et parfois très mauvaise, des photocopies présentes dans les dossiers thématiques. L'échec à la reconnaissance optique de caractères ne permettrait donc pas de systématiser le traitement LAD ou RAD, et paraît déjà être rédhibitoire à la mise en place de traitements automatisés.

Un autre facteur vient renforcer l'idée que la reconnaissance optique de caractères ne serait pas fiable concernant les références des articles : sur plus de 35 % des photocopies, ces références ne sont pas typographiées (en haut ou en bas de page) mais écrites à la main (et de plus à des positions aléatoires). Or, on sait que les algorithmes OCR et ICR sont

complètement inefficaces sur l'écriture manuscrite excepté si les caractères sont écrits en majuscules, bien droits et bien espacés. Ce qui n'est pas du tout le cas des écrits sur les photocopies stockées dans les dossiers.

Dernière difficulté majeure : les articles dans les dossiers sont issus d'une multitude de périodiques, dont les mises en forme et les publications sont non seulement très différentes les unes des autres, mais qui, surtout, ont parfois sensiblement variées dans le temps. D'où une impossibilité d'attribuer, pour chaque périodique représenté dans les dossiers, un masque de reconnaissance et un gabarit qui seraient fixes et définitifs.

Par exemple, la position des références sur les pages d'une revue et leurs formes (police, taille de caractères, etc) ont pu changer en fonction des nouvelles formules de publication, et seront probablement encore amenés à évoluer.

Ainsi globalement, vu le nombre assez élevé de périodiques qui sont représentés dans ces dossiers thématiques, la tentative de mise en place de LAD et RAD impliquerait un travail de paramétrage préalable extrêmement lourd et précis, et pour des résultats de reconnaissance optique de caractères (OCR et ICR) qui, de toute façon, ne seraient pas absolument fiables.

Et tout ceci obligerait donc à mener un véritable travail manuel et intellectuel de vérification, de correction éventuelle, après les traitements automatisés. Ce qui éloigne toute idée de gain de temps substantiel grâce aux traitements LAD et RAD, qui pourraient même engendrer l'égarage de certains documents numériques (mal classés), ou la fourniture d'informations non pertinentes.

- Extraction de données lors de la dématérialisation au fil de l'eau

Concernant le traitement du flux entrant, au fil de l'eau, la situation est légèrement différente.

Tout d'abord, la numérisation au fil de l'eau se fait à partir des périodiques originaux, qui présentent donc une qualité d'impression irréprochable. Certaines revues sont même déjà reçues en version électronique. Ensuite, le problème des références écrites à la main disparaît.

On pourrait donc, uniquement dans le cadre de la dématérialisation au fil de l'eau et en se limitant seulement à quelques revues très présentes dans les dossiers thématiques – CB news et Stratégies seraient par exemple très appropriées -, essayer de mettre en place une extraction automatique des références des revues – par exemple le titre, le numéro et la date de publication – qui deviendraient ensuite des métadonnées de l'article numérisé.

Néanmoins, des tests préalables - dans le cadre d'une démonstration commerciale personnalisée par exemple - devront probablement être effectués avec des éditeurs de solutions de GED, afin d'évaluer très exactement la faisabilité d'une telle extraction de

données. Sans résultats de ces tests, il est donc aujourd'hui impossible de se prononcer avec certitude sur les possibilités en termes de LAD et RAD pour la dématérialisation au fil de l'eau.

3.8.1.2 Organisation et planification de la dématérialisation rétrospective du stock

Que la dématérialisation soit réalisée en interne ou au contraire par un prestataire spécialisé, ce sera le rôle de la documentaliste de la publicité – seule responsable et gestionnaire du fonds - d'organiser et planifier la numérisation des documents. Il s'agira par exemple de procéder soit en suivant l'ordre du plan de classement, soit en sélectionnant par fréquence de consultation des dossiers thématiques -autrement dit la popularité des documents.

3.8.1.3 Cas d'une dématérialisation rétrospective en interne

Dans l'hypothèse d'une dématérialisation en interne et sur le plan des moyens humains, la seule documentaliste responsable de la publicité ne pourra très certainement pas prendre en charge tout le processus. Ainsi, il pourra être nécessaire et utile de recourir à d'autres personnels – stagiaires, emplois en interim, mobilisation ponctuelle des autres documentations, toutes ces options devront être évaluées par la responsable du centre de documentation – au moins pour les phases de préparation et de numérisation physique des documents.

La préparation des documents consistant notamment à :

- Retirer les agrafes ;
- Décorner les pages ;
- Déplier les pages ;
- Eventuellement paramétrer et imprimer des codes à barres, et constituer des lots de documents similaires (qualité d'impression, couleur ou noir et blanc, recto seul ou recto/verso, etc) ;

En revanche, on ne pourra pas inclure le désherbage dans cette phase de préparation, puisque seule la documentaliste est apte à réaliser cette tâche.

Ensuite, la numérisation physique des documents – n'incluant pas le classement ni l'indexation manuelle - pourrait aussi être effectuée sans difficulté par d'autres personnels, à qui la documentaliste aura au préalable fait une démonstration et une courte formation en interne.

Ces personnels pourraient par exemple assurer les tâches suivantes :

- placer le document dans le chargeur ou sur la vitre d'exposition ;

- déclencher la numérisation en appliquant le profil d'utilisation (autrement dit les paramètres et les traitements automatisés) adéquat en fonction du document à traiter ;
- nommer et stocker le fichier de sortie ;

La phase de numérisation physique des documents et celle d'indexation manuelle pouvant être découplées et désynchronisées, on peut donc imaginer que la documentaliste puisse traiter intellectuellement le document et l'indexer après sa numérisation physique. Et l'utilisation d'un tableau de bord – option disponible dans la quasi-totalité des solutions de GED - pourrait l'assister dans la mise en place d'un planning et d'une organisation qu'elle aura fixés et qui lui sembleront les plus judicieux.

3.8.1.4 Cas d'une dématérialisation par un prestataire en externe

Actuellement, l'offre de services en numérisation est très large et se décline pour tous les types de documents et de besoins. Il est donc nécessaire de décrypter le marché actuel des prestataires spécialisés en dématérialisation, et d'essayer d'en tirer une première analyse.

Un panorama des principaux prestataires actuels a été réalisé fin 2008 par la revue Archimag⁴⁰. Il montre que la maturité des technologies d'OCR homogénéise les offres des prestataires, et que la qualité de la numérisation - particulièrement dans le cas de notre projet qui concerne des documents textuels et non des photos - en termes de résolution n'est plus discriminante. Ainsi, une numérisation autour 300 dpi - environ la valeur maximale requise pour le projet au centre de documentation - est aujourd'hui une norme standard proposée par tous les prestataires. Sur le plan technologique donc, les machines sont à peu près au même niveau de capacités et les prestataires se sont normalisés. Un constat identique se dresse concernant l'élaboration des tarifs, puisque le nivellement des performances techniques et la multitude de prestataires débouchent logiquement vers une égalisation des tarifs pratiqués.

Globalement donc, si les ressources humaines et techniques dont disposent les prestataires ne sont pas réellement de nature à faire actuellement une véritable différence lors du choix final, en revanche les études ont montré que, pour certains projets, la préparation des documents pouvait représenter jusqu'à 70 % du coût total de la numérisation. Or, on sait que le fonds des dossiers thématiques de la publicité nécessiterait une lourde préparation. C'est donc bien ce coût de préparation des documents qui représente le plus grand obstacle, en termes financiers, à une numérisation par un prestataire spécialisé.

⁴⁰ Cf. Trouver un prestataire spécialisé. Archimag, 2008, La dématérialisation : guide pratique, p. 78-90. ISSN 0769-0975

A titre indicatif, et au regard des tarifs moyens actuellement pratiqués, on évalue le coût de dématérialisation d'une page A4 entre 0,70 et 0,90 euros, pour une prestation qui comprend :

- prise en charge ;
- déconditionnement et préparation (sortir des pochettes, enlever les agrafes, ...) ;
- numérisation à 300 dpi et en couleur ;
- saisie pour chaque document (titre publication, titre article, référence page, référence unique) ;
- reconditionnement en pochette papier neuve et numérotée ;
- remise en boîte ;
- post numérisation : redressement, OCR, suppression des pages blanches ;
- production d'un PDF par document ;
- retour des documents et livraison des documents sur support au choix (Disque dur, CD, DVD) ;

D'une manière générale, pour obtenir très précisément le devis d'une prestation, voici quelques indications à respecter concernant les informations et les éléments qu'il est indispensable de communiquer au prestataire :

- faire une rapide présentation de l'organisation des musées des Arts Décoratifs et du centre de documentation (activités, directions, départements et documentations directement concernés par le projet, etc) ;
- faire une synthèse de l'objectif du projet (quoi, pour qui, quand, retour sur investissement attendu) ;
- indiquer ce que le projet n'inclut pas (par exemple exclure le reconditionnement des dossiers papier si on souhaite les détruire après la numérisation) ;
- indiquer les acteurs du projet, notamment les équipes projets ;
- donner les contraintes qui ont été identifiées et pouvant influencer sur le projet (par exemple la présence d'agrafes, ou une qualité d'impression très inégale) ;
- proposer un planning, mais assez large ;

Ensuite, pour la rédaction d'un véritable cahier des charges, on livrera également aux représentants du prestataire, avec suffisamment de détails et de précisions, certaines spécifications fonctionnelles et techniques du projet.

Notamment :

- une description précise de l'existant : volumétrie, typologie, caractéristique des documents à reprendre ;
- des échantillons significatifs ;

- une description détaillée du résultat attendu (formats de restitution, pourcentage de reconnaissance, taux de rejet ou de perte toléré) et de l'usage projeté ;
- définir les modalités du contrôle de qualité attendu, c'est-à-dire préciser s'il est exigé par le centre de documentation une relecture systématique du document numérique par un opérateur, ou plutôt un contrôle par sondage (en précisant que certaines techniques informatiques proposent aujourd'hui des contrôles automatisés qui fournissent des résultats statistiques de ces contrôles) ;
- décrire toutes les contraintes pouvant influencer sur le projet et la prestation de numérisation. Par exemple, pour notre projet de dématérialisation des dossiers thématiques de la publicité, aucune clause particulière n'est à envisager sur le plan de la fragilité, ni sur le plan de la confidentialité ;

Tous les éléments qui viennent d'être cités peuvent avoir au final un impact significatif sur l'évaluation globale du budget nécessaire à la prestation de numérisation.

3.8.2 Le contexte d'un musée, et le prestige du support

Tout d'abord, il est important de bien prendre en compte le contexte et l'environnement dans lequel on met en place la dématérialisation. Il s'agit ici d'un centre de documentation dont la fonction principale est d'être au service des conservateurs et d'illustrer les collections des musées des Arts Décoratifs. C'est un lieu où l'attachement au papier et au support est forcément très présent. La rareté et l'originalité de certaines ressources du centre font qu'elles se prêtent bien à une dématérialisation dans le cadre d'un projet de conservation et dans le but d'assurer leur pérennité, et non à une dématérialisation visant à leur exploitation au format électronique - cette remarque ne concerne toutefois pas les dossiers thématiques de la publicité, constitués de panoramas de presse assez récents et ne comportant pas de documents d'une très grande rareté. Autrement dit, contrairement à des projets standards de dématérialisation en entreprise qui concerneraient des factures ou des formulaires, certains documents peuvent être ici de prestigieux et fragiles catalogues, de très vieilles et très rares photos, etc. Et naturellement, les conservateurs ainsi qu'une certaine partie du public accueilli éprouvent le besoin et le désir de manipuler, de toucher, ces vieux et prestigieux documents. Dans ces conditions, la perspective de dématérialiser paraît moins évidente ou indispensable, et peut même être perçue chez certains comme quelque chose de négatif plutôt que de positif, de réducteur plutôt qu'une source de progrès.

Autre spécificité qui concerne une petite partie du fonds, certains documents présents dans les dossiers documentaires sont en fait de véritables « petits objets » - par exemple des

cartons dépliés dans les trois dimensions. Il est donc impossible de les dématérialiser. La seule solution envisageable dans ces cas serait de conserver les objets dans les salles d'archives et de stocker dans la GED des photographies de ces objets prises sous plusieurs angles.

3.8.3 S'abonner aux versions électroniques de certains périodiques

Certains périodiques alimentant les dossiers documentaires du centre de documentation proposent déjà un abonnement pour une version électronique. On peut citer les exemples de CB news et de La Gazette de l'Hôtel Drouot.

Dans le futur, le nombre de ces périodiques disponibles en version électronique va très probablement augmenter, et la dématérialisation pourrait représenter une réelle opportunité de passer des abonnements au format papier vers des abonnements au format électronique.

En effet, les différentes documentations pourraient éventuellement en retirer les avantages suivants :

- Tout d'abord, certaines de ces documentations sont abonnées aux mêmes périodiques. Il y a donc des cas d'abonnements au format papier en doublon, on peut citer l'exemple La Gazette de l'Hôtel Drouot – un abonnement pour l'Art nouveau Art déco et un autre pour le 17^{ème} 18^{ème}. La GED pourrait ainsi permettre de faire l'économie d'un abonnement, en passant à une version électronique exclusive qui serait partagée par les deux documentations.

- Ensuite, la GED pourrait permettre, plus globalement, un partage systématique de tous les périodiques en version électronique par toutes les documentations des musées. Même si, très probablement, cela n'apportera rien dans la majorité des cas, il se pourrait que ce partage automatique en ligne soit tout de même intéressant dans certains cas particuliers. Par exemple, certaines revues dépouillées par la documentation mode et textile pourraient intéresser la documentation publicité, et réciproquement. Ce point devra faire l'objet d'une analyse et d'une réflexion de la part de tous les documentalistes.

En fait, un partage des périodiques pourrait même être instauré entre les différents services des Arts Décoratifs. Par exemple, la direction de la communication du musée - qui comprend notamment le service de presse des musées - est abonnée à des journaux de presse quotidienne (Le Monde, Le Figaro, Libération, etc) et à des revues d'information et d'actualités hebdomadaires (Le nouvel Observateur, L'express, Le point, etc), qui pourraient intéresser par exemple les documentalistes de la mode et du textile, de la publicité. Les dossiers thématiques de la publicité comportent ainsi de nombreux articles issus de la presse quotidienne.

Ces possibilités nouvelles offertes par la GED doivent tout de même être évaluées au regard des éléments suivants :

- Si le partage on-line de numéros de périodiques en version numérique ne présente à priori aucun obstacle ou difficulté technologique, il se heurtera en revanche très probablement à des habitudes de travail et à des réticences de la part de certaines documentations et de certains utilisateurs, qui sont très attachés au fait de pouvoir consulter et manipuler au format papier.
- De plus, l'exploitation de ces périodiques dans le cadre de la réalisation de panoramas de presse destinés à alimenter les dossiers documentaires doit être encadrée par des contrats, et faire au préalable l'objet d'une expertise et d'une analyse juridique très précise. Cet aspect du projet sera développé dans la partie 3.9 de cette étude.

3.8.4 La perspective d'une ouverture plus large au public externe

Les entretiens effectués avec les documentalistes ont mis en évidence quelques divergences d'opinion concernant certaines des perspectives offertes par la dématérialisation, notamment au sujet de :

- l'éventualité et l'opportunité d'ouvrir un peu plus largement le centre de documentation au public ;
- le désir de mettre en ligne sur Internet une partie des références des ressources du centre de documentation, dans le prolongement de la mise en ligne sur le site d'une partie des collections des musées ;

Le public externe qui est invité à consulter sur place se compose de chercheurs, d'universitaires, de créateurs, d'historiens, de critiques d'art, d'étudiants et de quelques particuliers ayant un besoin très précis et justifié. Le public est donc assez rigoureusement sélectionné, d'autant que l'accueil de ce dernier n'est pas la vocation du centre de documentation mais plutôt celui de la bibliothèque des Arts Décoratifs.

Actuellement, les faibles effectifs du personnel du centre de documentation ne permettent pas vraiment d'élargir ce public, néanmoins il a été exposé plus haut dans l'étude – cf. la partie 3.7.4 – que le système de GED pourrait éventuellement représenter un début de solution pour hausser la fréquentation du centre.

On constate cependant certaines disparités de fonctionnement selon les documentations du centre. On peut citer les exemples des documentations du verre et de la mode. La première a fait plutôt le choix de ne recevoir que des chercheurs renommés, les spécialistes de la discipline en quelque sorte, et limite ainsi au maximum les visites. La documentation de la

mode adopte elle une démarche presque opposée, en recevant près de 150 personnes par an, un public parfois très jeune ou composé d'étudiants pas réellement spécialistes. En matière d'accueil du public externe, les perspectives concernant la dématérialisation sont donc perçues de façon quelque peu distinctes par ces deux documentations.

Les Arts Décoratifs ont récemment mis en ligne sur leur site Internet un extrait des bases de données des collections et des revues. Cette initiative ayant rencontré un vif succès, il est envisagé par certains documentalistes – notamment la documentation de la publicité – de mettre en ligne sur Internet une partie des références - un catalogue en quelque sorte - des dossiers documentaires du centre de documentation. Potentiellement, ceci pourrait naturellement susciter l'intérêt et la curiosité de certains internautes, qui pourraient par la suite faire une demande de consultation auprès centre de documentation. La mise en ligne d'un tel catalogue des dossiers est donc susceptible d'augmenter la fréquentation du centre de documentation - tout comme la mise en ligne sur le site d'une partie des collections des musées contribue probablement à élever la fréquentation des musées. Mais toutes ces nouvelles sollicitations et demandes de visites pourront elles être gérées et assumées par le centre de documentation ? Sur ce point, certaines documentations ont des avis partagés. La documentaliste du verre ne souhaite pas consacrer plus de temps à l'accueil de visiteurs, et n'est donc pas très enthousiaste à la mise en ligne sur Internet de ce catalogue. Avec une perception différente, la documentaliste de la mode souhaiterait au contraire que la mise en ligne d'informations puisse se développer et prendre de l'importance. Sa motivation principale réside dans le fait qu'elle se dit consciente de la valeur et du caractère exceptionnel des ressources qu'elle gère. Ainsi de probables débats auront lieu, post dématérialisation, pour se prononcer sur l'adoption, ou au contraire l'abandon, de cette idée d'un catalogue on-line des dossiers documentaires.

D'une manière générale, il est évident que les réunions entre documentalistes autour de la faisabilité et l'opportunité d'une dématérialisation constituent une excellente occasion de provoquer des échanges, d'exprimer des opinions et de prendre du recul.

Les entretiens auront déjà permis de récolter les premières impressions des documentalistes.

De la même façon il pourrait intéressant et révélateur d'organiser deux petits sondages :

- L'un auprès des conservateurs, afin de connaître précisément leurs intérêts, leurs attentes et leurs appréhensions face à la dématérialisation, et pour essayer de mieux décrypter leurs rapports avec le papier.
- L'autre auprès du public, afin d'entendre ses préférences et ses attentes en termes de consultation électronique.

3.9 Une démarche qualité, des indicateurs de qualité

Après la rédaction du cahier des charges, et avant la mise en production du processus de dématérialisation, il existe une étape, un levier, qu'il ne faut pas oublier ou sous-estimer : celui de la démarche qualité.

Par exemple, les tests suivants pourront être effectués :

- Sélectionner des échantillons d'articles représentatifs parmi les dossiers documentaires à numériser, pour affiner les traitements à appliquer et éventuellement créer des profils d'utilisation. La préparation du document et le paramétrage du scanner peuvent être en effet spécifiques selon l'état et le type du document (niveaux de gris ou couleur, effet moiré, très faible contraste, caractères très petits, document corné ou plié, etc).
- Affiner les modes de préparation. La remise à plat des pages, la suppression des pliures excessives, l'enlèvement des agrafes ou trombones, l'orientation des pages et éventuellement le positionnement de codes barre sont quelques-unes des tâches qui devront être effectuées. Cette préparation concerne surtout la numérisation rétrospective, car, pour la numérisation au fil de l'eau, les flux entrants concernent essentiellement des périodiques neufs. L'affinage des modes de préparation est une étape extrêmement importante puisque cela permettra de fluidifier le processus et d'éviter une dégradation du scanner.
- Tester les différents modules de la chaîne de dématérialisation (numérisation, OCR, traitements d'images, indexation, archivage électronique). Chacun pourra faire l'objet de tests spécifiques visant à valider son niveau de qualité, sa rapidité d'exécution ainsi que la finesse des paramétrages.

Après avoir défini une démarche qualité, il faut s'attacher à déterminer des indicateurs. Quels sont les **indicateurs qualitatifs et quantitatifs** qui pourraient faire l'objet d'un suivi et d'une attention dès la mise en activité de la plateforme de dématérialisation ?

Sur le plan **quantitatif**, on peut d'ores et déjà isoler 4 indicateurs : 2 pour l'alimentation du fonds et 2 pour son exploitation.

- Premier indicateur quantitatif portant sur l'alimentation, celui du **temps de traitement d'un document**. Ce dernier étant bien entendu à comparer avec le temps de traitement actuel d'un document dans le système papier. Tout en précisant que le traitement d'un document numérisé comprendra une phase supplémentaire par rapport au système papier actuel : celle d'une indexation électronique très précise et très fine, alors que le traitement d'un article

papier se résume actuellement à un simple référencement manuel – une inscription directe sur la photocopie – suivi d'un classement dans une chemise et une boîte ;

- Un deuxième indicateur intéressant qu'il conviendrait d'observer repose sur **le nombre de documents qui seront injectés dans la GED par semaine (ou par mois)**, tout en le comparant au nombre moyen estimé de documents qui sont traités et photocopiés avec le système papier actuel (par semaine ou par mois). Cet indicateur serait un bon moyen d'évaluer le gain de productivité apporté par la dématérialisation. Evidemment cette évaluation ne devra porter que sur le dépouillement courant et au fil de l'eau des périodiques, sans prendre en compte les documents qui seraient éventuellement numérisés dans le cadre d'une opération de dématérialisation rétrospective ;

- Un troisième indicateur très intéressant pourrait concerner l'exploitation du fonds, et porterait sur **la fréquentation des utilisateurs**. Il serait révélateur de surveiller l'évolution du nombre de consultations électroniques par semaine (ou par mois) de la part des conservateurs. Ceci afin d'avoir un retour d'informations sur l'acclimatation et l'appréhension face à ce nouvel outil de travail. Néanmoins, cet indicateur ne pourrait être vraiment significatif qu'après avoir complètement dématérialisé le fonds, puisque, lors de la phase transitoire de numérisation rétrospective des dossiers thématiques (qui pourrait durer plusieurs mois, voire plus, selon les moyens et de l'engagement qui seront mis en œuvre), une partie des documents sera disponible uniquement au format papier ;

- Enfin, dernier indicateur quantitatif essentiel, toujours à propos de l'exploitation : **le temps d'accès aux documents électroniques**. Temps à comparer logiquement avec le temps d'accès actuel aux documents papier. Pour pouvoir comparer ultérieurement ces temps d'accès aux documents, entre le système papier et le système électronique, il serait nécessaire d'effectuer dès à présent des mesures avec le système de boîtes thématiques. Par exemple, commencer par décider arbitrairement d'une recherche ciblée, puis chronométrer le temps cumulé de la recherche dans les boîtes et de la remise en place des photocopies des documents trouvés ;

Le premier et le dernier des indicateurs cités plus haut sont particulièrement importants, puisqu'ils ont directement trait au gain de temps pour les documentalistes et/ou les conservateurs. Ils donneraient ainsi clairement une estimation du retour sur investissement après mise en œuvre de la dématérialisation.

Sur le plan **qualitatif**, on pourra se focaliser sur **le taux de satisfaction des utilisateurs**. Par exemple, actuellement une fiche est remise à tous les visiteurs extérieurs, et qui comprend

sur la dernière ligne une question au sujet des améliorations souhaitées. Afin de pouvoir établir quelques statistiques de satisfaction - à comparer ultérieurement et après mise en place de la GED -, il serait intéressant d'élargir les questions présentes sur ce formulaire. Cela permettrait de connaître l'avis des visiteurs sur la qualité de la recherche effectuée, et sur la pertinence des documents récoltés. Tout en restant bien conscient qu'une recherche électronique se trouve être forcément plus fructueuse, qualitativement et quantitativement, qu'une recherche dans des documents papier. Surtout dans l'hypothèse où la recherche en texte intégral est disponible dans le système de GED mis en place. Toujours sur le plan de la satisfaction et de l'adaptation des utilisateurs au nouveau système qui sera mis en place, un point important ne doit pas être négligé : l'ergonomie des interfaces. En effet, les utilisateurs des centres de documentation n'ont pas en général la pratique et l'expérience professionnelles d'un documentaliste pour la recherche d'un document, il s'agit donc d'optimiser au mieux l'ergonomie et le visuel du produit de GED. Ceci passe par une réflexion sur les interfaces de recherche et de consultation, et reste aussi conditionné par le choix de la solution logicielle qui sera adoptée. Cette dernière doit autoriser assez de souplesse et de liberté en termes de formatage et de paramétrage. Et après la mise en place de l'application de GED documentaire, une enquête pourra par exemple être menée pour recueillir les avis favorables ou défavorables des utilisateurs, ceci dans l'intention d'améliorer la solution en place. Concernant les conservateurs des musées des Arts Décoratifs, ils consultent déjà l'intranet pour l'accès aux bases des collections, mais leurs habitudes informatiques se limitent pour certains à cette seule consultation. La question reste donc posée quant à leur capacité d'adaptation au système numérique pour tout ce qui concerne la recherche dans les dossiers documentaires illustrant les collections.

3.10 Les aspects juridiques du projet

Les aspects juridiques du projet sont cruciaux et déterminants, ils conditionneront sa réussite mais surtout sa légalité. La problématique juridique tourne ici essentiellement autour du statut du « panorama de presse », notion apparue avec l'adoption en 1995 de la loi relative à la gestion collective du droit de reproduction par reprographie. Cette loi a permis d'éclaircir la législation, mais même rebaptisée « panorama de presse » de nombreuses interrogations continuent de se poser au sujet de ce dernier, tout particulièrement dans l'environnement électronique qui concerne directement le projet au musée.

La dématérialisation des dossiers papiers en version électronique s'apparente donc, sur le plan de la législation, au passage d'un panorama de presse papier à un panorama de presse dit « électronique ». Ce qui nécessite une réévaluation de la situation et de la gestion des droits de la part des responsables du centre de documentation.

3.10.1 Le droit de reproduction

La numérisation, tout comme la reprographie, est bien un acte de reproduction. Et, selon le code de la propriété intellectuelle (CPI), l'auteur d'une œuvre originale dispose d'un droit d'autoriser ou d'interdire la reproduction de son œuvre. Ce droit est bien évidemment cessible et négociable.

Il s'agit donc ici de démontrer que les centres de documentation ne peuvent pas bénéficier d'une exception au droit de reproduction, et que s'ils souhaitent procéder à des photocopies ou à des numérisations d'articles, ils ne peuvent éviter de procéder à des demandes expresses auprès des auteurs ou éditeurs.

Précisons tout d'abord la notion de « copiste » : depuis 1984, le copiste est défini comme étant celui qui, détenant dans ses locaux le matériel nécessaire à la confection des photocopies, exploite ce matériel en le mettant à disposition de ses clients. Les documentalistes des centres de documentation, qui font des photocopies pour les utilisateurs en interne ou pour les visiteurs extérieurs, sont considérés comme étant des copistes.

Le centre français d'exploitation du droit de copie⁴¹ (CFC) précise aussi que l'usage « privé » du copiste ne correspond qu'à un usage strictement personnel ou familial, et qu'il exclut tout usage interne, même destiné à une seule personne, qui serait fait dans un cadre

⁴¹ www.cfcopies.com

collectif. Les centres de documentation ne peuvent donc pas invoquer le bénéfice de l'exception de copie privée. Autrement dit : en termes de « droit de reproduction », le centre de documentation ne peut prétendre à aucune exception au droit patrimonial des auteurs concernant les photocopies faites en interne à un seul exemplaire et à finalité professionnelle, puisque la notion d'usage à titre personnel au sein d'une collectivité n'est pas reconnue comme exception au monopole de l'auteur.

De plus, dans le cas du centre de documentation des Arts Décoratifs, les documentalistes facturent les photocopies des articles de périodiques aux visiteurs extérieurs, ce qui peut être assimilé au regard de la loi à une forme « d'exploitation commerciale ». Dans le système actuel, le centre de documentation se fait en effet rétribuer pour ses prestations de photocopies. Un prix arbitraire a été fixé pour les photocopies en noir et blanc, un autre pour les photocopies couleur, et même un troisième pour les prises de vues avec un appareil photo. Or le musée n'a jamais considéré véritablement cette pratique comme relevant d'un quelconque « usage commercial » de ses dossiers, mais plutôt comme une forme d'indemnisation au titre des frais de fonctionnements et d'entretiens de ses imprimantes-photocopieurs. Pourtant, n'importe quel auteur ou éditeur concerné est en droit de considérer légitimement que ces copies sont faites à des fins commerciales. De plus la facturation des photocopies aux visiteurs est en réalité supérieure aux frais d'entretiens réels induits - achats de cartouches, maintenance et entretien courants, remplacements de pièces mécaniques, etc -, ce qui pourrait venir mettre en doute la bonne foi du centre de documentation en cas de procès pour acte de contrefaçon.

Cet aspect sera développé plus loin, mais on peut d'ores et déjà préciser que toute forme de copie faite à des fins commerciales exige non seulement une autorisation du CFC - dont ne dispose pas le musée actuellement puisque ce dernier n'a jamais contracté auprès du CFC - mais également une autorisation expresse des éditeurs.

3.10.2 Les dernières modifications apportées au « droit de reprographie »

Le droit de reproduction par reprographie est encadré par la loi de janvier 1995, qui organise sa « cession légale ». La reprographie fait ainsi l'objet d'une cession automatique et obligatoire au CFC. La loi impose le recours à la gestion collective, avec le monopole du CFC, mais les conditions de versements des droits et les tarifs ne sont pas fixés par la loi. Et dans ce cadre, l'ayant droit peut autoriser ou interdire une exploitation de son œuvre.

A titre de comparaison, on peut citer un autre système de gestion des droits d'auteur : la « gestion collective volontaire », qui est proposée aujourd'hui par le CFC pour les panoramas de presse diffusés sur intranet pour le compte des éditeurs qui l'ont mandaté à cet effet.

Dans la définition donnée dans l'article L122-10 du CPI, « la reprographie s'entend de la reproduction sous forme de copie papier ou support assimilé par une technique photographique ou d'effet équivalent permettant une lecture directe ». Cette définition exclut les procédés qui, comme les ordinateurs, exigent le recours d'un appareil pour pouvoir les consulter. Elle exclut les œuvres consultables uniquement sur support numérique ainsi que la transmission d'un article par courrier électronique. Elle englobe dans son champ les sorties papier des imprimantes lorsque les copies sont identiques aux originaux sur support papier.

Cette mise à jour en 1995 de la définition de la reprographie était destinée à faire une distinction plus nette entre la législation pour le papier et la législation pour le numérique, dans une période où l'information sur support numérique et en ligne commençait déjà à se développer. Mais aujourd'hui, avec par exemple la généralisation d'un format numérique comme le PDF, cette loi de 1995 apparaît comme relativement ambiguë et complexe. Ainsi, en plus des traditionnelles photocopieuses, le CFC indique dans ses contrats que « les appareils recourant à la numérisation d'une œuvre sur des supports optiques ou magnétiques en vue de la seule réalisation d'une copie identique à l'original » sont également concernés.

Pour essayer d'éclaircir ce point manquant de transparence et difficile à interpréter, on trouve ainsi dans la loi de 1995 les deux clauses suivantes :

- une copie numérique effectuée à des fins de stockage et destinée à une consultation exclusivement sur ordinateur n'entre pas dans le champs de la loi ;
- en revanche, un acte de numérisation réalisé en amont pour proposer des copies papier identiques à l'original serait en effet redevable de droits au CFC ainsi que toutes les sorties papier faites en aval ;

Voici un petit tableau récapitulatif :

Typologie des reproductions selon le support	
Champ de la loi de 1995	<ul style="list-style-type: none"> - Photocopie - Transparent - Télécopie - Sortie imprimante d'une copie identique à la version papier (par exemple une page en format pdf reproduisant l'article sous la même forme qu'en version papier)
Hors champ de la loi de 1995	<ul style="list-style-type: none"> - Numérisation pour lecture sur écran - Impression papier d'un article créé numériquement - Transmission par courrier électronique

Depuis 1995, le droit de reprographie intègre une condition supplémentaire : celle de l'affichage des mentions obligatoires. Ainsi sur chaque panorama de presse doit en théorie apparaître une mention rappelant que l'organisme bénéficie de l'autorisation du CFC.

Enfin, sur le plan de l'usage commercial des reprographies, et comme cela a déjà été abordé plus haut, la règle en la matière est la suivante : lorsque les copies sont faites par un tiers à des fins commerciales, une double autorisation est requise, celle du CFC ainsi que celle de l'auteur ou de son ayant droit. Et c'est au CFC qu'il appartient d'obtenir des ayants droit l'autorisation requise.

Classification selon la destination de la reprographie	
Usage collectif non commercial	CFC
Usage collectif commercial	Signalement au CFC qui est chargé d'obtenir l'autorisation des ayants droit

Il faut donc obligatoirement s'adresser au CFC pour l'informer de la commercialisation des copies. Mais, contrairement à la photocopie collective faite à des fins non commerciales, où aucune dérogation n'est possible, les ayants droits ont la possibilité de refuser un mode d'exploitation commerciale.

Sur ce point, et dans le cadre de ce projet de dématérialisation au centre de documentation, il faut analyser et traiter les alinéas 3.e et 8 qui sont inscrits dans l'article L122-5 du CPI, modifié par la loi n°2009-669 du 12 juin 2009 – art 21.

- Alinéa 3.e : l'auteur ne peut interdire la reproduction d'extraits d'œuvres des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche, dès lors que le public auquel cette reproduction est destinée est composé majoritairement d'élèves,

d'étudiants, d'enseignants ou de chercheurs directement concernés, que l'utilisation de cette reproduction ne donne lieu à aucune exploitation commerciale

Tout d'abord, les articles photocopiés et proposés aux visiteurs du centre ne sont pas des extraits d'œuvres mais bien des œuvres complètes. Ce sont en effet des articles entiers et non pas des extraits d'articles. Ensuite, le centre de documentation fait-il vraiment ces photocopies dans le cadre « d'un enseignement et de la recherche », ou, plus exactement, est-ce que le centre de documentation « illustre » un enseignement et une activité de recherche ? La réponse est clairement non.

Enfin, le fait de facturer les photocopies peut être assimilé à une forme d'exploitation commerciale.

- Alinéa 8 : l'auteur ne peut interdire la reproduction d'une œuvre à des fins de conservation ou destinée à préserver les conditions de sa consultation à des fins de recherche ou d'études privées par des particuliers, dans les locaux de l'établissement et sur des terminaux dédiés par des bibliothèques accessibles au public, par des musées ou par des services d'archives, sous réserve que ceux-ci ne recherchent aucun avantage économique ou commercial

Là aussi, la situation des Arts Décoratifs ne rentre pas dans le cadre de cet alinéa. Premièrement, les photocopies faites par les visiteurs sont systématiquement emportées par ces derniers, et non pas consultées uniquement « sur place » au centre de documentation. Deuxièmement, ce dernier facture les photocopies et en retire donc un intérêt commercial.

En fait, il n'existe qu'un seul alinéa dans cet article L122-5 qui donne presque toutes les autorisations en termes de reprographie. C'est l'alinéa 7 et il est très récent, mais il concerne uniquement les personnes handicapées.

Voici un extrait de cet alinéa :

L'auteur ne peut interdire la reproduction d'une œuvre par les établissements ouverts au public, tels que les bibliothèques, archives, centres de documentation et espaces culturels multimédia, en vue d'une consultation strictement personnelle de l'œuvre par des personnes atteintes d'une ou de plusieurs déficiences des fonctions motrices, physiques, ..

3.10.3 Les « panoramas de presse » de la documentation publicité

Il s'agit ici de bien discerner les notions de « revue de presse » et de « panorama de presse », afin de clairement établir et démontrer que les produits documentaires proposés dans les dossiers thématiques de la documentation publicité constituent bien des panoramas de presse et non des revues de presse.

Après de multiples jurisprudences, la revue de presse fait actuellement partie des exceptions au monopole de l'auteur reconnues par l'article L 122-5 du CPI, et doit répondre à trois critères :

- représenter une comparaison entre plusieurs articles écrits par un journaliste (reproduits intégralement ou non) ;
- avoir trait à un même thème ou à un même événement ;
- donner lieu (éventuellement) à un commentaire ou comporter un élément original fournissant ainsi la matière lui permettant d'être cité dans une autre revue de presse ;

En fait, un 4^{ème} critère discriminatoire a même été ajouté, puisqu'on a spécifié que la revue de presse ne pouvait être qu'une rubrique journalistique réalisée par un organe de presse.

Ainsi, la simple reproduction d'une série d'articles qui concernent par exemple une entreprise ou ses produits - comme c'est le cas dans les dossiers de la publicité où l'on trouve par exemple des chemises renfermant des articles consacrés aux marques ou et aux produits phares - ne répond pas aux critères de la Cour de cassation qui exige des commentaires - autrement dit une valeur ajoutée autour de ces articles. De même, on peut citer le cas des press-books - présents notamment dans certains dossiers documentaires de la documentation de la mode et du textile - qui, du fait de leur non accessibilité au grand public et de leur finalité, ne peuvent pas non plus être considérés comme des revues de presse.

Enfin, pour exclure définitivement la possibilité pour le centre de documentation de prétendre à la mise en ligne de « revues de presse », on retiendra que ces dernières doivent répondre à une exigence d'actualité mais aussi de réciprocité et de comparabilité. Or ce n'est pas le cas pour les dossiers thématiques de la publicité – ni, d'une manière générale, pour les dossiers documentaires des autres documentations du centre.

On doit donc considérer que les dossiers thématiques de la publicité sont donc alimentés par des panoramas de presse, et non des revues de presse. Dans le système papier actuel, les dossiers thématiques sont élaborés grâce à juxtaposition d'articles qui relève du régime des « compilations » et non de la courte citation. Ils proposent en effet des copies de textes reproduits dans leur intégralité, or contrairement à la revue de presse réalisée dans un contexte journalistique, un panorama de presse qui reprend l'intégralité ou des extraits

d'articles ne fait pas partie des exceptions au droit d'auteur admises par la loi. Ainsi, dans la situation actuelle et avant toute régularisation des Arts Décoratifs qui passerait obligatoirement par la signature d'un contrat auprès du CFC, ces reproductions au centre de documentation sous forme de photocopies sont susceptibles de porter atteinte à l'exploitation normale de l'œuvre, et de causer un préjudice injustifié aux intérêts de l'auteur.

Le CFC dissocie en fait les panoramas de presse qui sont réalisés par les services de communication sur l'actualité et les panoramas de presse documentaires qui sont thématiques – catégorie à laquelle appartiennent donc les dossiers thématiques de la publicité élaborés par le centre de documentation des Arts Décoratifs. Néanmoins, cette distinction n'a juridiquement aucun impact car les deux produits, s'ils correspondent à des assemblages d'articles photocopiés, sont redevables de droits. Ces deux types de panoramas font cependant l'objet de deux rubriques⁴² bien distinctes sur le site Internet du CFC, et sont encadrés par des contrats différents en termes de redevance et de tarification. Le premier produit est intitulé sur le site du CFC : « copies réalisés dans le cadre de panoramas de presse », et ne concerne pas les centres de documentation. Le deuxième produit, qui concerne en revanche directement notre étude, est intitulé « copies réalisées dans le cadre d'un centre de documentation ». Il est clairement indiqué que ce contrat couvre les photocopies réalisées pour les besoins documentaires, sous forme de copies à l'unité, de dossiers thématiques ou documentaires.

Dès à présent, il est très important de souligner que cette distinction entre deux types de panoramas n'existe que pour le format papier, et disparaît avec le support électronique – qui fait donc l'objet d'une seule et même rubrique⁴³ sur le site. Cela s'explique notamment par les modes de gestion des droits qui sont différents pour les deux types de support – « cession légale » pour le droit de reproduction par reprographie, « gestion collective volontaire » pour les panoramas de presse électroniques.

Dans l'hypothèse où le centre de documentation continuerait à exploiter ses dossiers thématiques au format papier après le lancement du processus de dématérialisation – notamment pendant la période de numérisation rétrospective du stock des dossiers -, il est donc fortement recommandé qu'il régularise sa situation et contracte auprès du CFC, afin d'éviter toute condamnation pour acte de contrefaçon. Or pour les panoramas à base de

⁴² <http://www.cfcopies.com/V2/cop/home.php>

⁴³ http://www.cfcopies.com/V2/cop/cop_ea_pano_elec_home.php

« copies réalisées dans le cadre d'un centre de documentation », le CFC propose en fait deux options : un contrat « type »⁴⁴, et un contrat « copies à fin de vente »⁴⁵.

Afin d'éviter tout risque d'assignation en justice, et au regard de la jurisprudence, le « contrat type » n'est pas recommandé pour le centre de documentation qui facture les photocopies aux visiteurs au titre d'une indemnisation des frais induits. Néanmoins, il faut souligner qu'il est assez regrettable – tout particulièrement pour les centres de documentation dont le budget est en général limité - de ne pas pouvoir disposer d'une définition précise et exacte de la notion de « fins commerciales », ce qui n'est pas le cas à l'énoncé de l'article L122-10 du CPI. En effet, comme cela a été sous-entendu plus haut, cette notion implique-t-elle toute rémunération y compris celle qui permet de couvrir les frais d'achat du toner ou du papier, ou implique-t-elle qu'il faille faire un bénéfice ? Sur ce dernier point la jurisprudence montre qu'il n'en est rien, puisque par exemple lors de son procès la Chambre de Commerce de Paris a été considérée comme ayant une activité commerciale, alors même que cette dernière était déficitaire.

Enfin, d'une manière générale, il faut également souligner que toutes les décisions qui ont été prises dans les récents procès ont démontré que l'autorisation des ayants droit était requise pour « les fournitures payantes » de copies d'articles, afin d'éviter une concurrence déloyale. Et l'exploitation de la copie relevant du régime de la gestion collective du CFC, c'est ce dernier qui doit demander l'autorisation aux ayants droit et bien reprendre dans son contrat les réserves éventuelles formulées par les éditeurs.

3.10.4 Les aspects juridiques des panoramas de presse électroniques

Après avoir cerné les enjeux et les difficultés juridiques concernant l'alimentation et l'exploitation des dossiers thématiques de la publicité en version papier, il faut ensuite procéder à une nouvelle analyse dans le cadre du projet de dématérialisation. Il faut donc se pencher sur les spécificités juridiques du « panorama de presse électronique », objet d'une législation sensiblement différente de celle du panorama de presse papier. La situation pour l'environnement électronique est en effet plus complexe, et notamment la loi de janvier 1995 ne permet pas d'étendre la cession automatique et obligatoire du droit de reprographie à la numérisation. Les règles pour le document électronique sont très différentes. Pour les reproductions sur support numérique, aucun système de gestion collective obligatoire n'a été mis en place, ce qui signifie que, dans le cadre d'une réalisation d'un panorama de presse

⁴⁴ http://www.cfcopies.com/V2/cop/cop_ea_centre_type.php

⁴⁵ http://www.cfcopies.com/V2/cop/cop_ea_centre_vente.php

électronique, les reproductions des articles sont, sauf exception, soumises à l'autorisation de chaque auteur et/ou éditeur.

Un nombre assez important d'entreprises de presse ont néanmoins déjà confié au CFC la gestion de leur droits numériques pour les panoramas de presse proposés sur des intranets. Actuellement, une liste comprenant plusieurs centaines de périodiques dont les éditeurs ont passé des contrats avec le CFC est disponible⁴⁶ sur le site de ce dernier. Un extrait de cette liste est consultable en annexe 4, à la page 178 de ce document. Mais cette liste n'englobe évidemment pas l'ensemble des publications disponibles sur le marché, et de plus elle subit périodiquement des modifications assez importantes - ajouts ou suppressions de nombreux titres.

Concernant la nature et les clauses des contrats, le système actuel comporte une « chaîne de relations contractuelles à trois niveaux » :

- un 1^{er} niveau lie les entités utilisatrices des panoramas de presse avec les prestataires de presse ou le CFC ;
- un 2^{ème} niveau lie les prestataires de services de panoramas de presse avec le CFC ;
- un 3^{ème} niveau lie les éditeurs de presse avec le CFC ;

Dans cette chaîne, on voit bien que c'est un contrat du 1^{er} niveau qui concernerait directement le centre de documentation.

Et il est très probable que, vu l'offre désormais très large et classique du CFC, c'est plutôt auprès de ce dernier que les Arts Décoratifs contracteraient. Il faudra tout de même bien vérifier si le CFC possède le droit d'autoriser une reproduction électronique de chacune des publications concernées par le panorama de presse.

En revanche, si les contrats de panoramas de presse électroniques proposés par le CFC ne satisfont pas le centre de documentation, il existe aussi d'autres possibilités. Premièrement, on pourra prospecter auprès de prestataires de presse ou d'agrégateurs de presse. Deuxièmement, il est également envisageable de procéder à une « exploitation directe », ce qui signifie contacter directement les éditeurs en vue de négocier et rédiger un contrat particulier et adapté aux besoins spécifiques du centre de documentation. Tout en s'assurant que ces éditeurs disposent bien de tous les droits numériques.

Précisons deux points cruciaux :

- les contrats passés avec le CFC sont des « contrats-type » dont les dispositions peuvent toujours être renégociées ;

⁴⁶ à l'adresse suivante :

www.cfcopies.com/V2/repertoire_numerique/PUBLICATIONS_CONTRAT_UTILISATEUR.pdf

- ce type de contrat n'autorise pas la reprographie, dont les modalités doivent être réglées auprès du CFC par un « contrat d'autorisation de reproduction par reprographie d'œuvres protégées » (on verra plus loin quelles sont les possibilités d'impression offertes par les contrats relatifs aux panoramas de presse électroniques) ;

D'une manière générale, l'autorisation de réaliser un panorama de presse électronique qui est accordée par le contrat vise :

- la numérisation et le stockage technique temporaire des articles sur un support informatique ;
- la transmission des reproductions réalisées vers des postes de consultation autorisés ;
- la représentation sur écran informatique desdits articles ;
- leur impression en un exemplaire par consultation à partir desdits postes ;

Le premier et le dernier point soulèvent immédiatement de sérieuses interrogations, et semblent déjà poser problème.

Pour être le plus complet possible, voici ci-dessous une description et un commentaire de chacune des clauses fondamentales du contrat-type, dont on verra que certaines sont donc apparemment assez incompatibles avec le projet de dématérialisation tel qu'il a été pensé et imaginé par le centre de documentation.

• **La définition de l'intranet et l'interdiction d'une rediffusion**

Le CFC définit un intranet comme étant un « réseau local informatique dont l'accès et l'usage sont réservés aux salariés ou personnels d'une même organisation ». Cette définition donnée par le CFC pose problème et n'est pas vraiment adaptée à la situation, puisque le centre de documentation des musées souhaite naturellement proposer une consultation électronique de ses dossiers thématiques, via son intranet, à ses visiteurs extérieurs. Cette clause est donc un réel obstacle, que rencontrent en fait bien d'autres structures. La définition du CFC concernant le réseau intranet est trop étroite et ne répond pas aux besoins concrets des centres de documentations.

• **L'article comme unité de base**

Contrairement aux contrats du CFC relatifs aux droits de reprographie, les contrats autorisant la diffusion de panoramas de presse électroniques considèrent que c'est l'article qui est l'unité de base du calcul, et non la page A4. De ce fait, les coûts induits pour un produit documentaire électronique s'avèrent souvent bien plus élevés que pour les versions papier.

• **Un stockage temporaire et une impression papier en un exemplaire**

Le contrat-type proposé par le CFC autorise la numérisation des articles et leur stockage sur un poste central, mais uniquement pendant la durée du contrat. Autrement dit, la rupture du contrat

impose que l'on détruise tous les fichiers stockés. Cette clause est donc très contraignante. En effet, vu le coût financier élevé de ces contrats relatifs aux panoramas de presse, vu la quantité d'articles qu'il est projeté de mettre en ligne, et vu que le centre de documentation souhaite entreprendre cette dématérialisation pour une conservation de type patrimoniale et à très long terme, alors cette clause de stockage temporaire apparaît comme contradictoire avec la philosophie du projet de dématérialisation au musée. Précisons que ce type de clause n'existe pas pour le papier, où il n'y a jamais eu de contrôle de ce type. Pour contrebalancer, le CFC autorise néanmoins l'impression en un exemplaire de l'article à partir de chacun des postes connectés à l'intranet. Dans le cas du centre de documentation, si on imagine par exemple que les consultations électroniques se fassent exclusivement sur les postes de la salle de consultation, c'est-à-dire que l'on se trouve dans une situation et un système où les postes personnels des conservateurs ne soient pas configurés pour accéder aux dossiers thématiques numérisés, alors, là aussi, le nombre autorisé de « une seule » impression papier par poste connecté paraît complètement absurde et inadapté aux usages imaginés par le musée.

- **Une limitation du nombre de titres concernés par le dispositif**

On a vu que l'autorisation accordée par les contrats du CFC concerne les articles parus dans les seules publications dont le CFC gère les droits. En cas de signature d'un tel contrat, les Arts Décoratifs devront savoir avec certitude ce qu'ils peuvent diffuser ou non de façon licite, et par là même se prémunir contre le risque de poursuites pour contrefaçon.

- **Des articles susceptibles d'être proposés par le CFC sans illustrations**

Certains titres de périodiques interdisent, faute d'avoir pu négocier les droits, la reproduction de photographies ou de graphiques qui accompagnent les articles. Là encore, ce genre de clause ne figure pas dans les contrats signés pour la reprographie, alors même que le coût d'achat pour les panoramas de presse électroniques est plus élevé que pour les versions papiers.

- **L'interdiction d'une réindexation**

Cette clause soulève également une très grosse interrogation, et paraît vraiment contradictoire avec les objectifs visés par la dématérialisation. Sur le plan de la législation, l'indexation correspond à la reprise des idées, et est libre. Pourtant, dans le paysage électronique tel qu'il est perçu et réglementé par le CFC, toute « réindexation électronique » est interdite. Or, la documentation publicitaire voit au contraire dans la dématérialisation une formidable opportunité de mettre en place une indexation numérique puissante et souple, afin d'optimiser le traitement documentaire et la qualité de ses dossiers thématiques. Une telle clause apparaît donc comme étant quasiment rédhibitoire à la signature d'un tel contrat par le musée, puisqu'une telle interdiction est totalement abusive et incompatible avec l'activité d'un centre de documentation dont l'objectif est d'insérer les documents dans un cadre de classement homogène.

- **Des quotas d'articles imposés dans certains contrats**

Il existe, dans la majorité des contrats proposés par le CFC, une limite au nombre d'articles consultables chaque mois. Limite fixée par le titre de presse lui-même, et non par le CFC. Cette clause paraît également inacceptable pour un centre de documentation qui établit sa sélection sur des critères qualitatifs et non quantitatifs.

- **Une tarification en fonction du nombre d'utilisateurs potentiels**

Le système de tarification des droits d'auteur mis au point par le CFC est établi en fonction de trois éléments : la redevance moyenne par article, le nombre d'articles mis en ligne, et le nombre de postes informatiques ayant potentiellement accès au panorama. Dans le contrat tel qu'il a été pensé par le CFC, pour tenir compte des personnes qui ne liront pas les articles, des abattements sont prévus. Mais cette tarification, fondée sur l'usage potentiel, est un paramètre très aléatoire et, malgré les abattements qui sont proposés, elle entraîne des coûts souvent excessifs.

Pour le projet des Arts Décoratifs, dans l'hypothèse d'une signature d'un contrat, il faudra donc au préalable bien déterminer les postes informatiques qui auront accès aux dossiers thématiques numérisés. Notamment il faudra faire un choix important : soit l'on souhaite donner l'accès à cette base d'articles électroniques seulement pour un ou deux postes informatiques de la salle de consultation, soit au contraire l'on préfère donner également l'accès à cette base pour tous les postes informatiques personnels de tous les conservateurs - et de leurs assistants ? - souhaitant consulter les dossiers thématiques de la publicité. Après avoir défini cette liste de postes autorisés, il faudra limiter et sécuriser l'accès à la base grâce à un système d'identification classique et géré par le logiciel de GED. Ainsi, en cas de contrôle du CFC, ce système informatique de GED mis en place et permettant un filtrage des accès permettrait de justifier les termes du contrat signé avec le CFC, dans lequel le nombre d'utilisateurs potentiels déclarés serait très inférieur au nombre total de personnes connectés à l'intranet des Arts Décoratifs.

Conclusion

Cette étude de faisabilité, portant sur le projet de dématérialisation d'une partie des dossiers documentaires du centre de documentation des musées des Arts Décoratifs, a eu pour objectif de dresser un panorama de la situation actuelle qui soit relativement complet, et de fournir des préconisations et des éléments d'information assez détaillés et approfondis. Ils permettront d'aider le futur responsable du projet dans ses choix et ses stratégies, et l'orienteront vers un numériseur et une solution logicielle de GED qui soient les plus adéquats et les plus appropriés possibles.

Dans ce mémoire ont été notamment abordés ou traités : les modalités de l'alimentation et du traitement documentaire des dossiers thématiques, les logiciels documentaires utilisés actuellement au centre de documentation, les besoins et les attentes de ce dernier en termes de numérisation et de stockage informatique, les dernières évolutions technologiques en matière de scanner et de GED – et leur impact sur les différentes étapes de la chaîne de traitement du document -, les nouvelles possibilités de travail collaboratif induites par la dématérialisation, les dernières modifications apportées aux contrats d'exploitation des panoramas de presse électroniques.

Cette étude ne peut toutefois pas prétendre à une quelconque forme d'exhaustivité, puisqu'elle a été rédigée par une personne extérieure à l'organisme. L'avantage étant de pouvoir ainsi apporter un regard plutôt neutre et objectif sur la situation, l'inconvénient étant de ne pas pouvoir recenser, deviner, ou anticiper absolument tous les besoins du centre de documentation - par exemple pour ce qui concerne les attentes précises des conservateurs des musées. Les promoteurs de ce projet au centre de documentation, qui portent l'idée d'une dématérialisation bénéfique, devront donc s'appliquer à finaliser et fixer définitivement ces besoins et ces attentes. Par exemple, il serait judicieux d'organiser rapidement des réunions à propos de ce projet avec tous les futurs utilisateurs concernés, ou bien de livrer à ces derniers de petits questionnaires, et ce avant la phase de rédaction précise d'un cahier des charges. Dans la perspective de cette tâche de rédaction, l'étude de faisabilité aura probablement permis d'apporter un certain éclaircissement, en même temps qu'une forme d'élan.

Toujours afin d'aider les décideurs et futurs responsables du projet, il est intéressant de citer le travail de l'Association des professionnels du numérique⁴⁷ (Aproged). Cette dernière a mené récemment des études dont il est ressorti très clairement que chaque projet de gestion de documents sous forme numérique était particulier, tenant naturellement compte des spécificités de chaque organisme. Néanmoins, il est apparu tout aussi nettement qu'un ensemble de questions étaient communes à tous ces projets. En partant de ce constat, et après

⁴⁷ www.aproged.org

un travail d'études et de synthèses auprès de plusieurs organismes, l'Aproged a ainsi pu réaliser un tableau de questions destinées à préciser la nature et les caractéristiques d'un projet de gestion documentaire, ainsi que son contexte technique, économique et organisationnel. Il s'agit en quelque sorte d'un guide de bonne conduite, qui est consultable en annexe 6 à la page 180 de ce document. Beaucoup de questions présentes dans ce tableau ont été abordées et traitées dans ce mémoire, mais d'autres appellent encore certaines réponses ou précisions, notamment sur le plan des budgets prévus ou sur l'organisation du pilotage du projet. Ce tableau de questions a vocation à aider la fixation des bases du projet stratégique que souhaite mettre en œuvre le centre de documentation, mais il permettra aussi, et plus simplement, d'apporter les preuves de l'intérêt et de la pertinence de la dématérialisation.

Parmi les bonnes règles pour la réussite du projet de dématérialisation, on rappellera que ce dernier possède trois volets d'importance quasi égale.

La technique constitue le premier volet, souvent le plus facile à gérer car il est essentiellement lié à des choix matériels et logiciels adéquats. On a vu dans cette étude de faisabilité que les choix d'une numérisation des documents textuels au format PDF, d'une diffusion de ces derniers sur des serveurs web XML, et d'un stockage sur des disques optiques numériques représentent la solution la plus adaptée et la plus simple – et ce dans l'optique d'un archivage à moyen ou long terme.

L'organisation constitue la deuxième facette du projet, déjà plus délicate à gérer car elle touche au fonctionnement même du métier de documentaliste. Et les principaux obstacles pour le centre de documentation sur ce plan organisationnel sont la difficulté à mettre en place – dès l'origine du projet – une véritable gestion de projet, ainsi qu'un manque de personnels.

Le troisième volet du projet est le facteur humain, le plus difficile à maîtriser et pourtant celui qui conditionne la réussite complète du projet. Dans ce mémoire, il a été ainsi clairement signalé que les difficultés culturelles pourraient être le premier frein à surmonter pour ce projet aux Arts Décoratifs. Concernant les documentalistes tout d'abord, leur métier est impacté par la dématérialisation dans le sens où des tâches de manipulation du papier comme le découpage et le rangement dans des boîtes sont aujourd'hui substituées par l'indexation électronique, la manipulation de logiciels pilotant les scanners et les processus OCR, ou encore la sauvegarde informatique. Il y a donc une nécessité de plus en plus évidente pour les documentalistes de disposer d'une double compétence documentaire et informatique, et cette mutation peut être ressentie par certains comme un réel bouleversement. Ensuite concernant les utilisateurs, d'une manière plus générale, beaucoup

sont encore attachés au toucher et au feuilletage manuel des dossiers papier – un processus non reproductible dans le monde numérique et donc irremplaçable.

Enfin, dans le cas du centre de documentation des Arts Décoratifs, la dématérialisation concernera – du moins pour ce premier projet - plus particulièrement des dossiers thématiques alimentés par des articles de presse, et l'on a vu que la faisabilité et la réussite du projet passent par une bonne connaissance et analyse de la législation actuelle en matière de panoramas de presse électroniques. Les aspects juridiques représentent donc ici un quatrième volet, probablement le plus critique. D'une manière générale, on trouve actuellement relativement peu de dossiers outils numériques en ligne sur le web, et l'explication tient clairement aux problèmes de droits – droit d'auteur et droit de reproduction - inhérents au support numérique. En effet, de nature assez différente de la majorité des projets qui sont actuellement entrepris – par tous les types d'organismes, mais portant en revanche essentiellement sur la numérisation des factures, des courriers et des dossiers clients –, les projets de dématérialisation de dossiers documentaires se heurtent à une législation trop stricte et manquant de transparence, alors que, paradoxalement, ce sont justement les différentes réglementations et lois qui ont permis un véritable décollage du marché de la dématérialisation - reconnaissance dans le code civil de la valeur probante des documents électroniques en mars 2001, loi sur la sécurité financière (LSF), décret du 18 juillet 2003 modifiant le code général des impôts et légalisant la dématérialisation fiscale de la facture. On soulignera notamment que le nombre de contrats d'autorisation signés par le CFC avec des centres de documentation pour la réalisation de panoramas de presse diffusés sur intranet est faible, en raison de tarifs trop élevés. Pour le projet du centre de documentation des Arts Décoratifs, en cas d'impasse, le dernier recours résidera dans une négociation directe avec les ayants droit.

Bibliographie

La bibliographie a été arrêtée au 2 octobre 2009.

Elle est conforme aux règles de ponctuation simplifiées de l'INTD, par application des normes sur les références bibliographiques :

- Z44-005. Décembre 1987. Documentation.
Références bibliographiques : contenu, forme, structure et à la norme.
- NF ISO 690-2. Février 1998. Information et documentation.
Références bibliographiques : documents électroniques, documents complets et parties de documents.

Son classement est thématique. Les références sont répertoriées par ordre alphabétique d'auteurs à l'intérieur de chaque thème.

Dématérialisation

ARCHIMAG. La dématérialisation : guide pratique. Paris, SERDA, 2008. 96 p. ISSN 0769-0975

BELAID A, CECOTTI H. La numérisation de documents : Principe et évaluation des performances [en ligne]. Article, Université Nancy 2 - LORIA, 2006 [consulté le 2 octobre 2009]. 46 p.

<<http://hal.archives-ouvertes.fr/docs/00/11/09/27/PDF/evaluation-OCR.pdf>>

BOUCHARD Karinne. Numérisation du patrimoine écrit et graphique : guide de projet, cahier des charges [en ligne]. L'institut du Patrimoine, 2003 [consulté le 2 octobre 2009]. 104 p.

<http://www.livre-poitoucharentes.org/section-bibliotheque/patrimoine/docsressourcespat/doc_download/18-numerisation-documents-ecrits-et-graphiques.html>

DINE Aurélie. Mise en ligne de dossiers documentaires : les techniques, les attentes et les craintes. Etude menée à partir de l'exemple du service documentation du quotidien économique Les Echos. 2005. 105 p. Mémoire DESS, INTD. 2005

FOURCHER Frédéric. La mise en oeuvre des programmes de numérisation dans la fonction publique : étude de cas du centre de documentation du Ministère des sports. 2003. 60 p. Mémoire DESS, INTD. 2003

MINERVA (Ministerial Network for Valorising Activities in Digitisation), Groupe de travail Minerva numéro 6. Guide des bonnes pratiques, version 1.3 : Identification des bonnes pratiques et des centres de compétence [en ligne]. 3 mars [consulté le 2 octobre 2009]. 65 p. <http://www.culture.gouv.fr/culture/mrt/numerisation/fr/eeurope/documents/bonnes_pratiques.pdf>

REMIZE Michel, BREBION Patrick. Dématérialisation : le nouvel espace. Archimag, février 2006, n°191, p. 21-29.

WESTEEL Isabelle, AUBRY Martine (textes réunis par). La numérisation des textes et des images : techniques et réalisations (actes des journées d'études organisées à la Maison de la Recherche les 16 et 17 janvier 2003). Villeneuve-d'Ascq, Éd. du Conseil scientifique de l'Université Charles-de-Gaulle, Lille 3, 2003. 190 p. ISBN 2-84467-050-4

Gestion Electronique de Documents

ALIDIERES Agnès. Du dossier documentaire papier au dossier documentaire numérique : le cas de la Joie par les livres. 2005. 105 p. Mémoire DESS, INTD. 2005

ARCHIMAG. GED et solutions logicielles : guide pratique. Paris, SERDA, 2002. 96 p. ISSN 1242-1367

BUSCAL Caroline. Conduire un projet de GED. Archimag, mars 2003, n°162, p. 40-42

CHAUMIER Jacques. La gestion électronique de documents. Paris, Presses universitaires de France, 1996. 123 p. Collection Que sais-je ?. ISBN 2-13-047766-6

COUDERC Bruno, PREVEL Laurent. La maîtrise du cycle de vie du document numérique. In Site de l'Aproged, Publications [en ligne]. Puteaux, Aproged, Octobre 2007 [consulté le 2 octobre 2009]. 4 p. <<http://www.aproged.org/index.php/Voir-details/Publications/8-Memo-La-maitrise-du-cycle-de-vie-du-document-numerique.html>>

DUPOIRIER Gérard. Technologie de la GED : l'édition électronique. Paris, Hermès, 1994. 228 p. ISBN 2-86601-432-4

MARY Sylvie. La GED. In Site de Business Process Management Services, Articles "Métier" [en ligne]. Paris, CERPEG, 12 mars 2009 [consulté le 2 octobre 2009].
<http://www.bpms.info/index.php?option=com_content&task=view&id=4521&Itemid=113>

NUTTIN Guillaume. Systèmes de Ged : évaluation de l'impact réel. Archimag, juin 2007, n°205, p. 37-38

PRAX Jean-Yves, LARCHER Simon. La gestion électronique documentaire. 3ème édition. Paris, Dunod, 2004. 341 p. ISBN 978-2100078912

RANDRIANARIJAONA Bodosahondra. Transformation d'une chaîne documentaire papier en chaîne numérique : le cas d'un service Questions / Réponses. 2008. 122 p. Mémoire Titre professionnel de niveau 1, INTD. 2008

ROUMIEUX Olivier, REMIZE Michel. Format : l'aventure du PDF. Archimag, mars 2003, n°162, p. 23-34

TOSCA CONSULTANTS, LENART, Michèle (étude réalisée par). La gestion documentaire : évolutions fonctionnelles et description de dix logiciels. Paris, ADBS Editions, 2004. 185 p. ISBN 2-84365-075-5

TRANCHART Mélanie. Mise en oeuvre d'une GEIDE et intérêt pour une activité de veille : le cas pratique de l'AFPA. 2007. 74 p. Mémoire Titre professionnel de niveau 1, INTD. 2007

VERNAY Jean-Philippe. Gestion de documents pour la fonction juridique : cahier des charges pour un projet de GED Veolia Environnement. 2008. 74 p. Mémoire Titre professionnel de niveau 1, INTD. 2008

Indexation électronique, métadonnées

BAUDRY de VAUX Marie, DALBIN Sylvie. Métadonnées et valorisation de l'information. Documentaliste – Sciences de l'Information, 2006, vol. 43, n°2, p. 144-147. ISSN 0012-4508

MENON Bruno. L'indexation à l'heure du numérique. Documentaliste – Sciences de l'information, 2004, vol. 41, n°6, p.340-342. ISSN 0012-4508

OCHANINE Hélène. Technologies de reconnaissance optique : des progrès notables à ne pas perdre de vue. Archimag, octobre 2006, n°198, p. 40-44.

Archivage électronique

AFNOR Z 42-013 : Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes

ARCHIMAG. L'archivage électronique : guide pratique. Paris, SERDA, 2007. 96 p. ISSN 1242-1367

Cornwell Management Consultants plc (traduction française par CHABIN Marie-Anne). MoReq : Model Requirements for the management of electronic records. In Site de l'ADBS [en ligne]. Commission Européenne, 2004 [consulté le 2 octobre 2009]. 140 p. <http://www.adbs.fr/servlet/com.univ.collaboratif.utils.LectureFichiergw?CODE_FICHER=1202984252455&ID_FICHE=3413>

RIETSCH Jean-Marc, CHABIN Marie-Anne, CAPRIOLI Éric. Dématérialisation et archivage électronique : mise en oeuvre de l'ILM, information lifecycle management. Paris, "01 informatique", Dunod, 2006. 207 p. InfoPro. Management des systèmes d'information. ISBN 2-10-050077-5

Les dossiers documentaires

QUAIRE Françoise, VAISSAIRE Clotilde. Les dossiers documentaires : dossier outil, dossier produit, dossier électronique. Paris, ADBS Editions, 2005. 61 p. L'essentiel sur... ISBN 978-2843650819

Informatique documentaire

RIVIER Alexis. Aide-mémoire d'informatique documentaire. Paris, Éditions du Cercle de la librairie, 2007, 156 p. ISBN : 978-2-7654-0953-3

Aspects juridiques

Association des professionnels de l'information et de la documentation, Commission Droit de l'information (rédigé par Michèle Battisti). L'accès à l'information électronique : le contrat en questions. Paris, ADBS Editions, 2004. 110 p. Sciences et techniques de l'information. ISBN 2-84365-068-2

BATTISTI Michèle. Le panorama de presse: aspects juridiques. Paris, ADBS Editions, 2006. 64 p. L'essentiel sur... ISBN 2-84365-084-4

BATTISTI Michèle. Le panorama de presse. Bulletin des Bibliothèques de France [en ligne]. 2007, n° 2. [consulté le 2 octobre 2009], p. 125-125.
<<http://bbf.enssib.fr/consulter/bbf-2007-02-0125-001>>

BATTISTI Michèle. Droit et traitement de la presse dans les centres de documentation. Bulletin des Bibliothèques de France [en ligne]. 2006, n° 5. [consulté le 2 octobre 2009], p. 26-30.<<http://bbf.enssib.fr/consulter/bbf-2006-05-0026-004>>

CFC. Le droit de reproduction par reprographie. In Site du CFC [en ligne]. Paris, CFC [consulté le 2 octobre 2009] <http://www.cfcopies.com/V2/leg/leg_repr_repr.php>

CFC. Le droit de reproduction électronique. In Site du CFC [en ligne]. Paris, CFC [consulté le 2 octobre 2009] <http://www.cfcopies.com/V2/leg/leg_repr_elec.php>

Collectif AAF, ABF, ADBDP, ADBGV, ADBS, ADBU, ADDNB, AIBM, FFCB. Le projet de loi sur le droit d'auteur et les droits voisins dans la société de l'information : attention, danger !. Documentaliste-Sciences de l'information, 2005, vol.42, n°1, p. 62-63

Commission Juridique de l'Aproged. Le devoir de conseil des professionnels du numérique. In Site de l'Aproged, Publications, Livre blanc [en ligne]. Puteaux, Aproged, 6 mars 2009 [consulté le 2 octobre 2009]. 28 p.
<<http://www.aproged.org/index.php/Voir-details/Publications/31-Livre-Blanc-Devoir-de-conseil-des-professionnels-du-numerique.html>>

GESTE. Guide pratique pour les utilisateurs de presse : panoramas de presse, résumés et archivages d'articles de presse en intranet. In Site du GESTE [en ligne]. Paris, GESTE [consulté le 2 octobre 2009]. <<http://www.geste.fr/pdf/Guide-pratique-utilisateurs.pdf>>

GESTE. La charte d'édition électronique. In Site du GESTE [en ligne]. Paris, GESTE [consulté le 2 octobre 2009]. <http://www.geste.fr/9_commissions/commi_pres1.htm>

LEMU Michèle. Diffuser un panorama de presse en ligne : quels droits ?. Documentaliste – Sciences de l'information, 2003, vol. 40, n°3, p. 220-225. ISSN 0012-4508

STRA Claire. Les panoramas de presse. Bulletin des Bibliothèques de France [en ligne]. 2004, n° 6. [consulté le 2 octobre 2009], p. 51-55. <<http://bbf.enssib.fr/consulter/bbf-2004-06-0051-007>>

SYNDICAT NATIONAL DE L'EDITION. La loi du 1er août 2006 (DADVSI). In Site du Syndicat National de l'Édition [en ligne]. Paris [consulté le 2 octobre 2009].
<<http://www.sne.fr/pages/les-enjeux/evolutions-du-droit-d-auteur/loi-du-1er-aout-2006-dadvsi-.html>>

Annexes

Annexe 1 : Numérisation et numériseurs

Critères de choix pour le numériseur de documents

Caractéristiques	Importance	Valeurs typiques ou remarques
Format des documents	Haute	A4,A3,A2,Microfilm...
Vitesse de numérisation	Haute	15 pages/mn, 200 pages/mn,...
Volume quotidien	Haute	1000 pages/jour, 20000 pages/jour...
Facilité de maintenance	Haute	Remplacement des rouleaux, des lampes
Résolution	Moyenne	300 dpi
Mode de numérisation	Moyenne	Chargeur automatique, à plat,...
Type de numérisation	Moyenne	Bitonale, Gris, Couleur
Particularité	Moyenne	Recto/verso, tourne-pages,...
Prix	Basse	Le prix est le dernier critère de choix

Tableau de questions permettant de s'orienter vers des choix technologiques

Etape	Choix
Acquisition numérique	Scannage : entraînement automatique ou manuel ? Quel scanner ? Quelle résolution ? Quel format de fichier en sortie ? A réaliser en interne ou en externe ?
Stockage	Disque dur, CD, DVD, DON WORM, Bande magnétique ?
Interprétation	Reconnaissance Automatique de Documents (RAD) ? Lecture Automatique de Documents (LAD) ? OCR ? (avec ou sans relecture humaine) / ICR ? Saisie manuelle ?
Traitement documentaire	Quels types et combien de fichiers à produire pour répondre aux contraintes d'exploitation ? Quel identifiant ? Quelles métadonnées ? Quelle structuration des métadonnées ? Quelle norme adopter ?
Exploitation	Quel format de fichier pour l'affichage, pour l'impression, pour la conservation ? Quels accès ? Quelles fonctions ? Quelle diffusion ? Quelle gestion des accès ? Portail ?

Tableau récapitulatif des définitions de numérisation adéquates en fonction des modes numérisation et des types de documents

Type de document	Mode de numérisation	Définition
Texte		
texte et dessin au trait	bitonal	600 dpi
texte illustré noir et blanc	bitonal	600 dpi
	niveau de gris	300-400 dpi
texte illustré couleur	couleur 24 bits	200-300 dpi
cartes en couleur	couleur 24 bits	200-300 dpi
Images		
images demi-tons	bitonal	600 dpi avec lissages à l'écran
images demi-tons en couleurs	couleur 24 bits	au moins 1.5 fois la linéature
Manuscrits et documents spéciaux		
dactylographie, impression laser, stylo bille	bitonal	au moins 300 dpi
stylo à encre, plume d'oie, feutre	bitonal	300 dpi
documents dégradés	bitonal	300 dpi
papyrus	couleur 24 bits	600 dpi

Tableau résumant l'adaptation fonctionnelle du numériseur au type de document

Numériseur	Caractéristiques	Type de document
Scanner à plat	Les plus courants	- Pages simples (ex : courriers)
Scanner à plat + chargeur	Permet la production de fichier multipages	- Ensemble de feuillets - Contrats
Scanners à main	- Peu de précision - Transportables	- Code à barre - Segments courts (identifiants)
Scanners à microformes	- Coûteux - Très spécifiques à la reprise d'archives anciennes	Reprise de fonds de films, microfiches, cartes à microfilm
Scanners à diapositives, à négatifs ou à transparents	Adaptés aux documents à traiter : taille, relief du cadre des diapositives, agrandissement (pellicule).	- Diapositives - Transparents
Les scanners à tambour	- Lents - Très haute résolution - Très chers	- Pré-presse - Industrie
Les scanners à livre ouvert ou scanner vertical		- Documents reliés et fragiles - Livres patrimoniaux
Bancs de reproduction	- Appareil photo numérique monté sur un support associé à des projecteurs, écrans de contrôle, éventuellement système de tourne - Lents	Pour documents fragiles, grands formats, 3D, volumes reliés

Caractéristiques du mode « Bitmap »

<p>Tif ou Tiff (Tagged Image File Format)</p>	<p>Standard (propriété d'Adobe) multi plateforme :</p> <ul style="list-style-type: none"> - scanners ; - photo numérique ; - imprimantes ; <p>Pas de perte de qualité.</p>	<ul style="list-style-type: none"> - Il stocke les caractéristiques d'une image dans un format balisé. - Standard très utilisé en reprographie, mais pas forcément intégré aux applications bureautiques (nécessite l'utilisation d'un plug-in). - Permet de stocker des images haute résolution sans déperdition de qualité. - Inconvénient : le poids du fichier. - Gère le multipage. 	<ul style="list-style-type: none"> - Recommandé pour l'archivage du fichier de référence. - Pas adapté à la consultation en ligne. - Recommandé pour la photo.
<p>Jpeg (Joint Photographic Experts Group)</p>	<p>Norme ouverte de compression numérique (taux de compression réglable).</p>	<ul style="list-style-type: none"> - Format de compression très efficace mais avec perte de qualité. - Très adapté au web (léger). - Adapté à la photographie (16,7 millions de couleurs, bon rendu des dégradés). - Ne gère pas les transparences. 	<ul style="list-style-type: none"> - Adapté à la consultation de photos en ligne. - Ne pas utiliser comme format de référence. - Ne pas utiliser pour des contrats.

Caractéristiques du mode « Vectoriel »

Postscript	Format mis au point par Adobe il y a une quinzaine d'années, devenu depuis un standard.	<ul style="list-style-type: none"> - Utilisé dans les imprimeries pour la production de tout le matériel imprimé : journaux, livres, plaquettes, brochures. 	<ul style="list-style-type: none"> - Plutôt destiné à l'impression ou à la photo composition plutôt qu'à l'archivage.
PDF Portable Document Format	Format ouvert, créé par Adobe comme évolution du format Postscript.	<ul style="list-style-type: none"> - Préserve les polices d'écritures, les images et les objets graphiques du format source. - Associe image et caractères. - Très répandu, le lecteur étant intégré à de nombreux appareils. - Gourmand en taille mémoire. - Interactif. - Gère le multipage. 	<ul style="list-style-type: none"> - Adapté pour une consultation à l'écran sur des plates-formes diverses (ordinateurs, PDA, mobiles). - Formulaires en ligne. - Conservation des contrats.

Tableau de facteurs pour la dématérialisation

Facteur	Options	Influence
Le type de document	<ul style="list-style-type: none"> - article de presse ou de revue ; - formulaire ; - livres ; - iconographie ; - manuscrit original ; - contrats ; - documents administratifs ; - plans ; - catalogues ; 	<ul style="list-style-type: none"> - La forme de l'objet à dématérialiser : Texte ou graphisme, imprimé ou manuel, couleur ou noir et blanc ; - Le contenu : Texte riche ou formulaire modélisé, structure forte ou non ; - L'usage futur : Preuve, durée de conservation, lecture in extenso, modification, achat, accès libre ou contrôlé, etc ;
Le format du document	<ul style="list-style-type: none"> - La taille : A4, A3, plan, format spécifique. - La couleur : de la typographie et du fond ; - Le papier : grammage, état ; - La pagination : document multipage, liasse, agraffage, reliure, collage, document façonné, plié, cartonnage, etc ; 	<ul style="list-style-type: none"> - Type et coût du scanner ; - Poids du fichier stocké ; - Ergonomie de consultation (taille de l'écran) ; - Faisabilité de l'OCR ; - Risque de dégradation de l'original ; - Automatisation de la dématérialisation ; - Coût de l'opération en raison des manipulations humaines à réaliser ; - Production de fichiers multipages ;

Les volumes à traiter et l'accroissement	Volume important	<ul style="list-style-type: none"> - Méthode et coût du scannage (automatisation ou sous-traitance) ; - Contrôles qualité par sondage ; - OCR et recherche plein texte plutôt qu'indexation manuelle ; - Classification ; - Système de stockage ; - Organisation de l'envoi des documents au scannage ;
L'homogénéité du fonds	<ul style="list-style-type: none"> - Documents de formats, polices et couleurs différents ; - Documents de natures différentes ; 	<ul style="list-style-type: none"> - Méthode de scannage (automatisation) ; - Qualité de l'OCR ; - Coût de l'opération ; - Pertinence de traiter ces documents dans un même projet ;
L'état et la fragilité du document initial	Documents fragiles ou précieux.	<ul style="list-style-type: none"> - Choix du scanner ; - Choix de la méthode ; - Choix de traiter les documents sur place (création d'un atelier, coût de la main d'œuvre en interne) ou d'externaliser ;
La juridiction liée au document	<ul style="list-style-type: none"> - Document non reproductible ; - Document preuve ; - Droit d'auteur ; 	<ul style="list-style-type: none"> - La faisabilité du projet peut être remise en question ; - Pérennité du support, du format de fichier, du logiciel de consultation ; - Signature électronique ; - Datage, identifiant unique ; - Identification des accès, des profils d'utilisateurs ;
L'exploitation prévue et la pérennité	<ul style="list-style-type: none"> - Pas de recherche sur le texte du document ; - Le document électronique sera lu ; - Le document sera modifiable ; - Le document ne doit pas être modifié ; - Le document doit être détruit ; - Le document ne doit pas être détruit ; 	<ul style="list-style-type: none"> - Le traitement OCR est-il utile ? - La lecture doit-elle se faire sur le document OCRisé ? - Prévoir le système de gestion des contenus et des versions ; - Prévoir le système de sécurisation ; - Prévoir les alertes (workflow) ; - Assurer la pérennité avec des recopies régulières sur nouveaux supports ;

Annexe 2 : Les supports de stockage

Les formats courants de bandes magnétiques

Média	Capacité (en gigaoctets)	Transfert (en megaoctets par seconde)
AIT-1	de 25 à 70	3
AIT-2	de 50 à 130	6
AIT-3	de 100 à 260	12
DAT	de 4 à 8	1,5
DAT DDS3	de 12 à 24	3
DAT DDS4	de 20 à 40	6
DLT	de 40 à 80	5
LTO	de 100 à 220	15
LTO2	de 200 à 400	30
S-AIT	500	30
SDLT	de 110 à 220	11
VXA-I	de 33 à 66	5
VXA-2	de 80 à 160	6

Comparaison des supports de stockage

Technologie	Destination	Avantages	Inconvénients
Disques magnétiques	Données courantes, accès rapide en ligne à plusieurs téraoctets	Très rapides	Prix, non amovibles, non archivables, facilement effaçables
Disques optiques non-réinscriptibles	Archives, données sécurisées, accès en ligne à plusieurs dizaines de téraoctets	Amovibles, infalsifiables, grande durée de vie, taille illimitée sur étagère, prix	Mode d'écriture, temps d'accès
Bandes magnétiques	Sauvegardes, accès en ligne à plus de cent téraoctets	Grandes capacités, amovibles, prix du support	Temps d'accès très lent, durée de vie assez faible, facilement effaçables

Annexe 3 : Présentation de l'offre logicielle EverSuite

Au cours de la période de réalisation de cette étude de faisabilité, l'éditeur Ever Team a organisé une journée de séminaire sur la dématérialisation, à laquelle il nous a conviés. Cette journée avait vocation à exposer les grands principes méthodologiques et techniques de ce type de projet. Elle fut illustrée par une présentation de EverSuite qui est le produit phare de l'éditeur, ainsi que par l'explication d'un projet de dématérialisation ayant été réalisé pour le compte du CNRS⁴⁸, et dont la maîtrise d'œuvre avait notamment été confiée à la société BT technologies, spécialiste de l'intégration - le CNRS et BT technologies étaient donc participants et co-organisateur de ce séminaire.

Ever Team est aujourd'hui le premier éditeur Européen de solutions intégrées de gestion de contenu d'entreprise (ECM Entreprise Content Management), et un des atouts de ses solutions est qu'elles sont fortement « orientées métier », couvrant l'ensemble du spectre fonctionnel de la gestion de contenu d'entreprise, culturel et patrimonial.

Actuellement, toute « application métier » peut être amenée à traiter, partager ou utiliser des documents issus de sources multiples, et l'information est désormais un composant à part entière des processus organisationnels. Cette information peut être structurée ou non structurée, et des outils de gestion électronique de documents peuvent être la clé pour l'intégrer dans les processus et les applications métiers de manière transverse au sein du système d'information.

Mais comment pourrait se traduire concrètement une telle gestion transversale de l'information au sein du centre de documentation des musées des Arts Décoratifs ?

Si on analyse l'existant et les bases de données des collections des différents départements du musée, gérées par les logiciels Micromusée et Mobytext de l'éditeur Mobydoc, on constate qu'il n'existe pas de système d'information globale pour toutes les collections. En effet, le musée n'est même pas doté de l'option MUM proposée par Mobydoc et qui permet l'intégration des deux logiciels spécialisés Micromusée et Mobytext en un système unique. Avec cette option, chaque logiciel garde pourtant son interface qui lui est propre et qui correspond au métier de l'utilisateur. Mais l'avantage crucial apporté par le MUM réside dans le fait que les données de chaque collection sont immédiatement accessibles à tous les services, et que par exemple les fichiers d'autorité et les thésaurus sont partageables.

⁴⁸ Centre national de la recherche scientifique.

Néanmoins, si sur le seul plan de la gestion des collections une option comme le MUM pourrait permettre et encourager une circulation transverse de l'information, en revanche, pour le reste du fonds - abonnements aux revues, dossiers thématiques, dossiers d'artistes, etc - l'implémentation d'un logiciel de gestion de contenu comme EverSuite (qualifié de « solutions métiers ») paraît indispensable. On a vu en effet que les logiciels de Mobydoc n'avaient pas vocation à gérer des documents aussi variés – en termes de contenu, de degré aléatoire de structuration, de nature, de description - et volumineux que ceux que l'on peut trouver dans les dossiers documentaires du centre de documentation. D'autant que le spectre fonctionnel de ces outils est très incomplet sur le plan de la gestion de contenu, puisqu'ils ne contiennent pas de module d'acquisition, ni de module de processus et de travail collaboratif.

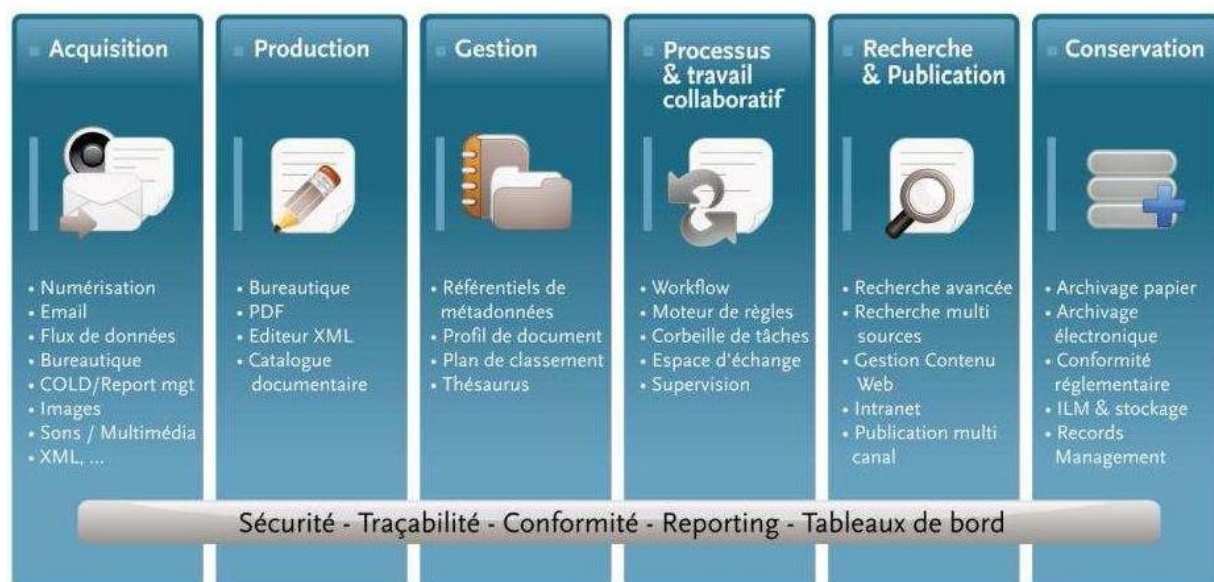
Un produit comme EverSuite dispose en revanche d'un spectre fonctionnel complet, permettant de déployer et d'adapter à chaque besoin et à chaque attente :

- une application de gestion documentaire
- une dématérialisation des processus
- une solution d'archivage

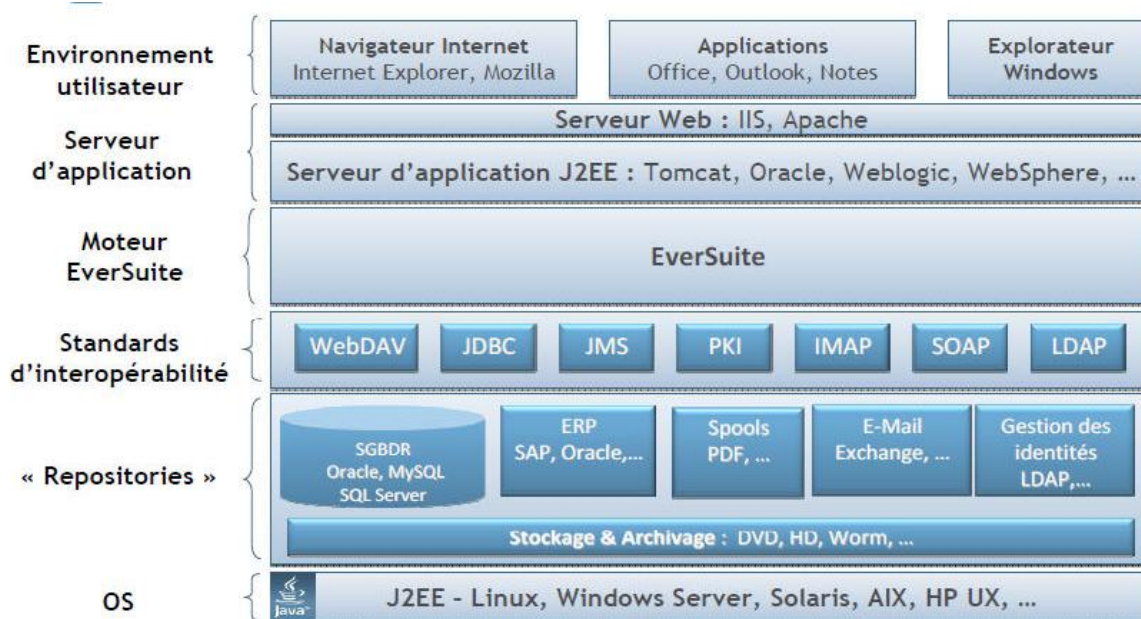
Ce qui rend possible :

- une valorisation du « capital informationnel et patrimonial »
- une généralisation et une facilitation de sa mise à disposition auprès de tous les utilisateurs et consultants potentiels

Voici la **décomposition du spectre** de EverSuite :



Voici l'**architecture technique ouverte** de EverSuite, qui assure une intégration facile au système d'information :



EverSuite, comme la plupart des offres ECM actuelles, présente 3 dimensions :

- Sur le plan purement informatique et technique, c'est une plate-forme d'ECM dite « **J2EE full SOA** ».

J2EE signifie Java 2 Enterprise Edition, et constitue la plate-forme (ou « framework ») Java 2 destinée à un usage professionnel avec mise en œuvre des serveurs. J2EE s'appuyant entièrement sur le Java, il bénéficie des avantages de ce langage, en particulier il assure aux applications une excellente portabilité et maintenabilité.

J2EE repose sur des composants distincts, interchangeables et distribués, ce qui signifie :

- qu'il est très simple d'étendre l'architecture ;
- que le système présentera une haute disponibilité et garantira une bonne qualité de service ;
- que la maintenance des applications est facilitée ;

Quant à la notion de SOA (Services oriented architecture), elle est apparue en 2003 dans la foulée de l'émergence des Web Services, et décrit une nouvelle manière d'implémenter un plan « d'urbanisation informatique ». Elle permet la réalisation des objectifs stratégiques grâce à l'alignement du système d'information sur les processus métiers de l'organisme. Plus globalement, la SOA a été en fait conçue pour jouer un rôle de langage de communication universelle entre environnements informatiques. Mais le point le plus important et à retenir en matière de SOA, c'est que sa mise en œuvre nécessite d'exploiter des outils de

développement et de déploiement particuliers, et d'appliquer un mode de gestion de projet spécifique, caractérisé notamment par un travail conjoint de la DSI avec les directions opérationnelles pour définir et optimiser les processus métiers.

- Deuxième dimension des offres ECM actuelles : les solutions dites **fonctionnelles**. Il s'agit en fait d'une décomposition de la solution en fonctions - traduites par des « modules » en termes de packages logiciels : capture, gestion de contenu et des processus, archivage, travail collaboratif, publication. On a vu au cours de cette étude de faisabilité que le projet de dématérialisation du fonds de la publicité ne nécessiterait pas la mise en place d'un véritable workflow, ni d'un moteur de règles, ni d'une corbeille de tâches. Au stade actuel de l'étude et du projet, la fonction ou module « processus et travail collaboratif » ne s'imposerait pas dans la solution éventuellement retenue.

- Troisième dimension des solutions ECM actuelles : leurs **spécificités métiers** leur permettant de s'adapter à chaque secteur d'activités.

Après avoir exposé ces 3 dimensions, il est également possible de synthétiser en 3 points la force d'une plate-forme de gestion de contenu comme celle d'Ever Team. Un 4^{ème} atout - relatif aux connecteurs - ne concerne pas directement la configuration actuelle au centre de documentation, néanmoins il est très important pour les entreprises et est souvent mis en avant par l'éditeur Ever Team.

• Une conception modulaire

Tout d'abord, le logiciel présente et propose une conception modulaire, afin de fournir des applications les plus personnalisées et les plus évolutives qui soient. La plate-forme comprend donc plusieurs dizaines de composants fonctionnels prêts à l'emploi, indépendants mais intégrés. Ce qui permet de couvrir l'ensemble des étapes de traitements du document, depuis la conception et la capture jusqu'à l'archivage et la destruction du contenu. Et c'est cette approche « orientée services » qui permet à chaque organisation de sélectionner uniquement les composants nécessaires. Pour le projet au centre de documentation, il n'y a par exemple aucun besoin réel de disposer d'un composant de conception ou de production numérique, puisque le fonds actuel est totalement papier à l'origine. De même il n'est pas nécessaire de mettre en place une organisation de destruction systématique de certains contenus numériques, puisque tous les documents numérisés injectés ont vocation à l'archivage sur le très long terme. En revanche, l'application réalisée et livrée par Ever Team pour un client peut tout à fait être enrichie dans le futur par des composants complémentaires si un besoin apparaît ou si une évolution fonctionnelle s'impose. Aucune application du style

de celle de EverSuite n'est figée et définitive. Ce qui présente un avantage majeur – ayant été souligné à plusieurs reprises par les représentants d'Ever présents au séminaire - pour les organismes qui prévoient une dématérialisation progressive « à petit pas ». Cette dernière est la stratégie adoptée par la majorité des candidats actuels à la dématérialisation et qui n'ont souvent pas les moyens - en temps, et financièrement – de refondre et de numériser en une seule fois tout leur fonds et toute leur organisation. Il est également ressorti de ce séminaire que la plupart des projets actuels reposent sur une stratégie de « clonage numérique » des processus papiers existants.

- **Performance et robustesse**

Deuxième atout d'un produit comme EverSuite : la robustesse et la performance. L'application peut en effet gérer un nombre quasi illimité d'enregistrements - plusieurs centaines de millions de pages pour certains projets -, dans le cadre d'un workflow comportant plusieurs dizaines de milliers de tâches, et concernant plus 150000 utilisateurs potentiels. Pour le projet des Arts Décoratifs, les performances exigées en termes de workflow et de nombre d'utilisateurs sont très loin de ces chiffres. En revanche, concernant le nombre d'enregistrements et de pages PDF à gérer, les besoins en robustesse et en performance pourraient rapidement devenir très élevés si la dématérialisation venait à s'étendre et toucher toutes les documentations du centre. D'autant que, en plus des dossiers documentaires, le centre de documentation a commencé à réfléchir sur l'opportunité d'une numérisation des dossiers d'œuvres, et aussi de certaines revues mortes qui sont aujourd'hui considérées comme un patrimoine historique et culturel. Ces dernières sont actuellement stockées dans les compactus, et il a été fait un premier recensement de celles qu'il serait immédiatement intéressant de dématérialiser. Parmi celles-ci on trouve : Vendre, Arts et métiers graphiques, L'estampe et l'affiche. Dans cette perspective, l'accumulation de tous ces documents numérisés commencerait ainsi à représenter une masse de fichiers très importante, nécessitant pour leur gestion une application très robuste.

Et, d'une manière générale, dans tout investissement informatique il est non seulement indispensable de bien cerner les besoins actuels mais aussi de savoir anticiper les ambitions qui pourraient rapidement apparaître dans un deuxième temps. Il est donc fortement souhaitable que le logiciel déjà acquis puisse supporter ces nouveaux besoins et répondre aux nouvelles attentes.

- **Sécurité**

Troisième force de EverSuite, une bonne gestion de la sécurité avec les protocoles SSL⁴⁹ et une extension de la mise en œuvre de la signature électronique à tous les composants de la plateforme. Néanmoins, les dossiers thématiques de la publicité ne présentent pas de documents à valeur probatoire, institutionnelle ou administrative. Et l'option de signature électronique, par exemple, ne figure pas parmi la liste des besoins établis, et ne sera donc pas exigée impérativement pour l'application de GED. Cependant, de la même manière que pour les exigences en matière de volumétrie des données manipulées, les besoins en sécurité pourraient être amenés à évoluer si l'utilisation de la GED se généralise à d'autres fonds. Il est donc nécessaire d'anticiper et de bien peser les options qui seront ou pas choisies initialement avec le logiciel. Par exemple, la décision de numériser une partie des dossiers d'œuvre induirait forcément une hausse des besoins en termes de sécurité, de contrôle et de surveillance. L'usage de signatures électroniques pourrait alors s'avérer indispensable, notamment pour assurer l'intégrité des documents numérisés et le scellement des données.

- **Les connecteurs**

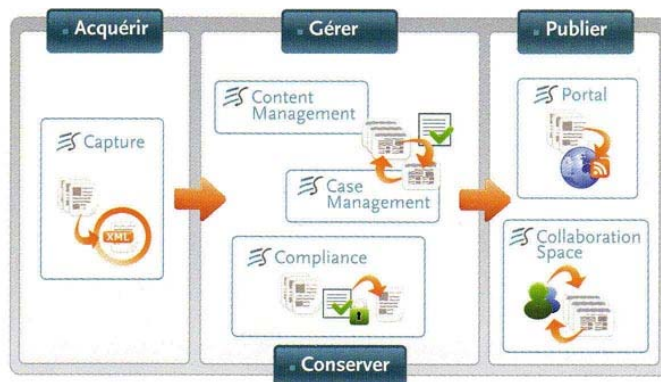
Enfin, EverSuite intègre un éventail de connecteurs qui sont fournis avec l'application, afin de faciliter son intégration au sein du système d'information existant. Concernant le système du centre de documentation, les seules véritables exigences en matière de « connecteurs métiers » concerneraient les applications métiers Micromusee et Mobytext, qui sont très spécifiques à la gestion de collections patrimoniales. Mais les solutions de GED actuellement sur le marché ne proposent pas par défaut de connecteurs avec ces applications de la société Mobydoc.

De manière plus générale, il n'y a pour ce projet au musée aucun besoin de dialogues avec des applications métiers du type ERP (Enterprise Resource Planning ou Progiciels de gestion intégrés) qui coordonneraient l'ensemble des applications du musée. L'étude révèle aussi qu'il n'existe aucun besoin de dialogues avec des applications du genre CRM (Customer Relationship Management ou Gestion de la relation client).

⁴⁹ Secure Sockets Layer.

Voici un document qui expose les solutions de gestion de contenu fonctionnelles proposées par le logiciel de la société Ever Team. On voit que ces solutions constituent une suite modulaire et intégrée, pouvant être combinées en fonction des besoins actuels et futurs.

Basées sur la plateforme **EverSuite**, les solutions ECM fonctionnelles constituent une suite modulaire et intégrée, permettant de résoudre les problématiques de gestion de contenu les plus fréquemment rencontrées. Elles peuvent être combinées en fonction des besoins actuels et futurs de chaque organisation.



EverSuite Content Management - Gestion des contenus et des processus métiers

EverSuite Content Management est un système complet de gestion de contenu et de dématérialisation des processus métiers associés. Intégrant nativement la notion de cycle de vie du

document, il permet de référencer, gérer, sécuriser et partager vos contenus au sein d'un référentiel unique et de mettre en place des processus de traitement adaptés à votre organisation.

EverSuite Compliance - Archivage et conformité réglementaire

EverSuite Compliance permet, à partir d'une solution unique, la gestion de l'archivage physique et de l'archivage électronique. Respectant les normes archivistiques et les exigences légales de traçabilité

et de conservation, elle intègre également des fonctions d'archivage sécurisées et propose des fonctions de remise en ligne et de recherche des éléments archivés.

EverSuite Capture - Acquisition de contenu

EverSuite Capture est une solution simple pour capturer tout type de flux de données entrant et circulant dans l'entreprise, paramétrer leur traitement en fonction de leurs caractéristiques et

les intégrer dans un référentiel unique et maîtrisé de gestion de contenu grâce à la mise en œuvre d'un format pivot XML.

EverSuite Collaboration Space - Travail collaboratif

EverSuite Collaboration Space permet de mettre en place des espaces de travail collaboratifs transverses. Elle simplifie et accélère l'échange et le partage de

tout type d'information au sein de l'entreprise, mais également avec des partenaires externes.

EverSuite Portal - Publication

EverSuite Portal permet de valoriser et diffuser des contenus produits et gérés par EverSuite au sein d'un portail d'entreprise. Grâce à **EverSuite Portal** chaque

utilisateur dispose d'un guichet unique d'accès personnalisé à une information agrégée et unifiée.

Il est important de souligner que la société Ever Team a également décliné sa plateforme ECM EverSuite J2EE en une autre version dédiée spécifiquement à la gestion de contenu culturel et patrimonial.

Cette solution, elle aussi « métiers », se dénomme **Flora** et se décline elle-même en 3 versions :

- Flora Library, qui offre une couverture fonctionnelle complète dédiée à la gestion et la valorisation de l'ensemble des collections des bibliothèques et des centres de documentation
- Flora SID, qui propose un service complet de production et de valorisation documentaire au sein du système global d'information
- Flora Archives, qui offre une couverture fonctionnelle complète de gestion des archives physiques adaptées aux organismes publiques

Même si EverSuite et Flora reposent sur une même plateforme technologique et disposent des mêmes standards décrits plus haut, il n'en reste pas moins que les solutions Flora sont donc des applications de gestion de contenu ayant été conçues à la carte pour les activités liées à la culture et au patrimoine.

Cependant, si on détaille précisément les fonctionnalités de ces trois produits Flora, on constate qu'aucun d'entre eux n'est vraiment adapté à la gestion de dossiers thématiques alimentés par des panoramas de presse. Pour notre projet de dématérialisation au centre de documentation, il est donc préconisé de s'orienter plutôt vers un produit de type EverSuite - tout en précisant que la bibliothèque des Arts Décoratifs exploite déjà Flora Library – qui est potentiellement capable de gérer tous les types de ressources documentaires, sans limite ni restriction – autrement dit les bases des collections, les dossiers thématiques, les dossiers d'artistes, les dossiers d'œuvres, le catalogage des ouvrages, le bulletinage des revues, etc.

L'exemple de la Documentation française avec son projet de Gestion de l'Information Publique sur Internet

Dans le cadre du projet de Gestion de l'information publique sur Internet (GIPSI), la Documentation française a choisi EverSuite pour améliorer la gestion de ses ressources documentaires hétérogènes – et en nombre croissants – et de son patrimoine documentaire, ainsi que valoriser l'information diffusée sur le portail tout en facilitant sa publication.

D'après les témoignages des documentalistes, les bénéfices immédiats d'un tel projet ont été :

- une simplification de l'outil de travail des documentalistes ;
- une meilleure qualité de l'indexation des documents ;
- un gain de productivité ;
- une valorisation de l'information actuelle et future ;

Afin d'apporter quelques précisions sur le plan technique, voici quelles étaient les spécifications :

- un serveur d'application Tomcat ;
- le SGBDR était MYSQL ;
- une intégration totale au système de publication Web existant ;

Concernant les enjeux de ce projet, et pour donner un ordre d'idée à propos des volumes, on peut préciser que la Documentation française est en charge d'un immense patrimoine, avec notamment près de 4000 rapports publics et plus de 75000 dépêches chronologiques.

Afin de corroborer les explications apportées plus haut et concernant les activités documentaires des Arts Décoratifs, il est utile de préciser que la solution retenue pour ce projet à la Documentation française fut une option dite « générique » de gestion des ressources documentaires, en privilégiant :

- la notion de paramétrage sur les développements spécifiques ;
- les standards (J2EE, XML, Oracle ou MySQL) sur les formats propriétaires ;

Cette solution générique peut donc évoluer ultérieurement par l'ajout de « briques » pour de nouveaux usages. La Documentation Française a ainsi décidé récemment la mise en œuvre d'une petite « bibliothèque numérique ».

La transition vers ce système EverSuite, répondant à l'ensemble des besoins fonctionnels de GED exprimés, a été réalisée avec souplesse grâce :

- à l'implication de tous les documentalistes ;
- à des réunions régulières avec les utilisateurs ;
- à une petite formation assurée par Ever Team auprès des documentalistes ;

Avec cet outil de GED, les documentalistes peuvent réaliser et valider les notices descriptives pour chacun des documents.

EverSuite, tout comme d'autres solution de GED puissantes, dispose notamment d'une « fonction de catégorisation » - fonction de plan de classement - qui peut même dans certains cas permettre un traitement automatique. En amont de sa mise en production, cela nécessite la structuration d'un arbre de catégories, c'est-à-dire un outil d'aide à la construction des plans de classement.

Dans le système mis en place à la Documentation Française, cette fonction est en fait couplé à un thésaurus qui est refondu et actualisé régulièrement, ce qui permet d'indexer les documents de façon détaillée.

Cette catégorisation se fait automatiquement grâce à l'identification de mots et de thématiques clés, et c'est un outil complémentaire à l'indexation manuelle. Pour le cas particulier des dépêches chronologiques qui sont très courtes, l'outil de catégorisation est devenu si performant que la qualité de l'indexation est désormais équivalente à celle réalisée par les documentalistes. Cet outil ne nécessite qu'une validation rapide de la part du documentaliste. Les documents sont donc classés automatiquement dans un plan de classement thématique et chronologique.

Sur le plan de la publication Web, la Documentation française se montre très satisfaite du choix de la puissance qui a été fait avec XML, ce qui permettra dans l'avenir de valoriser ce fonds patrimonial.

Au final, cette solution de GED a permis :

- un gain de productivité (en volume et en vitesse) ;
- une réactivité dans la mise à jour ;
- une simplification de l'outil de travail des documentalistes ;
- une meilleure qualité de l'indexation grâce notamment aux nombreux outils d'aide à la saisie, à l'assistance des tables et listes d'autorité, à l'outil de catégorisation automatique ;
- une organisation structurée et une vue plus conviviale de l'information grâce au plan de classement et à un thésaurus redevenu « vivant » ;

Annexe 4 : Extrait de la liste des publications dont le CFC gère les droits pour les reproductions et représentations numériques

Répertoire numérique professionnel général / Répertoire France

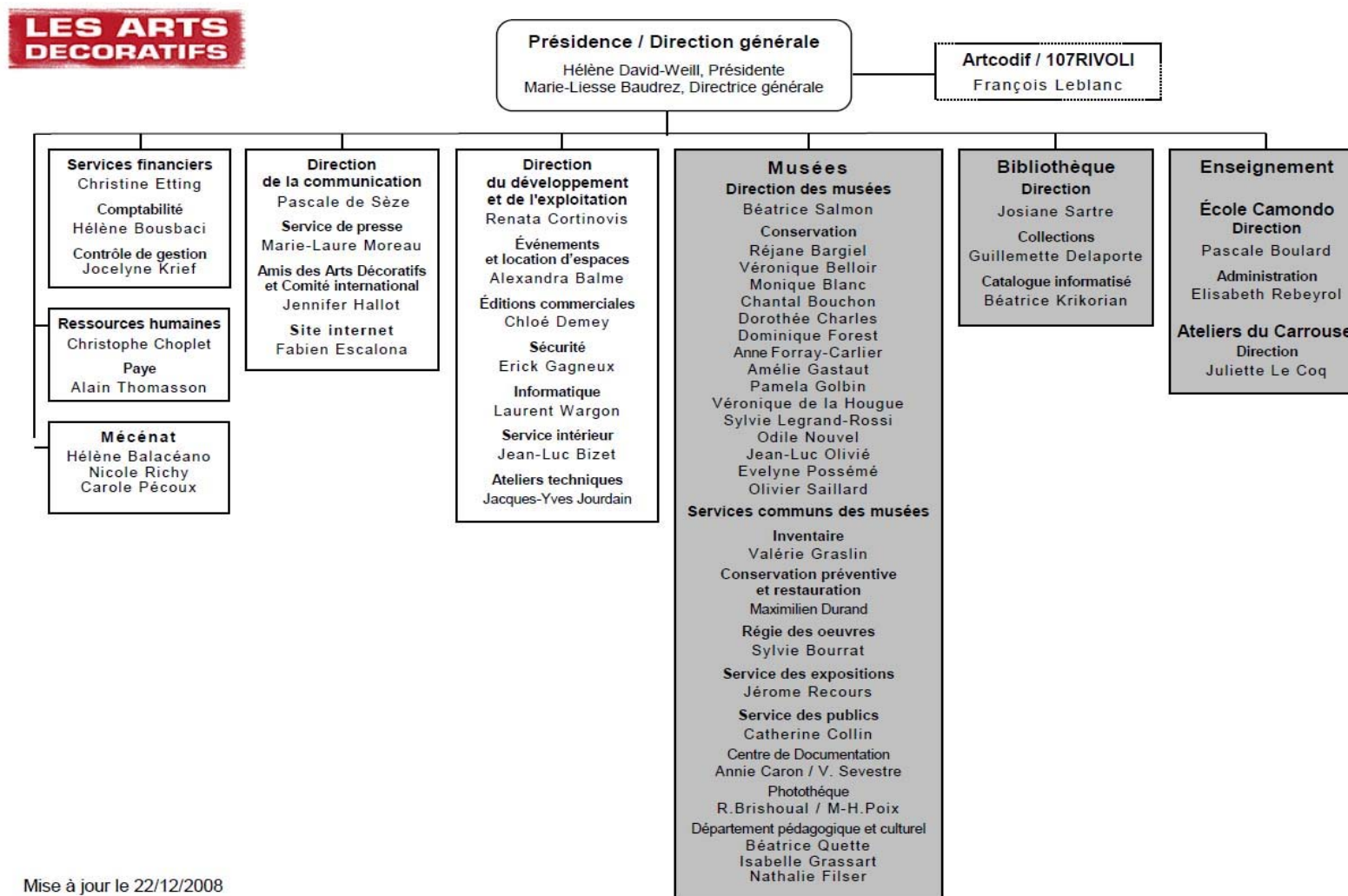
1 - Liste des publications dont l'utilisation est autorisée par le CFC

Publications	Redevance de référence par article en € HT PANORAMAS	Redevance de référence par article en € HT CLIPPING et COPIES INTERNES	Nombre maximum d'articles autorisés en moyenne par numéro du panorama de presse (moyenne calculée sur un mois)	Inclusion des infographies et photographies	Date d'effet du mandat	Date de dernière mise à jour
01 DSI	0,32	0,32	illimité	non	01-janv-04	29-avr-04
01 Informatique	0,32	0,32	5	non	01-janv-04	29-avr-04
01 Réseaux	0,32	0,32	7	non	01-janv-04	29-avr-04
01net.com	0,32	0,32	5	non	01-janv-04	29-avr-04
60 millions de consommateurs (mensuel + hors série)	0,16	0,24	5	infographies	01-janv-09	15-janv-09
60millions.mag	0,16	0,24	5	infographies	01-janv-09	15-janv-09
100% Exclusif	0,16	0,16	10	oui	01-janv-04	23-sept-04
3D Vision	0,32	0,48	illimité	oui	01-janv-08	17-mars-08
Abstract Cardiologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Dermatologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Diabétologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Gastroentérologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Gynécologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Neurologie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract ORL Pneumo Allergo	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Pédiatrie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Psychiatrie	0,32	0,32	5	non	01-janv-03	26-avr-03
Abstract Rhumatologie	0,32	0,32	5	non	01-janv-03	26-avr-03
acuite.fr	0,32	0,48	illimité	oui	01-janv-08	17-mars-08
Acteurs Publics (+ hors-série)	0,64	0,64	illimité	non	01-janv-04	15-oct-04
acteurspublics.com	0,08	0,08	illimité	non	01-janv-04	15-oct-04
Action Commerciale (L')	0,32	0,64	illimité	non	01-janv-05	04-mars-05
Actionco (Newsletter)	0,32	0,64	illimité	non	01-janv-07	17-déc-07
actionco.fr	0,32	0,64	illimité	non	01-janv-07	17-déc-07
Activeto	0,64	0,64	5	non	01-janv-04	05-avr-04
Actu Labo	0,64	1,60	5	oui	01-janv-06	21-juil-06
Actualité Juridique Droit Administratif	0,64	0,64	3	oui	01-janv-03	01-juil-03
Actualité Juridique Droit Immobilier	0,64	0,64	5	oui	01-janv-03	01-juil-03
Actualité Juridique Famille	0,64	0,64	5	oui	01-janv-03	01-juil-03
Actualité Juridique Fonctions Publiques	0,64	0,64	5	oui	01-janv-03	01-juil-03
ACTUALITES PHARMACEUTIQUES	0,32	0,32	5	oui	01-janv-08	17-juin-08
ACTUALITES PHARMACEUTIQUES HOSPITALIERES	0,32	0,32	5	oui	01-janv-08	17-juin-08
ActuBoissons	0,32	0,32	5	oui	01-janv-05	01-févr-05

CFC/Répertoire numérique professionnel général/Répertoire France - anx 1 - 2009.07.03

Web 1/54

Annexe 5 : Organigramme des Arts Décoratifs



Annexe 6 : Le guide de bonne conduite⁵⁰ de l'Aproged

N°	Questions
1	En quelques mots, pouvez-vous décrire en quoi consiste le projet portant sur la "gestion des documents sous forme numérique" ?
2	Quels sont les cinq principaux objectifs poursuivis ?
3	Le projet est-il considéré comme critique ? (expliciter les raisons, que la réponse soit positive ou négative)
4	Quels sont les documents (et/ou informations) visés par le projet ainsi que les services internes concernés ?
5	Quelles sont les réglementations génériques ou spécifiques impliquées du fait du projet ?
6	Etes-vous confronté à des obstacles juridiques majeurs ? Si oui lesquels ?
7	Quels référentiels et normes techniques sont retenus dans le cadre du projet et pour quelles raisons principales ?
8	Quelles solutions sont retenues (internes/externes – matérielle, logicielle, applicative, etc) ?
9	Quels sont les critères essentiels de choix pour retenir telle ou telle solution (couverture fonctionnelle, pérennité, évolutivité, indépendance (ex : logiciel libre), niveau de technicité, coût, conformité, etc) ?
10	Comment est organisé le pilotage du projet ? Quels sont les profils des personnes internes (et/ou externes) qui mènent le projet ? Quels sont les profils des personnes internes (et/ou externes) concernées par la mise en œuvre des « solutions » retenues à l'issue du projet ?
11	Quelle est la durée prévue entre la « prise de conscience » / volonté d'agir et la décision de lancement du projet ? Quelle est la durée prévue entre la décision de lancement du projet et le lancement effectif du projet ? Quelle est la durée prévue entre le lancement du projet et la mise en œuvre opérationnelle puis éventuellement le déploiement ?
12	Quels sont les budgets prévus pour réaliser le projet, par principal poste (étude, acquisition, services, formation, etc) ? Quels sont les gains envisagés du fait du projet puis de sa mise en œuvre, par principal poste ?
13	Quelle documentation a été produite du fait du projet ? (politique de gestion du cycle de vie du document numérique, politique d'archivage, plan de classement des documents, etc) ?

⁵⁰ Source : Direction générale de la modernisation de l'état. Gestion des documents numériques et de leur contenu. In Site de l'Aproged, Publications, Livre blanc [en ligne]. Puteaux, Aproged, 6 octobre 2008 [consulté le 02 octobre 2009]. *Guide de bonne conduite*, page 17.
<http://www.aproged.org/index.php/Telecharger/Publications/7-Livre-Blanc-Gestion-des-documents-numeriques-et-de-leur-contenu.html?chk=2d6c0c9f15ddb54085a5fe40093059a3&no_html=1>

14	Estimez-vous avoir suffisamment d'informations (notamment d'origine publique) ou d'aide pour mener le projet ?
15	L'appréhension du projet par les utilisateurs finaux a-t-elle été facilitée ? (communications, plan de formation, etc), est-elle suffisante ?
16	Pensez-vous être confronté à des obstacles majeurs pendant le projet ou durant la mise en œuvre des solutions, lesquels ?
17	Quels sont les résultats attendus du projet par les décideurs : <ul style="list-style-type: none"> - Volonté de productivité, - Mise en conformité légale et réglementaire, - Pression du marché ou d'entreprises partenaires, - Meilleure gestion « métier », - Etc.
18	Quelles sont les perspectives et les étapes suivantes du projet ?
19	Quels sont les deux ou trois principaux conseils que vous pourriez souhaiter avoir ?

Annexe 7 : Extrait du plan de classement des dossiers thématiques de la publicité

Achat d'espace
Affichage
Affichage dans le métro
Affichage mobile
Affichage route
Affiche de librairie
Afficheurs
Agriculture
Alimentation
Alimentation animale
Alimentation pour animaux
Ameublement
Animal et pub
Architecture commerciale
Art et pub
Automates
Automobile
B to B
Bande dessinée
Banque
Beauté
Boisson alcoolisée et non alcoolisée
Campagne collective
Célébrités dans la pub
Chromos
Cibles publicitaires
Cinéma : media planning
Cirque
Co-branding
Communication culturelle
Communication de crise
Communication institutionnelle
Consumer magazines
Conte de fées/mythologie/imaginaire
Création publicitaire
Danse
Design : design intérieur/fonctionnel, design objets, signalétique/urbanisme, théorie et stratégie du design/packaging, techniques et matériaux d'emballage
Détournements publicitaires
Développement durable
Diable à l'affiche
Disruption
Distribution

Écologie
Édition
Effets spéciaux (film publicitaire)
Efficacité/mémorisation publicitaire
Électroménager
Énergie
Enfant et publicité
Équipement de la maison
Est dans la pub
Expositions/manifestations
Fausse pub/pub fantôme
Femme dans la pub
Flacons de parfums (design)
Géomarketing
Glocalisation
Graphisme et publicité
Histoire (L') dans la pub
Histoire de la publicité
Histoire du graphisme
Historique de l'affiche
Homme dans la pub
Homosexuels (cibles publicitaires)
Humour et pub
Identité visuelle
Imprimerie/imprimeurs
Industrie du luxe
Industrie pharmaceutique
Informatique/bureautique
Intérêt collectif
Internet : média planning, fournisseurs d'accès, format publicitaire, mesure d'audience, logiciel anti-pub, publicité sur Internet
Jeunes (cibles publicitaires)
Jeux de hasard
Jeux jouets
Lancement de produits
Langage publicitaire
Loterie nationale
Manifestations publicitaires
Marionnettes
Marketing éthique
Marketing mobile
Marketing olfactif
Marketing sensoriel
Marques de distributeurs
Mécénat et sponsoring
Media planning : mesure d'audience, placement de produits sur différents supports
Media : chaînes de télévision, station de radio et presse
Mode
Mondialisation /globalisation

Annexe 8 : La structure des boîtes d'archives

Voici quelques photos illustrant la structure interne d'une boîte d'archives.

Il a été pris ici comme exemple la boîte portant le titre principal "Boissons non alcoolisées 2" (2ème des deux boîtes d'archives consacrées à ce thème).

Il est indiqué sur l'étiquette de cette boîte une série de titres secondaires, correspondant aux sous thèmes de la boîte :

- Sodas
- Jus de fruits
- Sirops
- Apéritifs sans alcool
- Boissons énergétiques

Photo illustrant en exemple deux articles découpés au ciseaux (et non recollés sur des feuilles A4). L'un à partir d'une photocopie, l'autre à partir d'une revue originale.



Photo illustrant des sous-chemises plastique transparentes dans lesquelles sont rangées des photocopies d'articles.

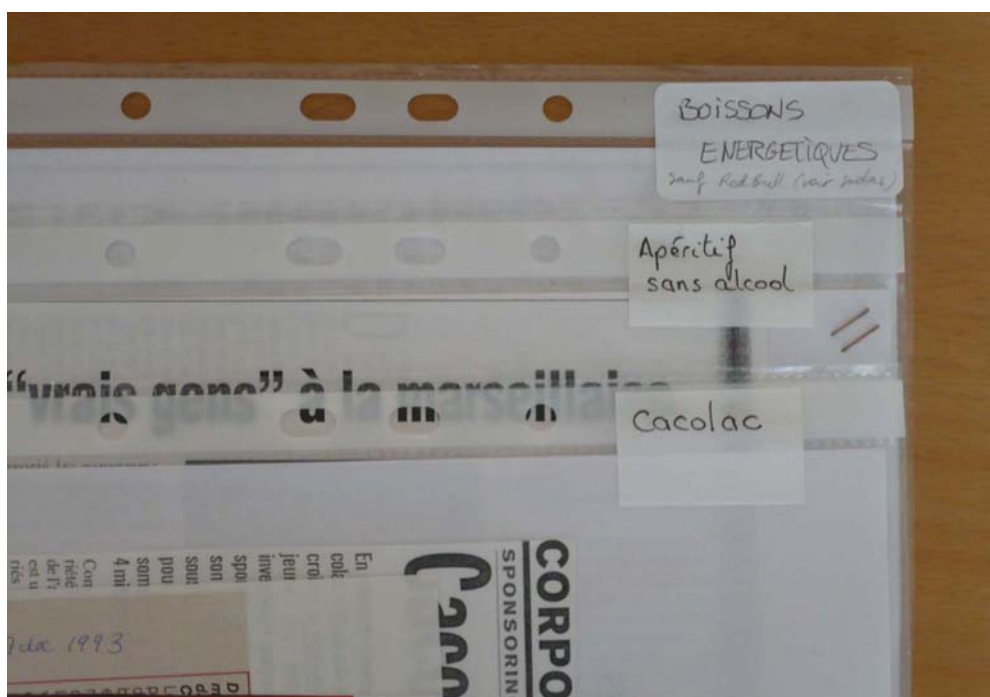


Photo illustrant la chemise papier intitulé "Jus de fruits" dans laquelle sont rangées diverses sous-chemises plastique transparentes.

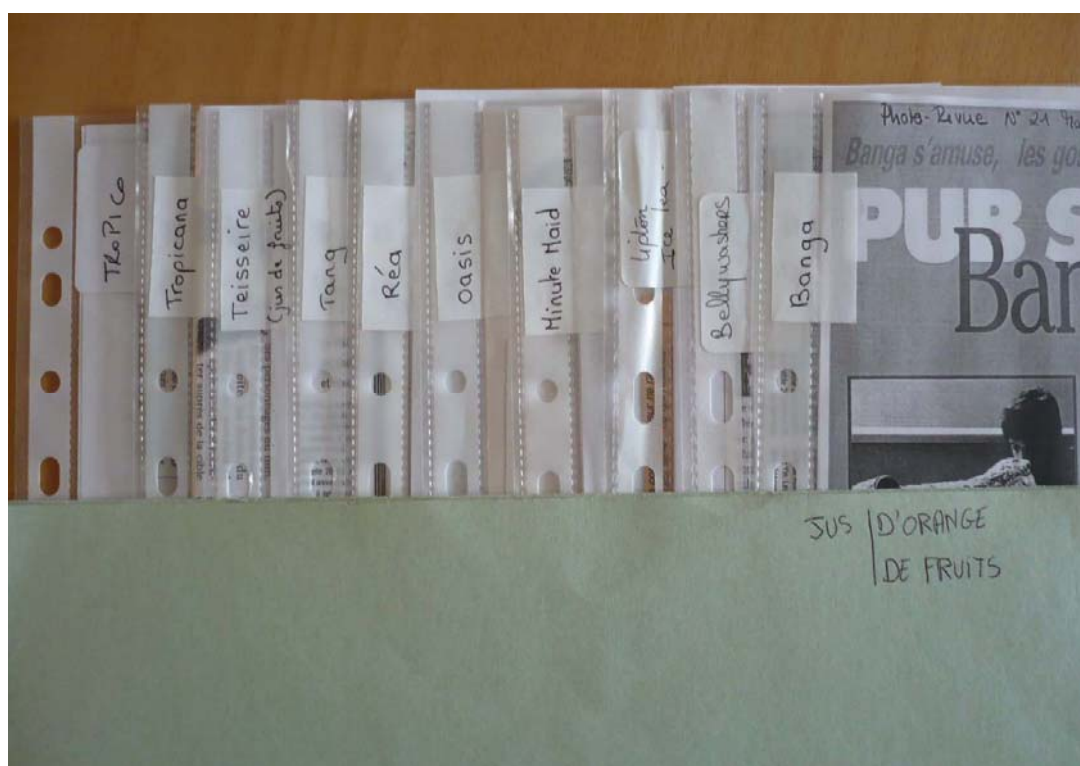


Photo illustrant diverses chemises papier qui sont rangées dans la boîte d'archives.



Photo illustrant l'intérieur de la boîte d'archives.



Photo illustrant les deux boîtes d'archives intitulées "Femme et publicité" et "Boissons non alcoolisées 2".



Annexe 9 : Comparaison entre un système de GED et un SAE

Un système de GED	Un Système d'Archivage Electronique
<ul style="list-style-type: none">- permet la modification des documents et la production de plusieurs versions ;- peut permettre la destruction des documents par leurs auteurs ;- peut comporter la gestion de durées de conservation ;- peut comprendre une structure organisée de stockage, sous le contrôle des utilisateurs ;- est à priori dédié à la gestion quotidienne des documents pour la conduite des affaires.	<ul style="list-style-type: none">- interdit la modification des documents ;- interdit la destruction de documents en dehors d'un contrôle strict ;- comprend obligatoirement un contrôle rigoureux des durées de conservation ;- comprend obligatoirement une structure rigoureuse de classement (le plan de classement), gérée et contrôlée par l'administrateur ;- peut faciliter les tâches quotidiennes mais est aussi destiné à la constitution d'un fonds sécurisé des documents probants de l'organisation ;

Annexe 10 : Tests de numérisation et tests OCR

Des tests de numérisation ont été effectués avec :

- le modèle de scanner/imprimante (autrement dit « imprimante multifonctions ») grand public **Canon Pixma MP170** ;
- le logiciel de numérisation (pilotant le scanner) **Canon MP Navigator 2.0** ;
- le logiciel professionnel **ScanSoft OmniPage Professionnal** pour le traitement OCR ;

Cette phase de tests a été effectuée à partir d'un échantillon de 19 articles extraits des dossiers thématiques de la publicité. Cette sélection a été spécialement élaborée avec le souci de représenter le plus largement possible la diversité des documents (taille de caractères, qualité de reprographie, vieillissement du papier, alignement des articles, etc) présents dans ces dossiers documentaires.

Le logiciel de numérisation a été utilisé en « mode avancé », permettant ainsi un maximum de paramètres manuels.

Concernant les paramètres d'entrée, on a procédé au choix à des numérisations en couleur ou en niveaux de gris.

Concernant les paramètres de sortie, on a procédé au choix à des numérisations à 200, 300, ou 400 dpi.

Concernant les paramètres d'image, toutes les numérisations ont été effectuées avec la même configuration suivante :

- tonalité automatique activée ;
- accentuation activée ;
- élimination du moiré désactivée ;
- aucune réduction des imperfections ;
- aucune correction de l'atténuation ;
- aucune correction du grain ;
- aucune correction du rétroéclairage ;

Par manque de temps, on voit donc que certains paramètres - nécessitant un réglage très fin et entièrement manuel, comme la correction du grain - de la numérisation n'ont pas pu être parfaitement optimisés. Néanmoins, la configuration décrite ci-dessus a tout de même permis de tester de façon très fiable le niveau de performance du traitement OCR. Les optimisations automatisées de la tonalité et de l'accentuation du contraste assurant en effet de procéder à une reconnaissance de caractères dans des conditions satisfaisantes de traitement.

De plus, afin que le traitement OCR puisse s'opérer de façon correcte, tous les documents présentant des articles non alignés – autrement dit les photocopies ayant été faites avec un manque de soins - ont été traités manuellement afin de réaligner les textes. Ce travail a été assez lourd puisque aucune option de correction automatique des désalignements n'est disponible dans le logiciel Canon MP Navigator.

Enfin, beaucoup de documents ont également été recadrés lors de la numérisation. Ce traitement était lui facultatif et n'a été fait, dans la majorité des cas, qu'avec un simple souci d'esthétisme. Ce recadrage n'ayant en effet aucun réel impact sur la performance de l'OCR. En précisant qu'aucune option de recadrage précis n'est disponible non plus dans MP Navigator. Celui-ci a donc dû être effectué manuellement avec Adobe Photoshop.

Cette phase de tests aura donc permis de confirmer que l'option de recadrage manuel précis et celle de correction automatique de désalignements paraissent indispensables dans la solution logicielle qui sera retenue par le centre de documentation des musées.

Sur le plan du format de sortie des fichiers, tous les documents ont été numérisés en PDF.

Ces fichiers PDF ont ensuite été soumis à un traitement OCR avec le logiciel ScanSoft OmniPage Professionnal (version 16). Une évaluation de la performance de l'OCR a ainsi pu être réalisée pour chacun des documents de l'échantillon de départ.

Pour chaque test, une indication est donnée sur la taille du fichier résultant de la numérisation. Ce sont des informations importantes pour l'évaluation des capacités de stockage informatique nécessaires pour le projet.

Quelques captures d'écran du logiciel ScanSoft OmniPage sont disponibles dans cette annexe, afin d'illustrer un peu toutes les possibilités offertes par ce type de traitement.

Document 1



Le premier test correspond à une numérisation d'un article de la revue Stratégies, directement à partir de l'original et non pas d'une photocopie.

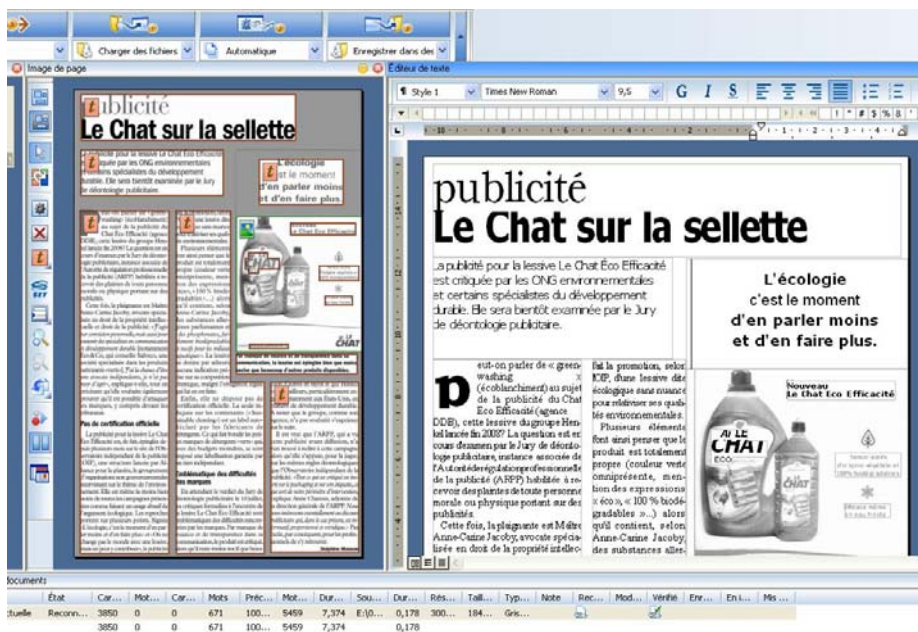
Le contraste et la qualité d'impression des caractères sont donc optimales pour ce document. L'article a été découpé numériquement du reste de la page grâce à une option de recadrage du logiciel.

Plusieurs numérisations ont été réalisées à partir de ce document :

- à 200 dpi en noir et blanc (plus exactement en « niveaux de gris »)
- à 300 dpi en noir et blanc
- à 200 dpi en couleur
- à 300 dpi en couleur

Ceci afin d'essayer de répondre aux deux questions suivantes :

- la résolution de numérisation a-t-elle un impact sur la performance de l'OCR ?
- dans quelle proportion la couleur alourdit-elle le poids du fichier pdf de sortie par rapport au noir et blanc ?



On voit sur la capture d'écran ci-dessus que le logiciel a bien distingué la zone « texte » et la zone « image ».

Avec une numérisation à 200 dpi en niveaux de gris, l'OCR produit 5 erreurs, mais le souci est que l'une d'elles porte sur le titre de l'article, ce qui est assez gênant. Et c'est très étonnant puisque la taille de ces caractères est élevée. Concernant les autres erreurs, elles portent toutes sur le texte inclus dans la zone « image » verte en haut à droite, et non pas sur le corps du texte de l'article. On pourrait donc presque déjà se satisfaire d'un tel résultat. D'autant que ces erreurs sont corrigibles manuellement.

Avec une numérisation à 300 dpi en niveaux de gris, le traitement OCR est en revanche parfait, avec 100 % de réussite. Absolument aucune erreur de reconnaissance.



Sur la capture ci-dessus, on voit un exemple de recherche plein texte avec le mot-clé « phosphonates ».

Les résultats de l'OCR obtenus avec la numérisation couleur sont exactement identiques à ceux obtenus avec le noir et blanc.

Poids des fichiers pdf :

-200 dpi, niveaux de gris : 165 ko (projection pour 80000 documents : 12,6 Go)

-200 dpi, couleur : 267 ko (projection pour 80000 documents : 20,37 Go)

Document 2



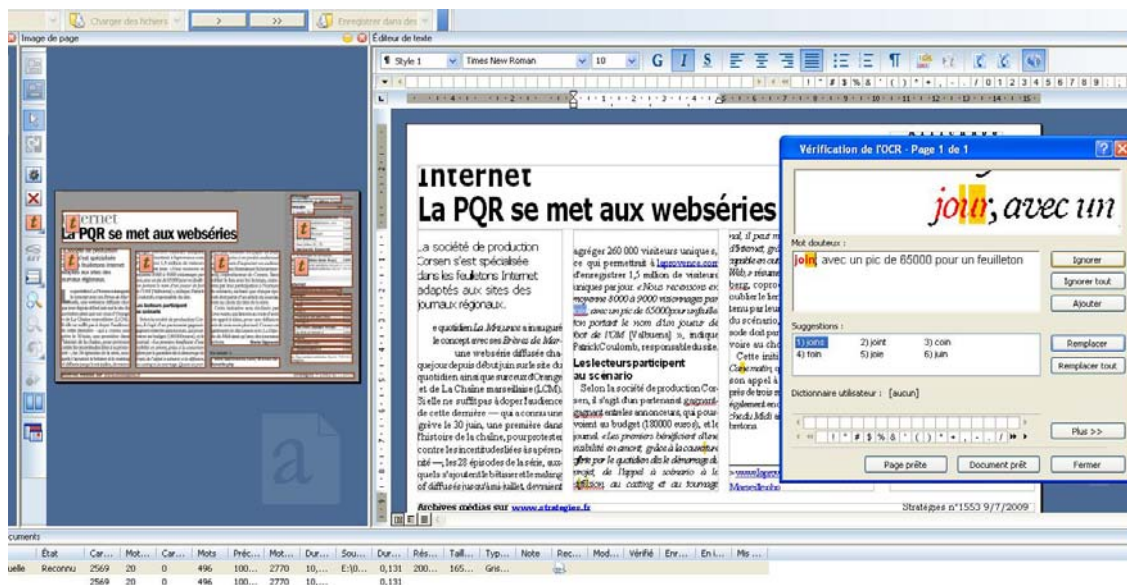
Le deuxième test correspond aussi à une numérisation d'un article de la revue Stratégies, directement à partir de l'original et non pas d'une photocopie.

Le contraste et la qualité d'impression des caractères sont donc optimales pour ce document. L'article a été découpé numériquement du reste de la page grâce à une option de recadrage du logiciel.

Plusieurs numérisations ont été réalisées à partir de ce document :

- à 200 dpi en niveaux de gris
- à 300 dpi en niveaux de gris
- à 200 dpi en couleur
- à 300 dpi en couleur

Avec une numérisation à 200 dpi en niveaux de gris, l'OCR commet seulement 7 fautes de reconnaissance. Néanmoins, tous les mots et les chiffres contenus dans les petits tableaux sur la partie droite du document sont parfaitement reconnus, ce qui est plutôt impressionnant avec une résolution de seulement 200 dpi. De même, les références de la revue – c'est-à-dire le titre du périodique, le numéro, la date de parution – présents dans la partie inférieure droite de la page ont été parfaitement reconnus. Ce qui montre qu'une indexation automatique de ces références – qui nécessiterait tout de même de désigner précisément la zone de la page qui contient toujours les références du périodique – est donc envisageable.



Avec une numérisation à 300 dpi en niveaux de gris, l'OCR est quasi parfait et ne commet qu'une seule faute.



Ci-dessus, recherche plein texte avec le mot-clé « Corsen ».

Les résultats de l'OCR obtenus avec la numérisation couleur sont exactement identiques à ceux obtenus avec le noir et blanc.

Poids des fichiers pdf en sortie :

-200 dpi, niveaux de gris : 192 ko (projection pour 80000 documents : 14,5 Go)

-200 dpi, couleur : 205 ko (projection pour 80000 documents : 15,7 Go)

Document 3

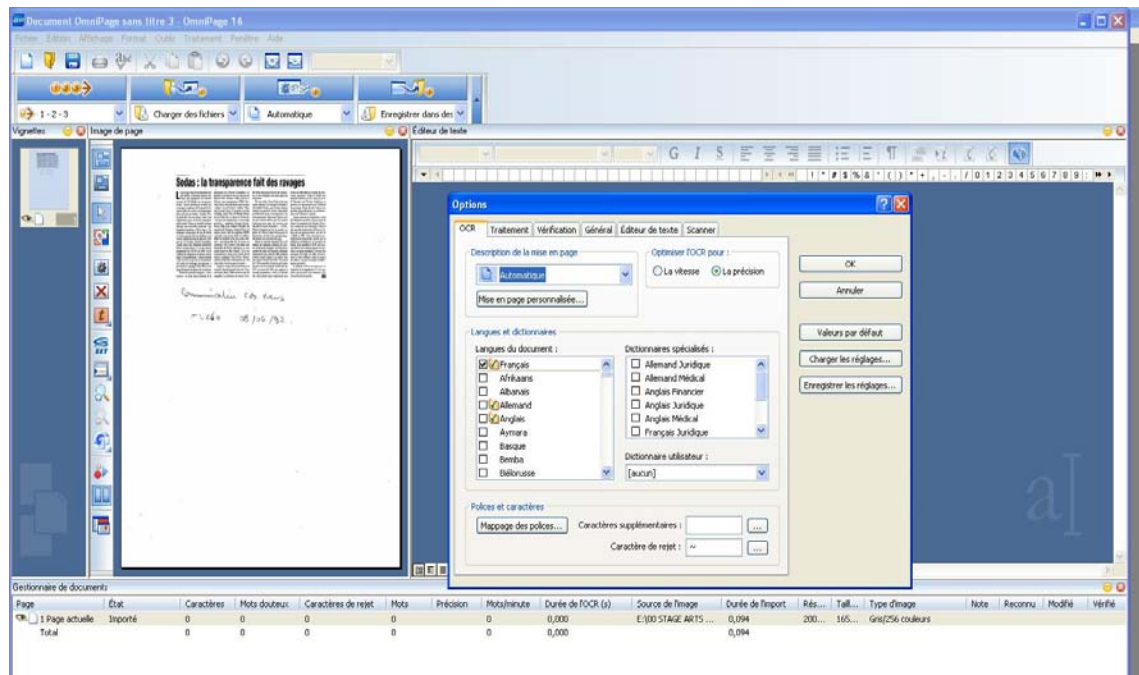
C'est une photocopie noir et blanc d'un article de CB news.

Ce document a été sélectionné à cause de la très petite taille de ses caractères. Cette photocopie est en effet une réduction – avec un taux de réduction visiblement trop élevé - de l'original.

En revanche la qualité de conservation du papier, ainsi que la qualité d'impression, et le contraste, sont très bonnes pour cette photocopie.
Trois tests de numérisation – tous en niveaux de gris -, à 200 dpi, 300 dpi, 400 dpi, ont été effectués avec ce document.

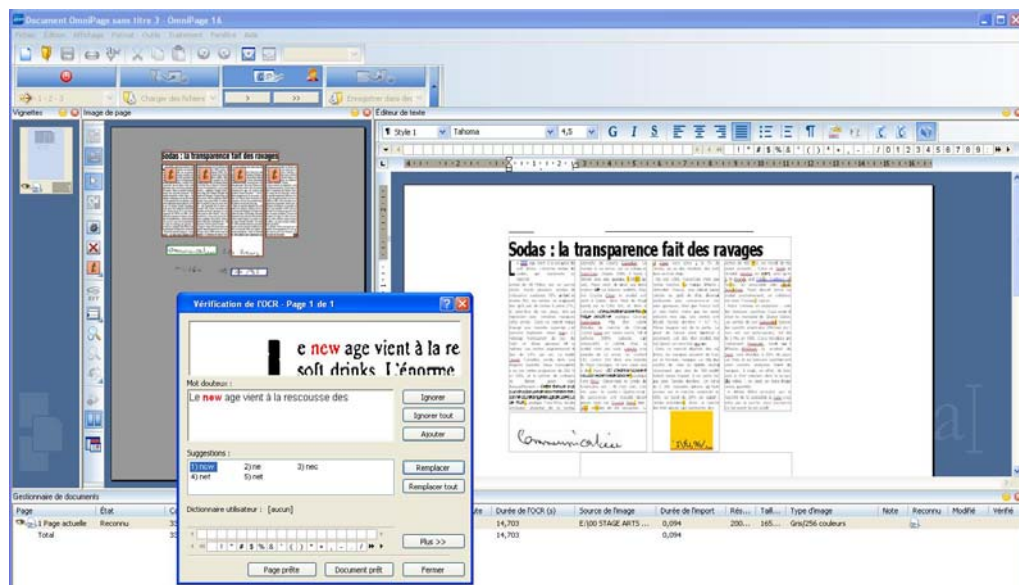
Capture écran 1- Options de l'OCR :

Sur la gauche se trouve le document original (très petits caractères).
Choix des langues et de la priorité entre vitesse ou précision.

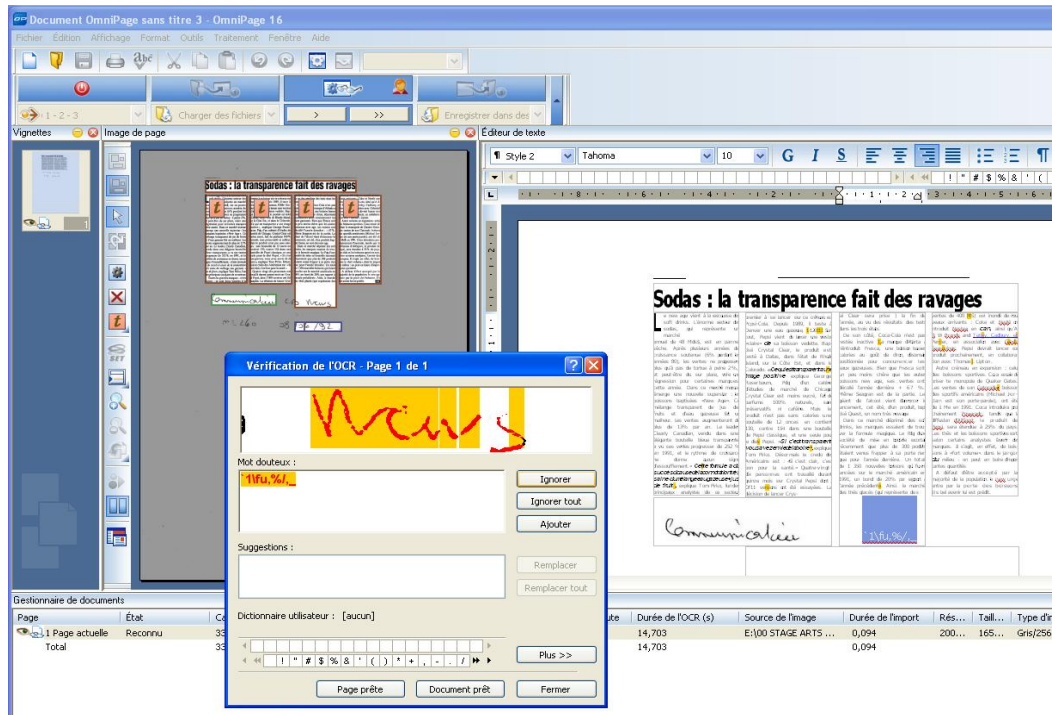


Capture écran 2 et 3 - Traitement OCR :

Sur la gauche on remarque les zones qu'a délimité le logiciel.
On a activé la vérification de l'OCR, qui détecte les mots douteux.

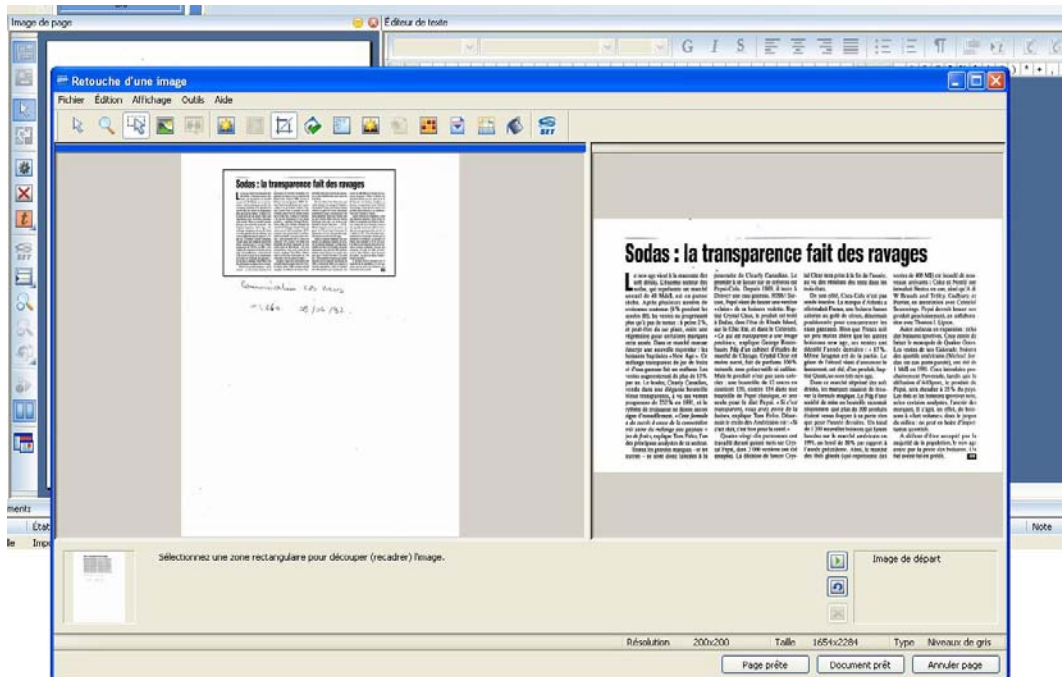


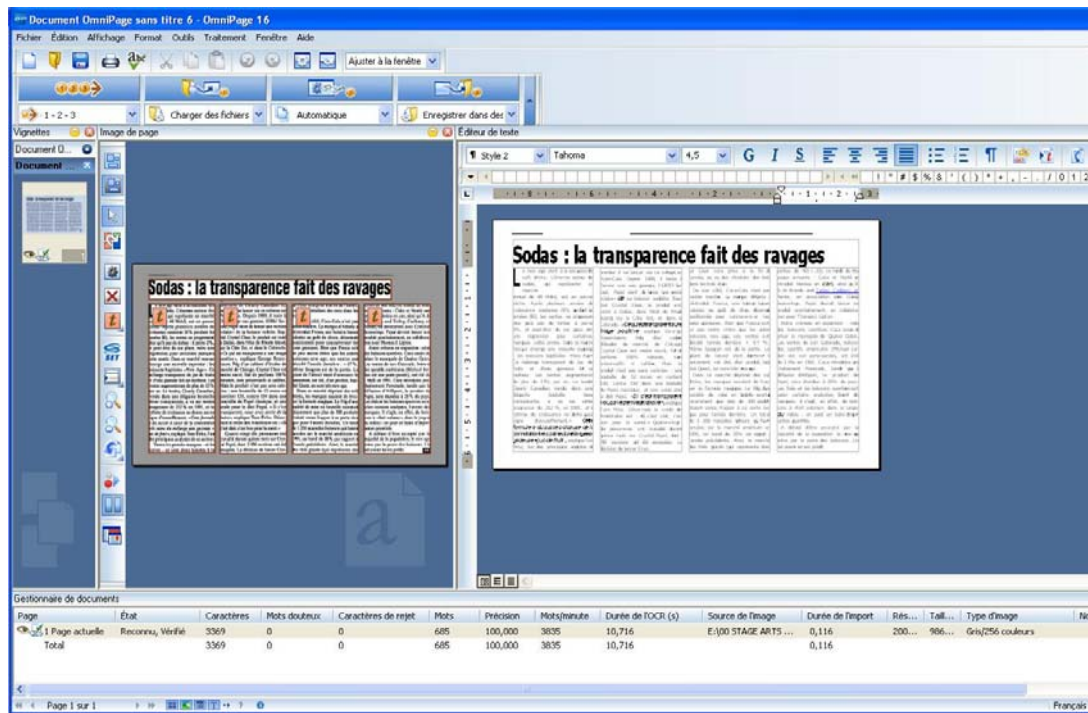
Ici, logiquement, le logiciel n'arrive pas à reconnaître l'annotation manuscrite « CB News ». Comme il a été expliqué plus haut dans cette étude, les lettres non séparées et non majuscules n'ont aucune chance d'être reconnues, peu importe la résolution de numérisation.



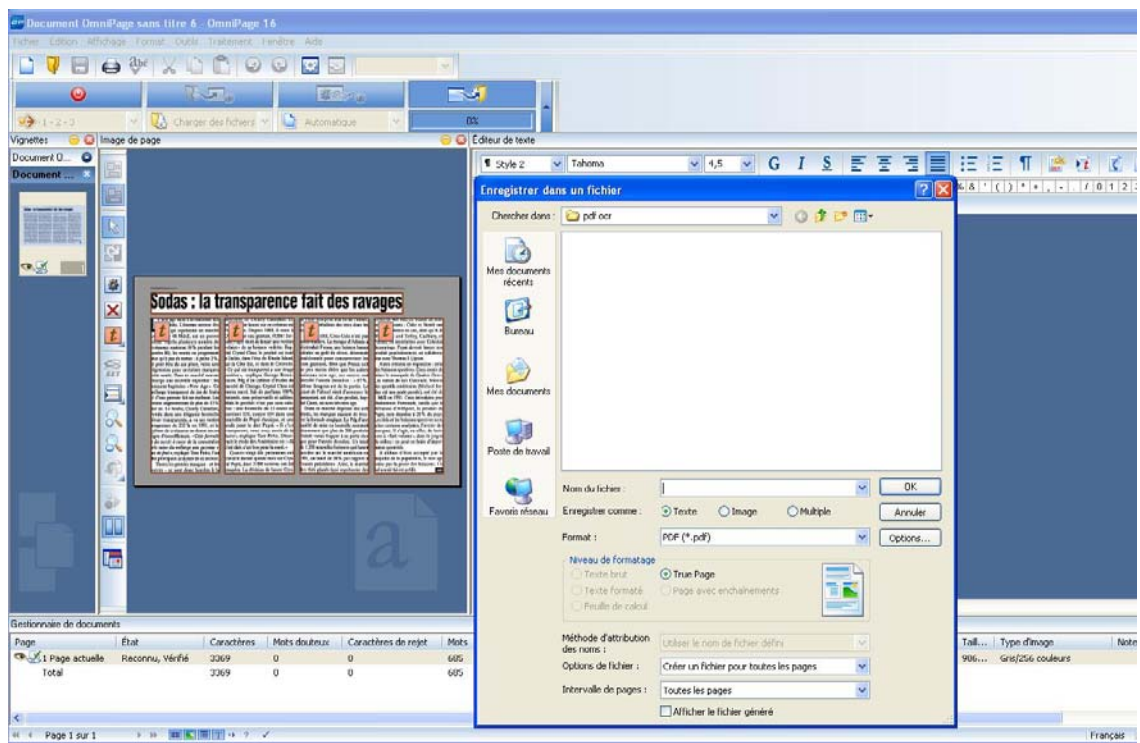
Capture écran 4 et 5 – Découpage et recadrage :

Dans ce logiciel, plusieurs options de traitement de l'image sont disponibles, dont la découpe et le recadrage. Ici cette option nous sert à enlever la partie inutile qui contient les références manuscrites (qui seront re-saisies via l'indexation), et surtout à supprimer toute la partie vierge de la page et qui alourdirait inutilement le fichier pdf final de sortie.



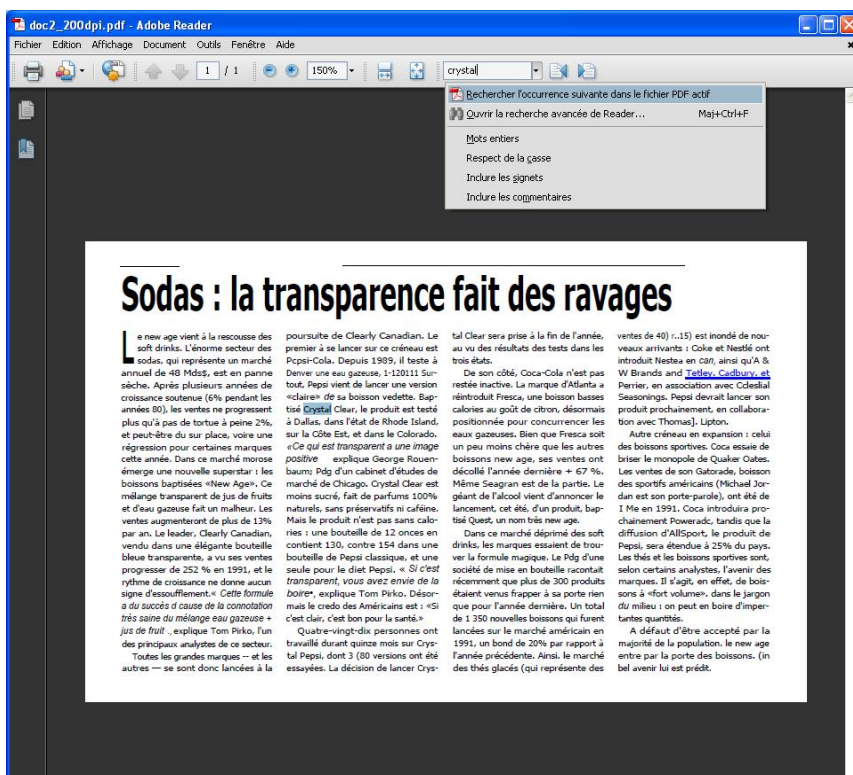


Capture écran 6 – Enregistrement du fichier final (toujours au format pdf)



Capture écran 7 – Ouverture du fichier avec Adobe Reader et tests de recherche plein texte

Les tests ont montré que, quelle que soit la résolution – 200, 300, 400dpi –, le traitement OCR a parfaitement fonctionné, puisque 100% des mots typographiés ont été reconnus. Même à 200 dpi et avec ces très petits caractères, la performance de l'OCR est excellente. On a fait ici un test avec le mot-clé « crystal ».



Sur le plan du poids des fichiers, voici les chiffres :

- à 200 dpi le poids du fichier est de 113 ko ;
- à 300 dpi il est de 121 ko ;
- à 400 dpi il est de 130 ko ;

Soit une différence de 15% entre la résolution la plus faible et la plus élevée – sur un document, précisons le, qui ne comporte pas d'image.

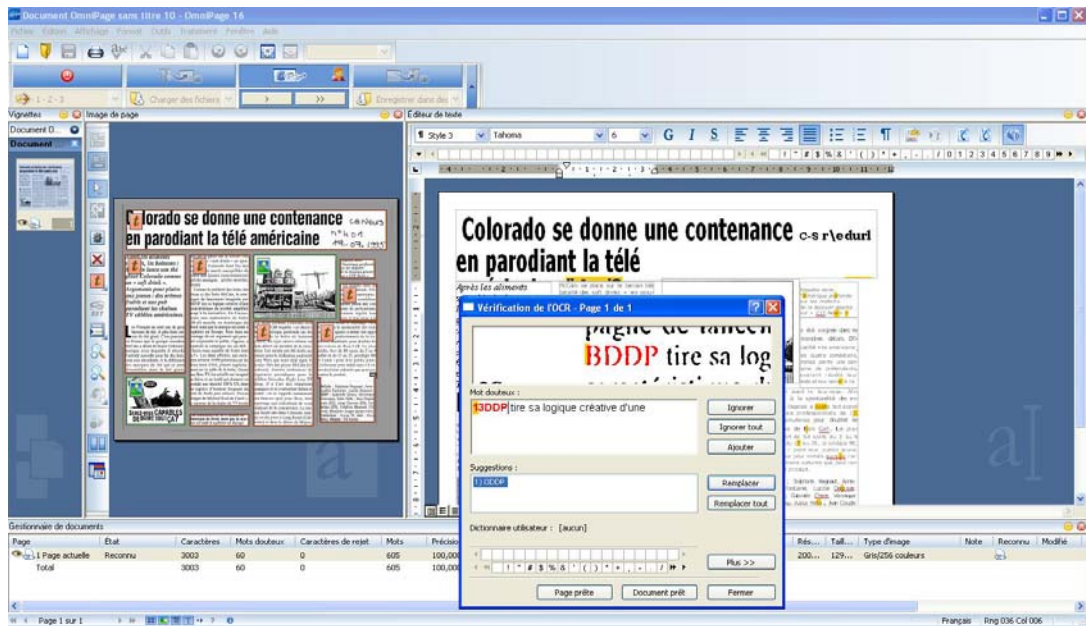
Estimations si on se projette avec 80000 documents (400 pages en moyenne pour 200 boîtes) :

- $113 \times 80000 / 1024 = 8828 \text{ Mo} = 8,6 \text{ Go}$
- $121 \times 80000 / 1024 = 9453 \text{ Mo} = 9,23 \text{ Go}$
- $130 \times 80000 / 1024 = 10156 \text{ Mo} = 9,91 \text{ Go}$

Document 4

C'est une photocopie en noir et blanc d'un article de CB News, lui aussi réduit. Mais cette fois l'article comporte des images. Les tests pour ce document se sont révélés très instructifs. En effet, bien que les caractères sur cette photocopie soient plus gros que ceux du document précédent, on a pu cette fois constater que la performance de l'OCR était liée à la résolution de numérisation du document. Ainsi, certaines erreurs de reconnaissance se produisant avec une résolution de 200 dpi ne se produisent pas à 400 dpi. Et c'est seulement à 400 dpi que l'OCR a obtenu 100% de reconnaissance pour ce document. L'explication tient à la qualité d'impression de cette photocopie, qui est légèrement moins bonne que pour le document 3.

Voici ici une capture d'écran montrant une erreur de la part du logiciel à la résolution de 200 dpi. L'OCR a interprété un 13 à la place d'un B. Erreur qui ne s'est pas reproduite à 400 dpi.



On voit bien sur la gauche de cette capture que le logiciel a bien distingué les zones de « texte » (en orange) des zones « d'image » (en vert).

A 400 dpi le résultat de l'OCR est parfait. Ici, test avec le mot « Cali ».



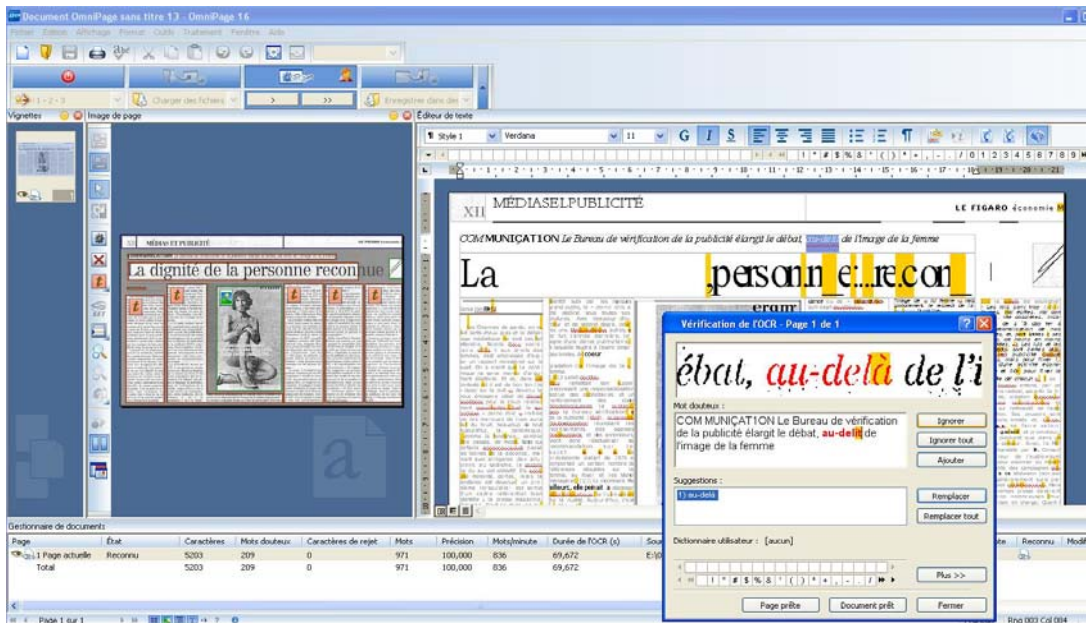
Pour ce qui est du poids : 170 ko pour le 200 dpi, 276 ko pour le 300 dpi, 362 ko pour le 400 dpi.
 Projections avec 80000 documents : 13 Go à 200 dpi, 21 Go à 300 dpi, 28 Go à 400 dpi.

Document 5



C'est une photocopie en noir et blanc d'un article du Figaro économie. Elle a été sélectionnée car elle présente un très mauvais contraste ainsi qu'une médiocre qualité d'impression. Trois tests ont été menés, à 200, 300 et 400 dpi. Précisons que l'article a été recadré au préalable.

L'état de cette photocopie (impression médiocre, mauvais contraste, textes volumineux et denses) a allongé considérablement le temps du traitement OCR par le logiciel, qui était ici d'environ 1mn 30 s.



On voit que le logiciel a bien distingué le texte et la photo, mais qu'il y a un souci avec la zone du titre.

De plus, aux trois résolutions se produisent de très nombreuses erreurs de reconnaissance, comme pour l'exemple ci-dessus. 120 mots douteux à 200 dpi, plus de 300 à 400 dpi. Cela provient de la mauvaise impression (caractères mal imprimés et pas assez « marqués »), plus que du mauvais contraste avec le fond de page.

On constate donc avec cette exemple qu'élever la résolution de numérisation ne solutionne pas toujours les problèmes de reconnaissance, et que la qualité de la photocopie reste primordiale. Ici il serait donc préférable de numériser à partir de l'original afin d'obtenir un taux de reconnaissance proche de 100%.

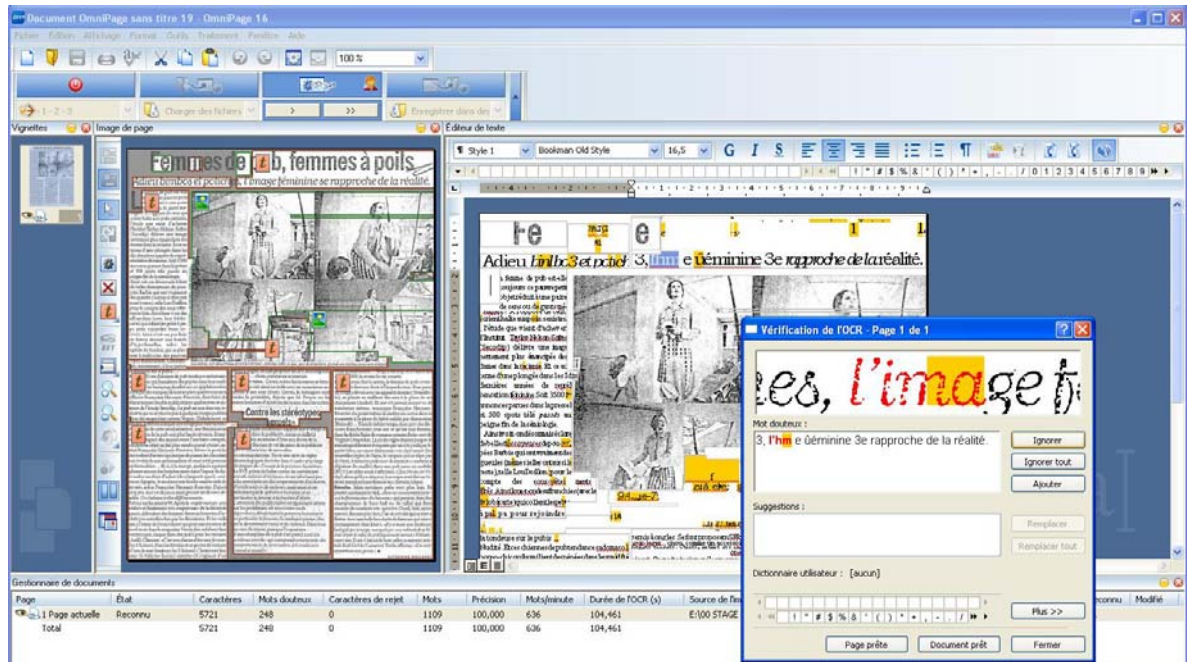


Les poids des fichiers de sortie sont ici très proches de ceux concernant le document 4.

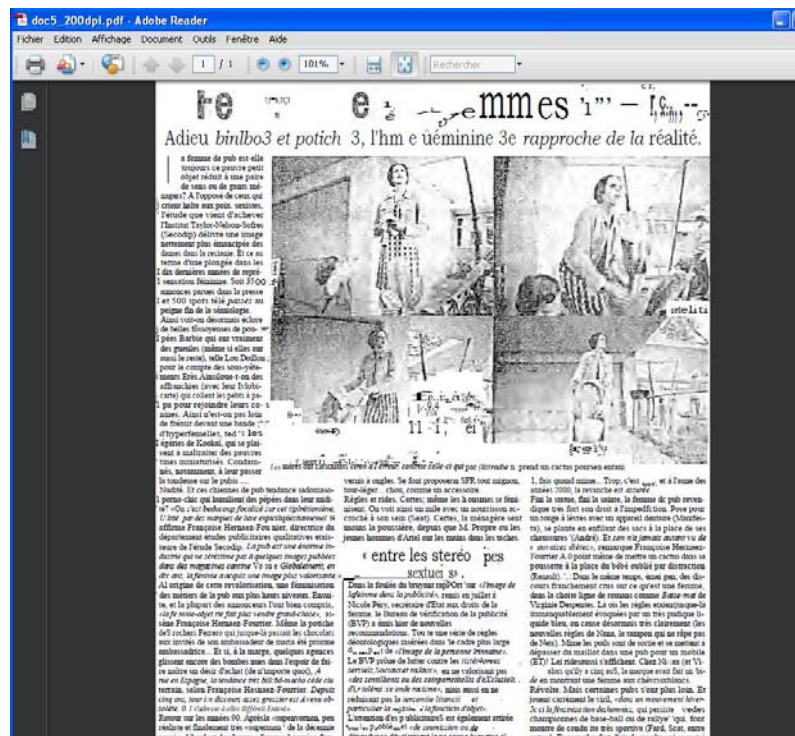
Document 6



C'est une photocopie en noir et blanc d'un article de Libération. Elle présente des caractéristiques très proches de celles du document précédent : mauvais contraste et mauvaise qualité d'impression – légèrement pire que celle du document 5 précédent. Trois tests ont été menés, à 200, 300 et 400 dpi. Précisons que l'article a été recadré au préalable.



La durée du traitement OCR a été ici d'environ 2mn. Comme pour le document 5, beaucoup d'erreurs de reconnaissance avec les trois résolutions (plus de 250 mots douteux à chaque fois).



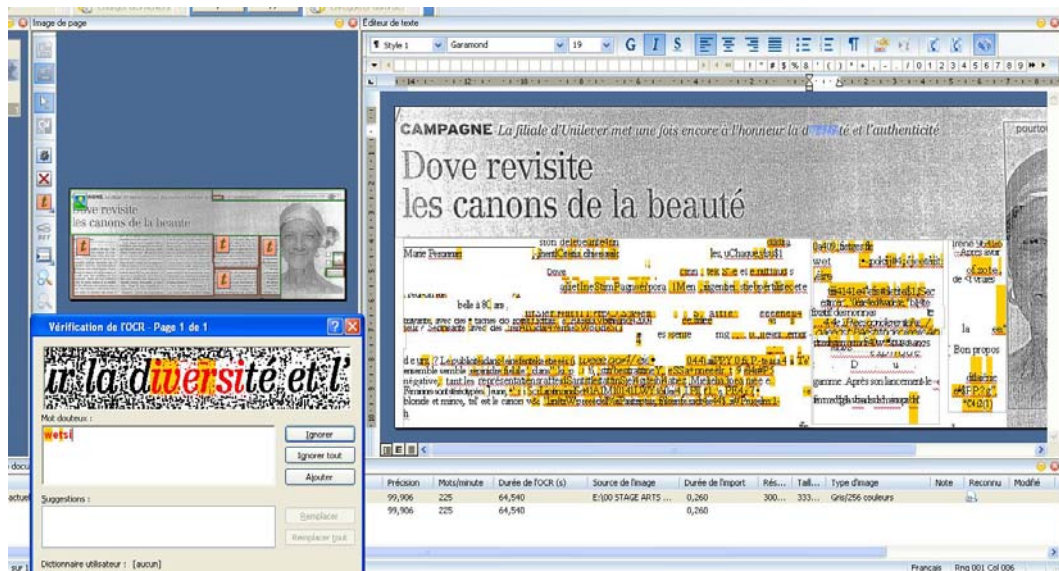
Le résultat final est donc très décevant (cf le titre en haut qui n'a pas été du tout reconnu malgré la taille élevée des caractères). Il est donc là aussi préconisé de renumériser l'article à partir de la revue originale.

Le poids du fichier de sortie est d'environ 250 ko. Soit 20 Go pour une projection à 80000 documents.

Document 7



C'est une photocopie en noir et blanc d'un article, dont la source est inconnue ! Elle présente un contraste particulièrement dégradé à cause d'un fond de page très « bruité ». Vu les conclusions des tests précédents – à savoir que la résolution n'a pas d'impact sur l'OCR si la qualité de la photocopie est trop mauvaise -, un seul test a été mené à 300 dpi. Précisons que l'article a été recadré au préalable.





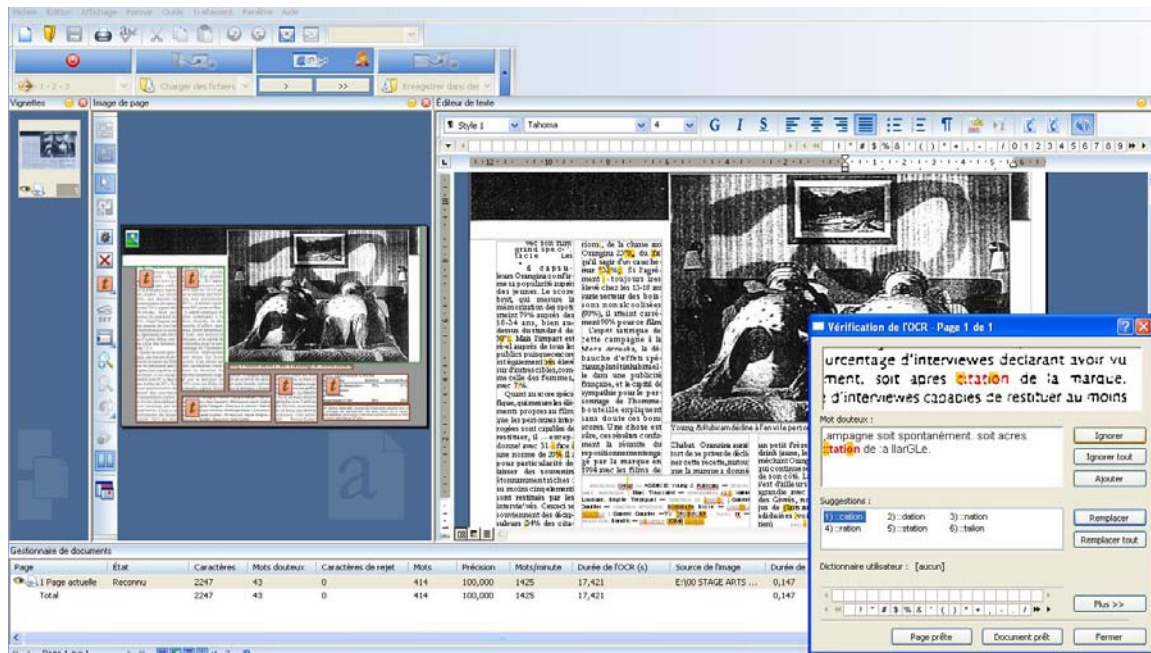
On constate – cf les captures d’écran - que le test OCR a ici complètement échoué. Le taux de reconnaissance est très faible et la recherche plein texte est ici inexploitable. La qualité de cette photocopie - notamment à cause du bruit de fond - est en dessous du seuil acceptable par le logiciel. Il y a donc obligation pour cet article de retourner à l’original (si le musée le possède) pour procéder à une nouvelle numérisation.

Document 8

C’est une photocopie noir et blanc d’un article de la revue *Stratégies*.

Le contraste est plutôt correct, mais la qualité d’impression des caractères est en revanche assez mauvaise.

La résolution de numérisation pour ce document était de 300 dpi.



Le taux de reconnaissance de l’OCR est ici - de façon attendue - très médiocre, avec environ une trentaine d’erreurs. Cf la capture ci-dessus.

En revanche les 300 dpi assurent une numérisation parfaite de l'image, sans aucune perte.



On voit ci-dessus par exemple que le premier caractère du texte – pourtant très gros -, en haut à gauche, n'a pas été reconnu par le logiciel.

Poids du fichier pdf de sortie : 184 ko.

Document 9

C'est une photocopie noir et blanc d'un article de la revue Médias.

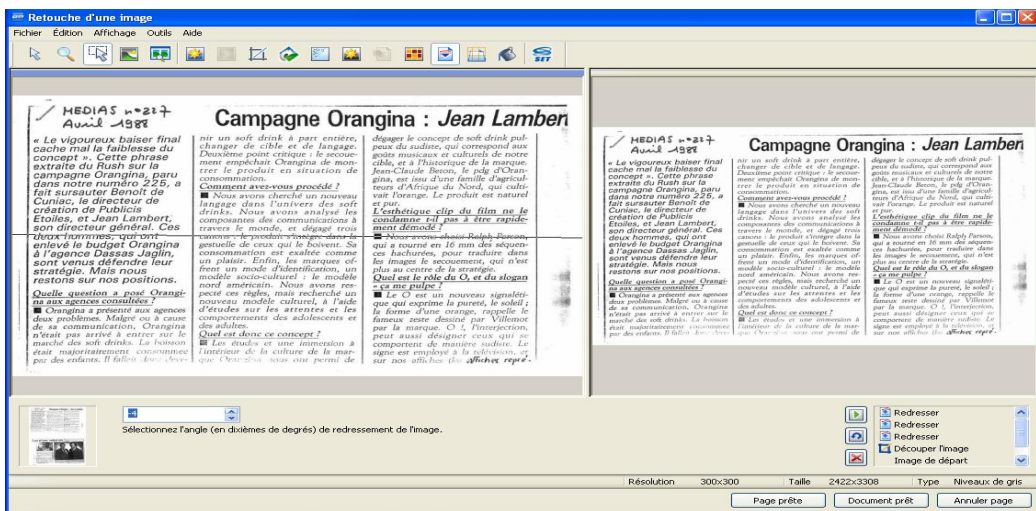
Le souci avec ce document est qu'il se compose de deux parties – suite à la photocopie de deux morceaux de papiers ayant été découpés aux ciseaux. La partie supérieure est inclinée alors que celle inférieure est bien alignée. Il faudrait donc isoler ces deux parties avec un découpage numérique, aligner la partie supérieure, puis refusionner ces deux parties en un fichier unique. On voit donc que cela peut être long et complexe, et qu'une photocopie faite sans un minimum de soin peut devenir problématique lors de la dématérialisation.

Une autre spécificité de ce document est que la qualité de la reprographie est inégale selon la région de l'article – probablement du encore une fois à une mauvaise manipulation lors de la photocopie. Ce qui ne facilite pas les traitements manuels potentiels comme la correction du grain ou l'amélioration du contraste.

Ce document était numérisé à 300 dpi.

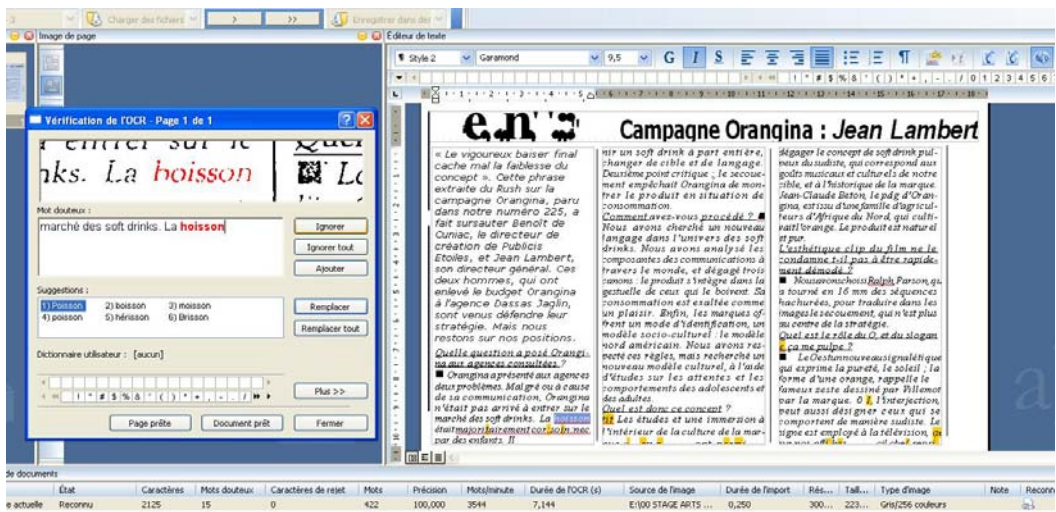


Outil de « redressement » disponible dans le logiciel ScanSoft OmniPage :



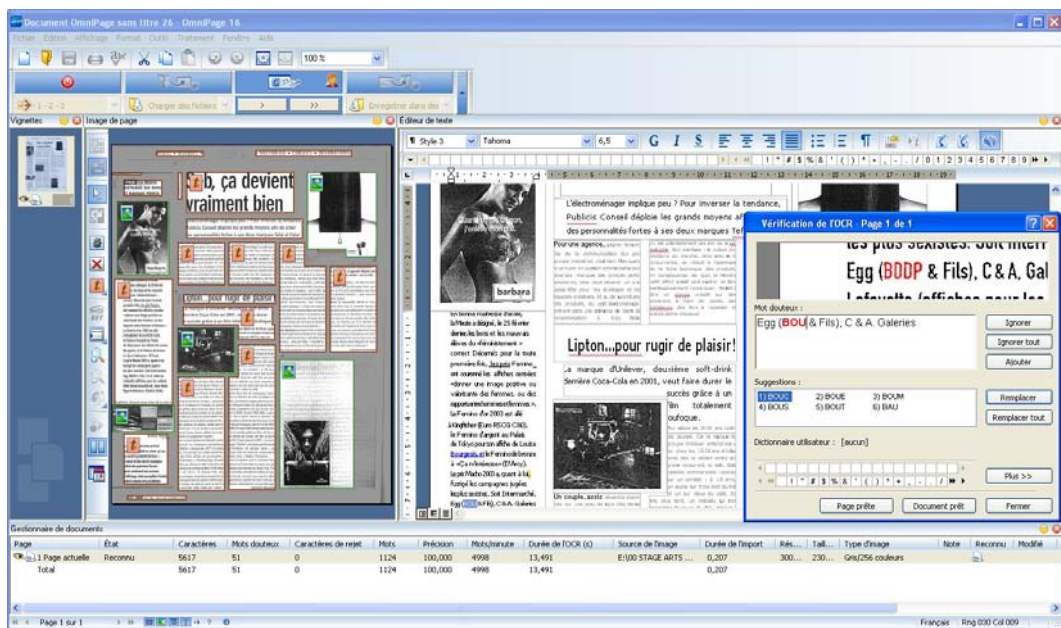
Avec cet outil, il est possible de redresser un document au dixième de degré près.

Malgré un très mauvais contraste dans la partie inférieure droite de l'article, l'OCR a bien fonctionné avec moins d'une dizaine d'erreurs. Cf la capture ci-dessous.





C'est une photocopie noir et blanc, avec réduction, d'une page de la revue CB news. La particularité de ce document est qu'il contient plusieurs articles sélectionnés sur cette même page, et que l'un d'eux – dans la partie inférieure droite – est particulièrement mal reproduit, à tel point qu'il est complètement illisible. Excepté cette partie mal reproduite – dont je ne m'explique pas l'origine ? –, le reste de la photocopie bénéficie d'un très bon contraste et d'une très bonne qualité d'impression. La résolution de ce document numérisé est de 300 dpi.



Ce test de numérisation est particulièrement intéressant puisqu'on voit sur la capture écran ci-dessus que le logiciel a interprété la partie illisible de la photocopie comme une zone « d'image » et non comme une zone de « texte ». Voir, tout en bas à droite, le cadre de couleur verte.

Pour toutes les autres zones de la photocopie, les résultats de l'OCR sont en revanche excellents (proche du 100%), avec moins de 10 erreurs recensées – dont l'une est représentée dans la capture ci-dessus, avec un BOU reconnu à la place de BDDP.

On peut donc se poser la question suivante pour ce document : doit-on faire le sacrifice de la partie non reconnue par l'OCR, ou doit-on recommencer la numérisation à partir de la revue originale ?

Poids du fichier pdf de sortie : 201 ko



Recherche plein texte totalement satisfaisante pour ce document, ici avec le mot-clé « christiansen ».

Document 11

C'est un document très semblable au document précédent, c'est-à-dire une photocopie noir et blanc d'une page de la revue CB news. Mais cette fois, absolument tous les articles contenus dans cette page ont été sélectionnés par le documentaliste.

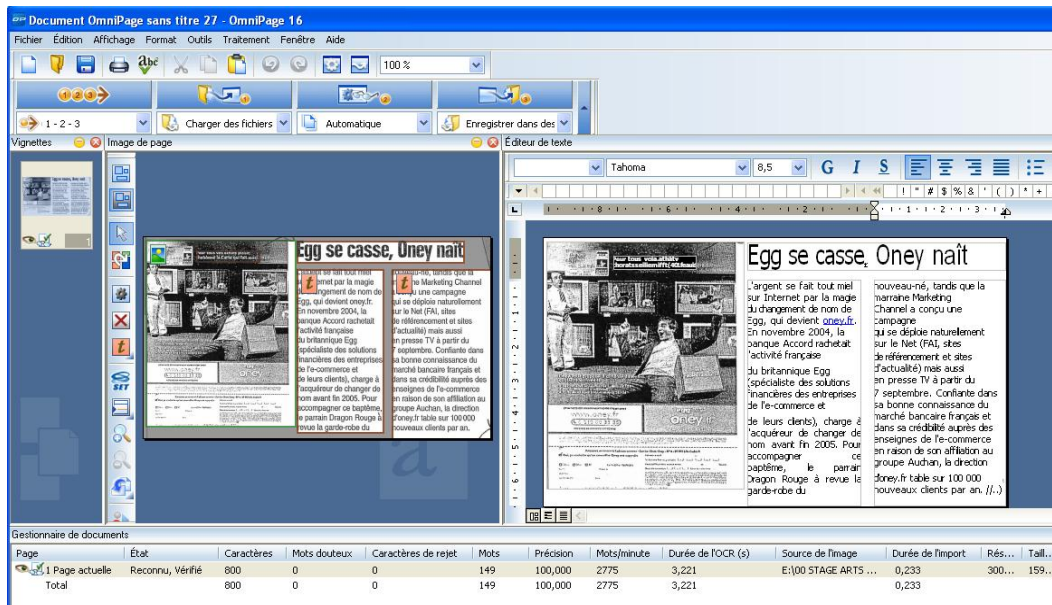


La question suivante se pose donc – question importante abordée plus haut dans un chapitre de l'étude - : doit-on conserver ce document sous la forme d'une page entière numérisée et contenant tous les articles, ou doit-on isoler chacun des articles sous la forme d'un fichier pdf indépendant ?

La première option a l'avantage d'être plus simple et rapide lors du traitement. La deuxième a l'avantage d'être plus confortable et esthétique pour le lecteur. Pour l'exemple, nous avons ici fait le choix d'isoler les articles.

On remarque là aussi que la qualité de reprographie est inégale selon les zones de la photocopie – la partie inférieure gauche présente un contraste fortement dégradé, on peut donc s'attendre à un OCR peu fiable.

La résolution de numérisation pour ce document est de 300 dpi.

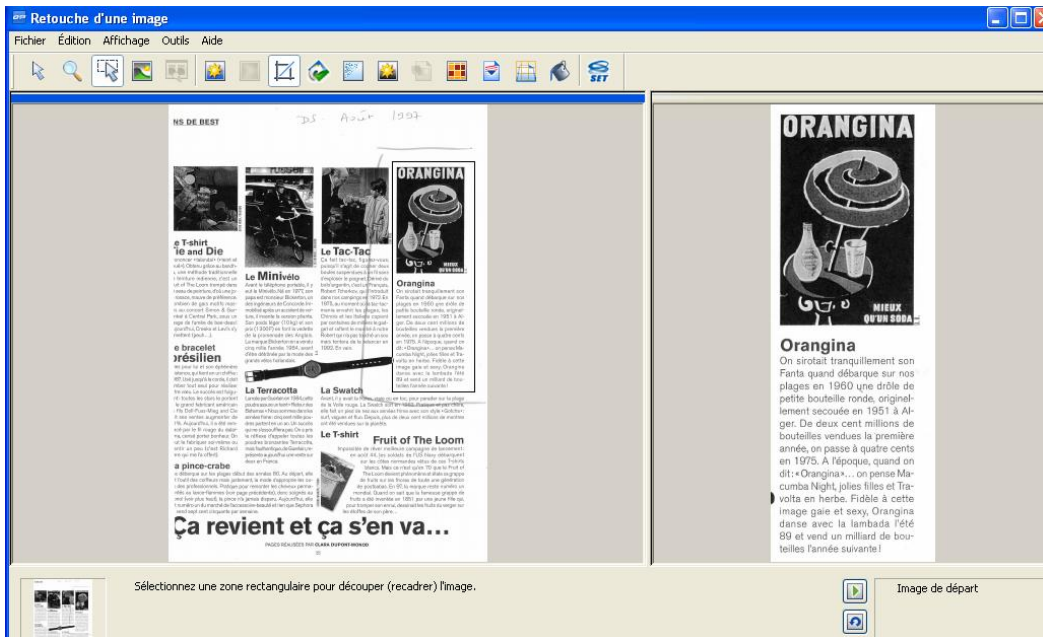


Avec cet article par exemple, l'OCR donne un résultat parfait (100% de reconnaissance).

Document 12



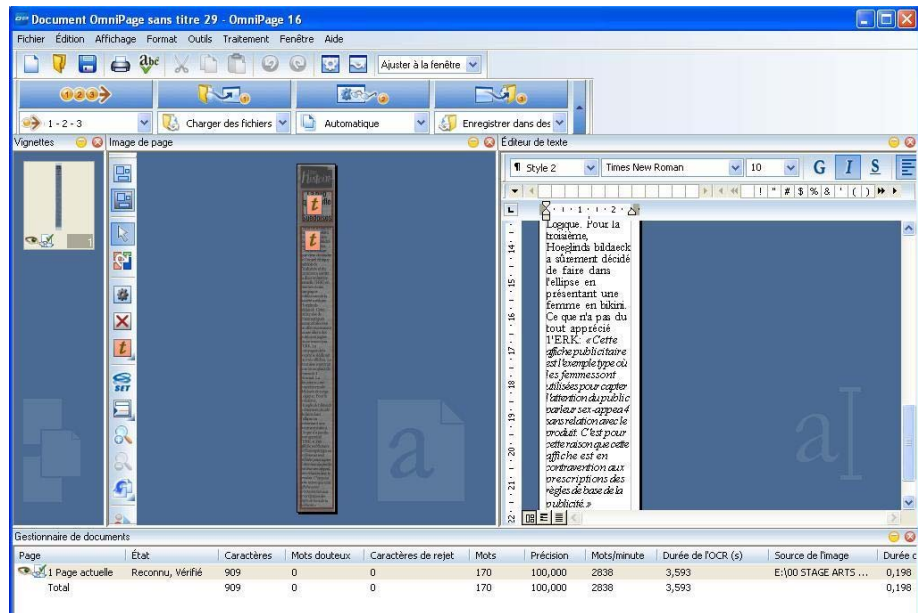
C'est une photocopie noir et blanc de la page d'une revue. Elle contient plusieurs articles dont un seul a été sélectionné par le documentaliste, et il semble donc naturel de le recadrer – même si cela allonge légèrement le temps de traitement, d'environ 20s avec OmniPage. Le contraste de cette photocopie est idéal, et la qualité d'impression est parfaite. La résolution de ce document est de 300 dpi.



Le taux de reconnaissance OCR est de 100% pour ce document, ce qui était attendu vu la qualité de cette photocopie.

Document 13

C'est une photocopie noir et blanc d'un article ayant été découpé dans une revue. Vu la très petite surface qu'occupe l'article sur cette photocopie, il convient bien naturellement de le recadrer, afin de réduire au minimum le poids du fichier de sortie – on rappelle que même une zone vierge blanche est mémorisée sous forme d'octets. Le contraste pour ce document est très médiocre (noir sur gris), mais en revanche il n'y a aucun « bruit » et le fond d'article est très homogène – ce qui est toujours un gros avantage pour la reconnaissance. La résolution de ce document est de 300 dpi.

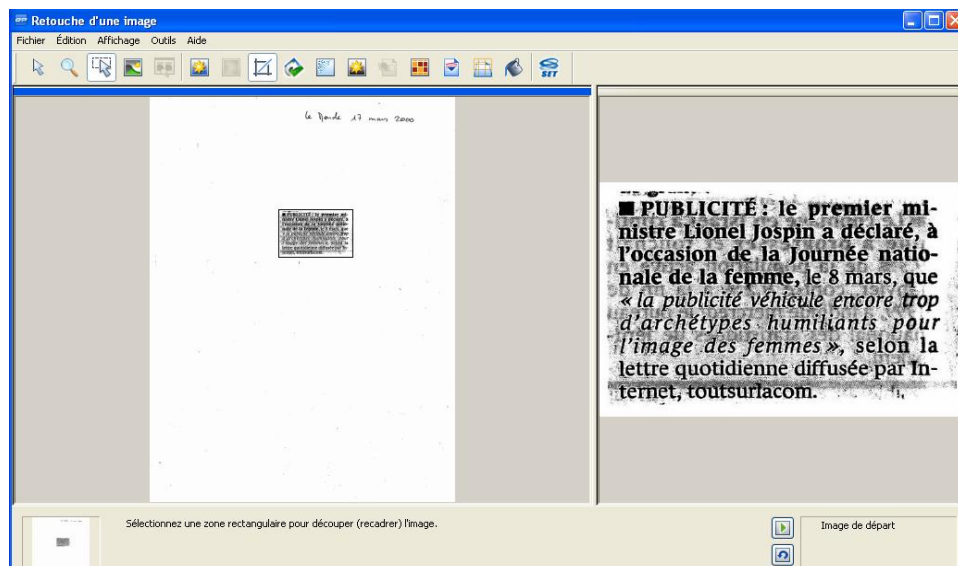


Le taux de reconnaissance OCR est de 100% pour ce document, alors que le contraste de la photocopie pouvait laisser craindre un résultat beaucoup plus mauvais.

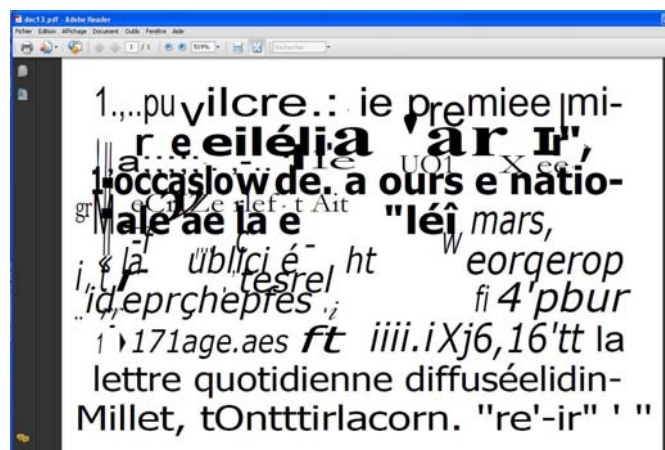
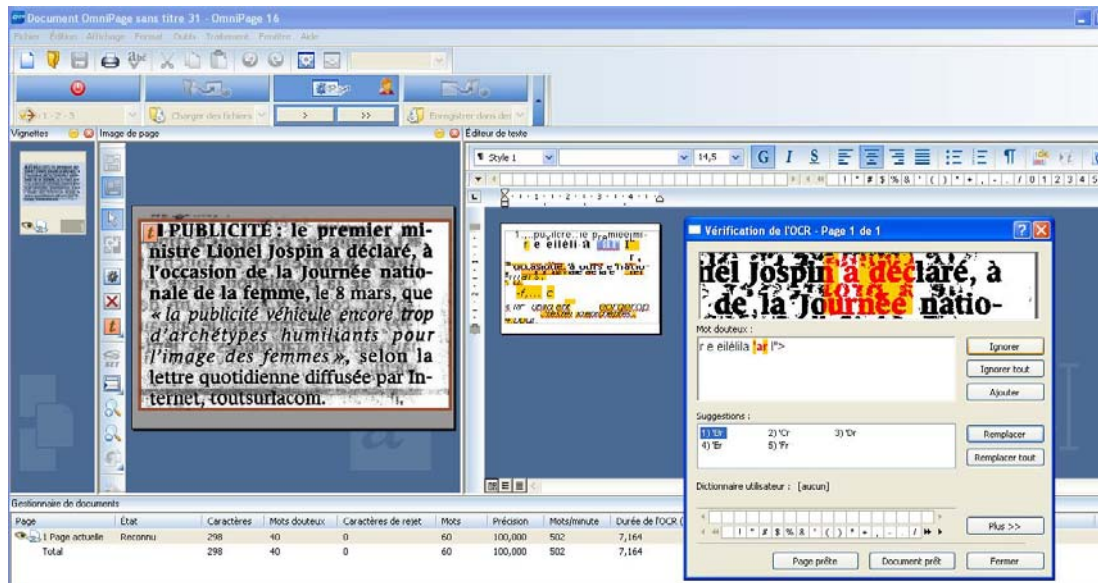
Le poids du fichier pdf de sortie est très faible, 60 ko.

Document 14

C'est un document assez semblable au précédent. Il a été redressé au préalable. Un recadrage s'impose également. La résolution de ce document est de 300 dpi.



Le contraste est faible, tout comme pour le document précédent, mais il est cette fois hétérogène, et présente beaucoup de « bruit ». En conséquence, l'OCR est un échec total, avec un taux de reconnaissance proche de 0%.



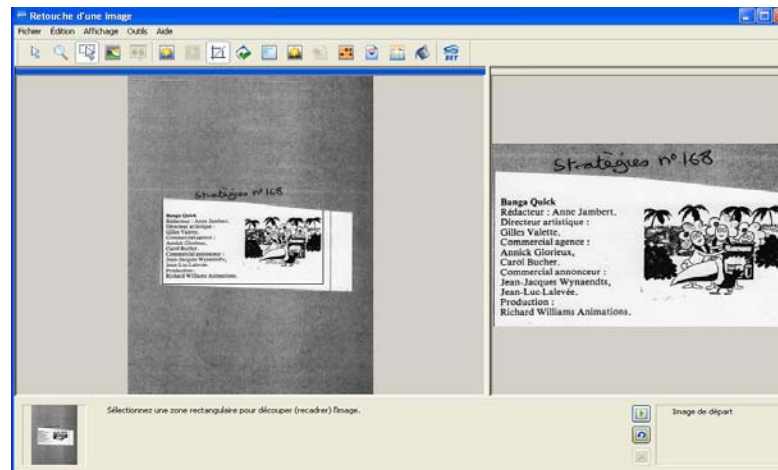
Document 15

Ce document ne présente aucune difficulté particulière – bon contraste, bonne qualité d'impression.

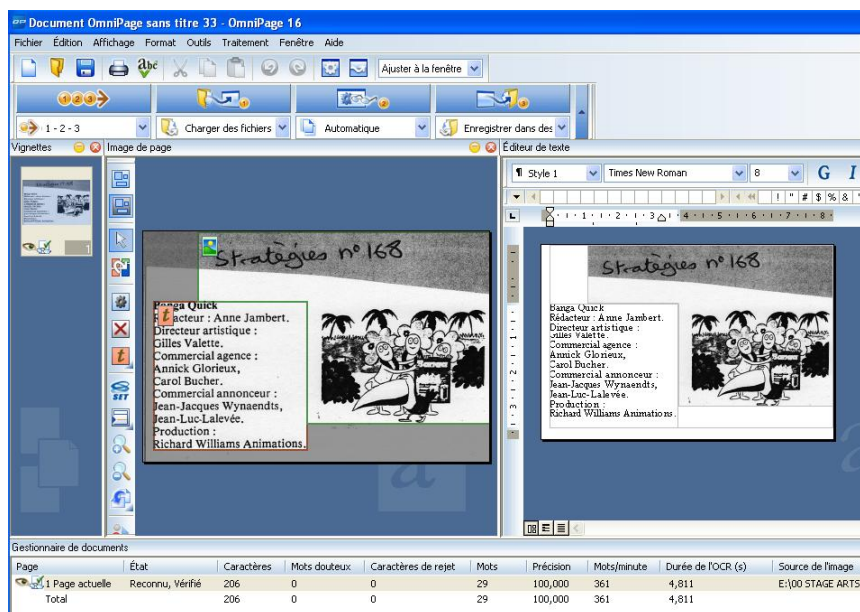
Il a seulement été sélectionné pour démontrer encore une fois que la **reconnaissance des caractères manuscrits non séparés et non majuscules est inefficace**.

La résolution de ce document est de 300 dpi.

Un recadrage de l'article s'impose aussi ici.

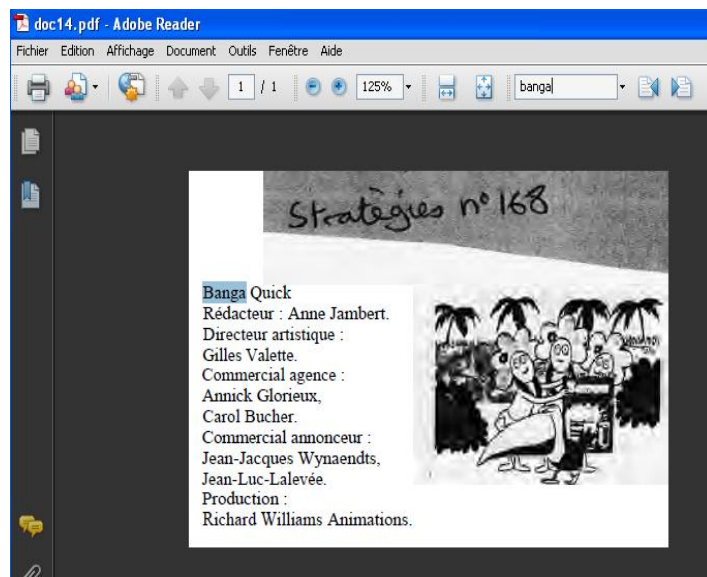


Ici encore, l'écriture manuscrite a été confondue avec une zone « image ».



L'OCR a donné ici 100% de réussite pour la zone de texte typographié.

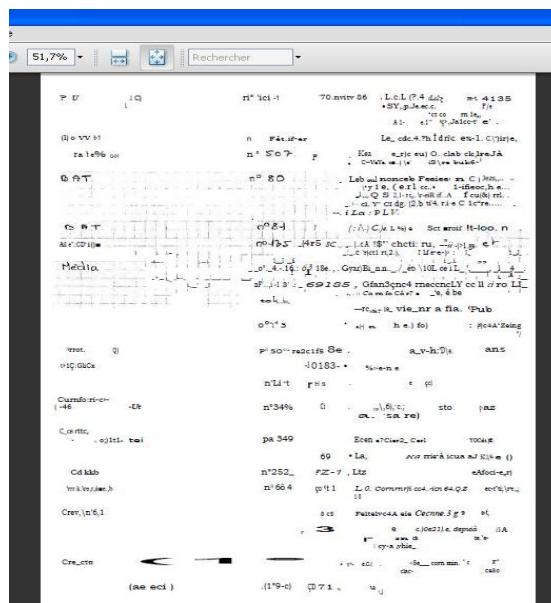
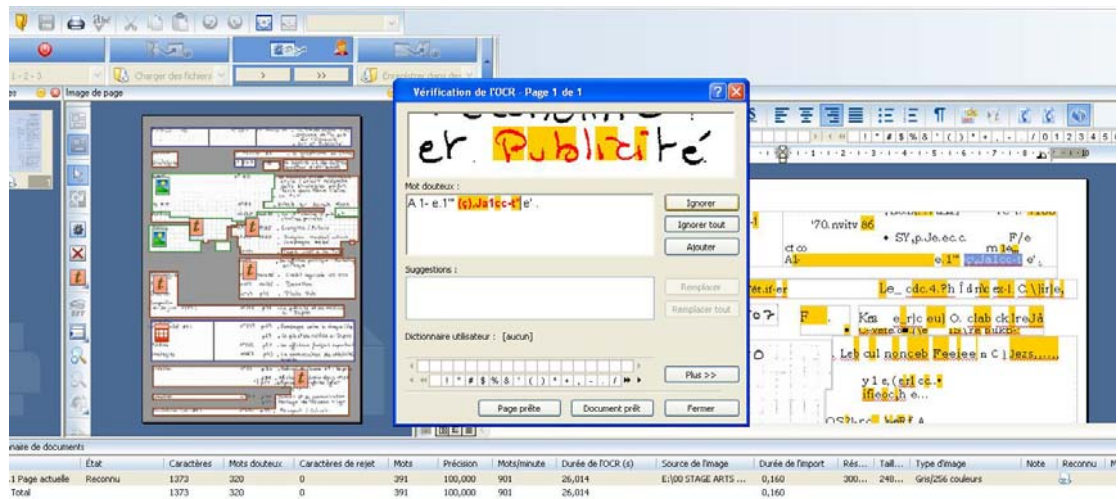
Ci-dessous, une recherche plein texte avec le mot-clé « Banga ».



Document 16

PUBLIC - IQ	n°204	24 Janvier 86	La Publicité chiffres 1985 - conjonction de la pub sur l'économie - Art of Publicité
NOUVA	n°	Fevrier 86	le graphisme en Chine
Stratégie	n° 507	p46	les agences US de chiffres - Carrefour distribution
OAT	n° 80		les annonces presse nouveaux style / en noir / beige dans stratégies médium Paris dans Marie Claire - La P&V
OAT	n°84		Article sur Sarah Moon
Media	n°135	mars 86	La 5 ^e Chine // pub et chiffres privés
Media	n°146	09/85	Grangina / Ballaro
	n°145	09/85	Grangina meunier colaro - Campagne bébé
	n°144		Tagline venir à la Pub
	n°143		les affiches politique : Marketing politique
Stratégies	n°209	mars 86	Crédit agricole 10 ans
Media	n°119	10/85	Benetton
Creation	n°49	p15	Photo Pub
Compagnie (de 20 Juin 1985)	n°213	p16	La publicité et ses médias au Japon
Compagnie (14-30 Juin 1985)	n°219	p19	Campagne contre le simple (ep) p69
	n°218	p71	les affiches français export
Media	n°669	p43	la communication des collectivités locales
Stratégies	n°56	p30	Festival de Cannes 87 : le prix d'Or - Affichage : depuis 1984 - Affichage : depuis 1984 - Affichage : depuis 1984
Creation	n°54	p36	Journal et communication p47
Stratégies (de 1985)	n°470	p22	Peugeot / Culver

Ce document ne comporte que des caractères manuscrits, et le taux de reconnaissance OCR est quasi nul.

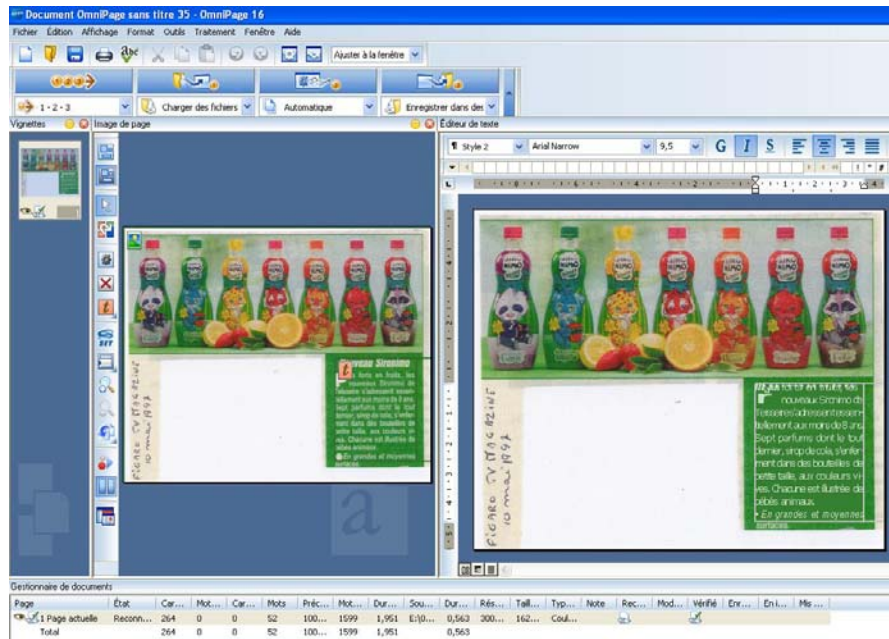


Document 17

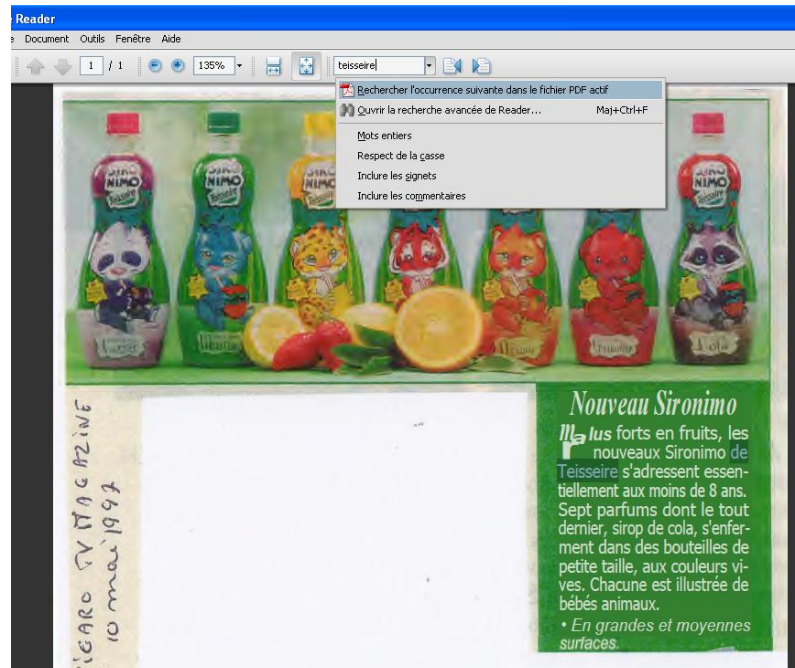
C'est une photocopie couleur d'un article découpé dans le Figaro Magazine. Le choix de numériser en couleur s'explique par un souci de bien conserver les informations et les messages véhiculés par l'image illustrant l'article - des bouteilles de sirop. C'est typiquement l'exemple d'un article où la photo est autant informative, sinon plus, que le texte, et où le choix de la couleur se justifie pleinement.

Pour cette photocopie le contraste est bon et la qualité d'impression des caractères (netteté) est bonne.

L'article a été recadré. La résolution de numérisation est de 300 dpi.



De façon attendue, le traitement OCR affiche un taux de reconnaissance de 100%.

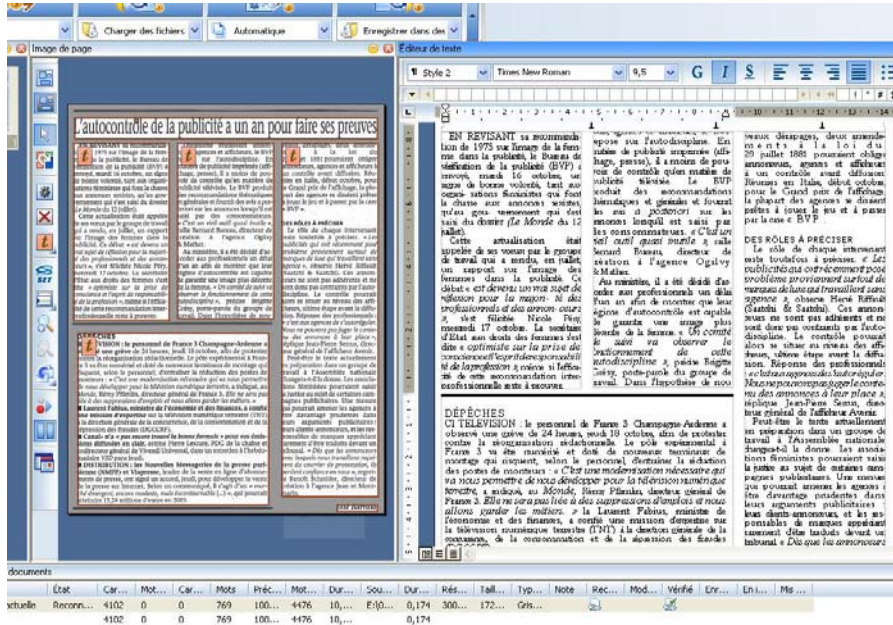


Recherche plein texte avec le mot-clé « teisseire ». Poids du fichier pdf de sortie : 150 ko.

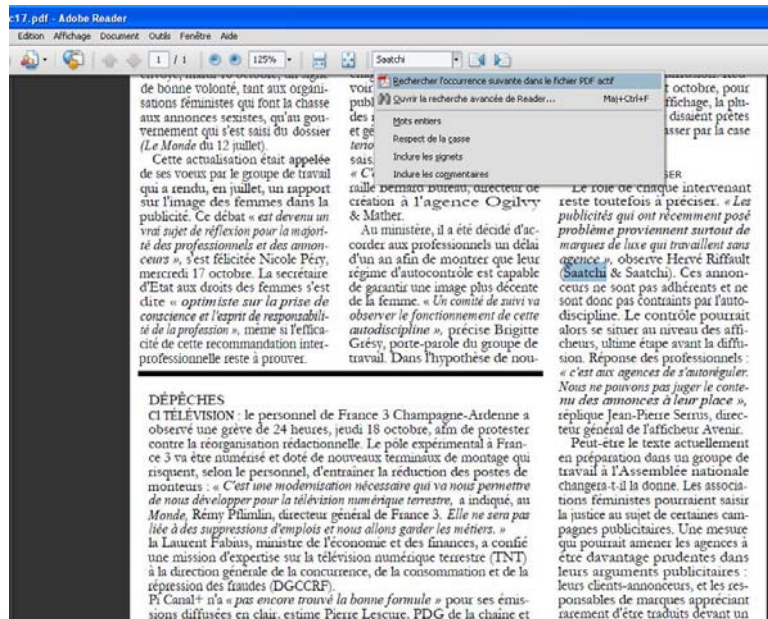
Document 18

C'est une photocopie noir et blanc d'un article du Monde. Le contraste pour cette photocopie est satisfaisant (seulement un très léger « bruit de fond »), et la qualité d'impression des caractères est bonne.

L'article a été recadré. La résolution de numérisation est de 300 dpi.



Sans surprise, l'OCR est quasi parfait, avec une seule erreur.



Recherche plein texte avec le mot-clé « saatchi ».

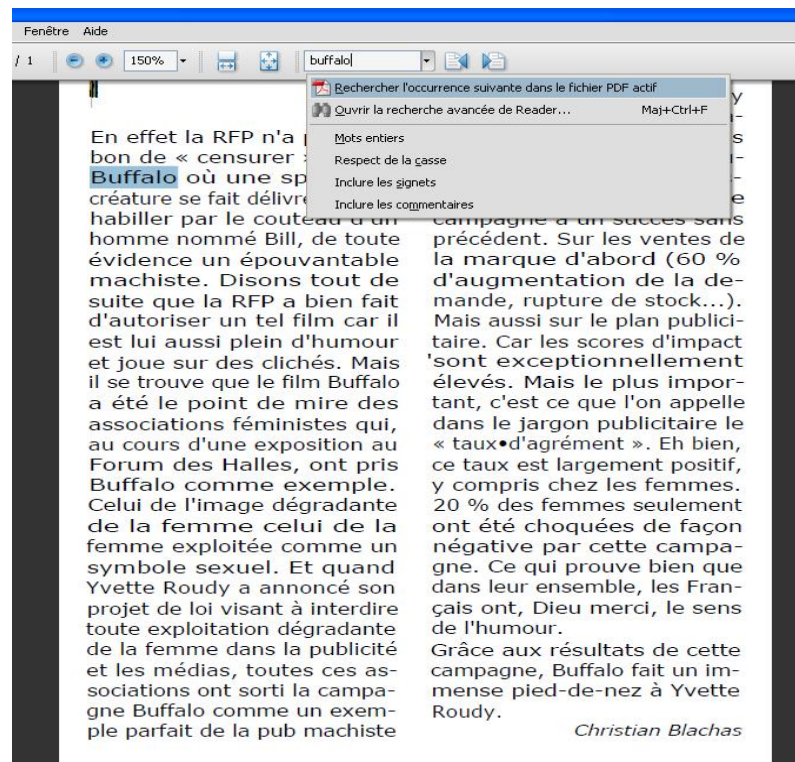
Poids du fichier pdf de sortie : 131 ko. Projection pour 80000 documents : 10 Go

Document 19

C'est une photocopie noir et blanc d'un article dont la source est inconnue ? De plus le titre est assez peu évocateur. A la lecture de l'article on comprend que c'est un article à propos d'une pub pour la société Buffalo. D'où l'intérêt d'une recherche plein texte pour cet article. Le contraste de cette photocopie est bon, et la qualité d'impression des caractères est bonne. L'article a été recadré. La résolution de numérisation est de 300 dpi.



Sans surprise, l'OCR est quasi parfait, avec une seule erreur – portant sur le titre.



Poids du fichier pdf de sortie : 78 ko. Projection pour 80000 documents : 6 Go.

Répertoire des sigles utilisés

Est indiqué en italique le développé en anglais quand il y a lieu, puis la traduction en français.

AFNOR	Association Française de Normalisation www.afnor.org
API	<i>Application Programming Interface</i> Interface de programmation applicative
APROGED	Association des PROfessionnels de la Gestion Electronique de Documents www.aproged.org
ASCII	<i>American Standard Code for Information Interchange</i> Code américain normalisé pour l'échange d'information
CCD	<i>Charge-Coupled Device</i> Dispositif à transfert de charge
CD	<i>Compact Disk</i> Disque compact
CFC	Centre Français d'exploitation du droit de Copie www.cfcopies.com
CMOS	<i>Complementary Metal Oxide Semiconductor</i>
CNRS	Centre National de la Recherche Scientifique www.cnrs.fr
CPI	Code de la Propriété Intellectuelle
DON	Disque Optique Numérique
DPI	<i>Dot Per Inch</i> Point par pouce
DSI	Direction des Systèmes Informatiques
DVD	<i>Digital Versatile Disc</i> Disque numérique polyvalent
GED	Gestion Electronique de Documents
GIF	<i>Graphics Interchange Format</i> Format d'échange d'images
ICR	<i>Intelligent Character Recognition</i> Reconnaissance intelligente de caractères
INA	Institut National de l'Audiovisuel www.ina.fr

ISBD	<i>International Standard Bibliographic Description</i> Description bibliographique internationale normalisée
ISO	<i>International Organization for Standardization</i> Organisation internationale de normalisation www.iso.org/iso/fr/home.htm
J2EE	<i>Java 2 Enterprise Edition</i>
JPEG	<i>Joint Photographic Experts Group</i>
LAD	Lecture Automatique de Documents
LED	<i>Light-Emitting Diode</i> Diode électroluminescente
MOREQ	<i>MOdel REquirements for the management of electronic documents and Records</i>
OCR	<i>Optical Character Recognition</i> Reconnaissance optique de caractères
PNG	<i>Portable Network Graphics</i>
RAID	<i>Redundant Array of Inexpensive Disks</i> Matrice redondante de disques indépendants
RAD	Reconnaissance Automatique de Documents
SAAS	<i>Software As A Service</i> Logiciel en tant que service
SAE	Système d'Archivage Electronique
SSII	Société de Services en Ingénierie Informatique
SSL	<i>Secure Sockets Layer</i>
SVG	<i>Scalable Vector Graphics</i> Graphique vectoriel adaptable
TIFF	<i>Tagged Image File Format</i>
TWAIN	<i>Technology Without An Interesting Name</i>
USB	<i>Universal Serial Bus</i>
W3C	<i>World Wide Web Consortium</i> www.w3.org
WMRA	<i>Write Many Read Always</i>
WORM	<i>Write Once Read Many</i>
XML	<i>eXtensible Markup Language</i>