

SOLUTION GED POUR L'ARCHIVAGE DES MANUSCRITS

A.W.MEDERREG, H.HOCINI, A.SEHAD

Centre de Développement des Technologies Avancées

128, chemin MOHAMED GACEM BP.N°245 EL MADANIA, ALGER

email : amederreg@cdta.dz

I. Introduction

Le livre manuscrit a été quelque chose de vivant avant de devenir sur les rayons de notre bibliothèque mort et réservé à l'usage de quelques érudits. Pourquoi les manuscrits sont morts. Ne devraient ils pas être au contraire une matière toujours vivante. Pour cela, ce patrimoine exige actuellement une vision moderne de conservation, reposant sur la distinction entre le texte réservé aux chercheurs avisés et le manuscrit lui même ; d'où l'archivage électronique de documents.

L'archivage électronique, est une technique informatique qui remplace les traditionnelles archives et redonne vie et valeur aux gisements d'informations qui s'accumulaient dans les caves. Le terme en vigueur est gestion électronique de document et d'information ou GED.

II. LA GESTION ELECTRONIQUE DE DOCUMENTS

Un refrain existait chez tous les archivistes : comment classer, stocker et retrouver un maximum de documents en un minimum de temps et d'espace. Les documentalistes sont-ils en plein rêve, ou va-t-on enfin leur proposer un outil capable de gérer par l'informatique leurs tonnes de papier Jusqu'alors on abordait le plus souvent la GED sous l'angle de l'archivage électronique de l'information, mais cette technique va au delà de cette simple application, elle est maintenant abordée sous l'angle d'une technique de gestion stratégique du patrimoine principal de l'entreprise : ses données et leur disponibilité maximale :

II.1. DEFINITION

La GED, Gestion Electronique de Documents n'est à vrai dire ni une application informatique ni une fin en soi, mais plutôt un ensemble d'outils et de techniques qui permettent de dématérialiser, classer, gérer et stocker des documents à partir d'applications informatiques dans le cadre normal des activités de l'organisme [3].

On perçoit facilement le gain de temps apporté par cette technique. Bien entendu, la GED peut aller au delà de cette simple utilisation; elle permet de gérer des flux d'informations, d'automatiser des processus de diffusion des documents, de construire des bases d'informations multimédia, etc. Néanmoins, il apparaît que ce sont des applications de type classement administratif ou archivage électronique qui, grâce à la GED, connaissent les gains de productivité les plus importants et dégagent le meilleur retour sur investissement.

II.2 LES DIFFERENTES CATEGORIES DE GED

Les applications de la GED se classent en cinq grandes familles qui sont les suivantes :

• La GED Bureautique

Cette GED s'applique au traitement du document dans sa forme d'échange et de gestion des flux d'informations. Les outils de GED Bureautique permettent de manipuler des documents dans leur format Bureautique d'origine (Word, Excel ...), de centraliser leur classement sur un ou plusieurs serveurs, d'échanger ces documents par messagerie électronique... Bien souvent, ils intègrent aussi des fonctions de distribution et de télécopie, d'agenda électronique...

• La GED Technique

La GED Technique, aussi appelée dans certains cas GED Métier, concerne toutes les applications dans lesquelles le concept de Gestion Electronique de Documents est directement liée à une profession.

• La GED Administrative

Ce type d'application concerne généralement ce qu'il est convenu d'appeler la **Gestion de Dossiers Electroniques** et correspond au classement de documents administratifs divers et variés, souvent sous une forme d'image

numérisée. On trouve parmi ces documents des bons de commandes, des factures fournisseurs, des correspondances, des fax...

• La GED Documentaire

La GED Documentaire est directement issue de la recherche documentaire particulièrement utilisée dans le cadre d'applications du type bibliothèques, documentation scientifique, éventuellement revue de presse. La GED a apporté à cette catégorie d'applications l'accès à l'image du document physique ou encore l'accès à des photos (application de type photothèque).

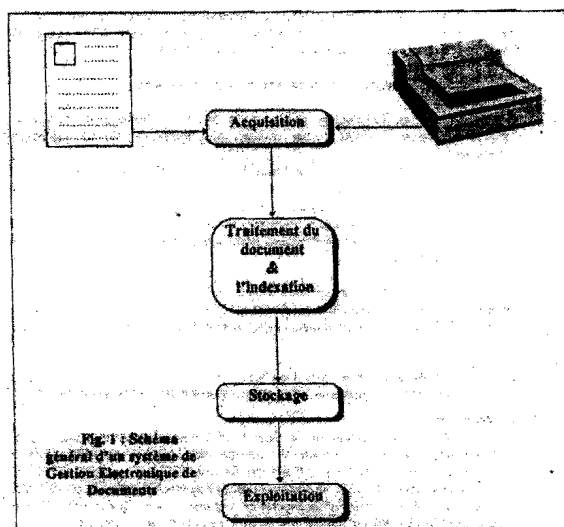
• La GED COLD

Les documents concernés sont les états comptables, les factures, les bulletins de salaire, les relevés de compte...

Cette classification a pour premier objectif de clarifier le principe de la gestion électronique de documents ainsi que ses champs d'application.

II. 3 LES ELEMENTS CONSTITUANT UN SYSTEME GED

Les éléments qui constituent un système de gestion électronique de documents, sont directement liés aux aspects techniques et fonctionnels des plates-formes choisies, car il n'y a pas de système de GED sans études préalables. (Fig. 1)



II.3.1. L'ACQUISITION DES DOCUMENTS

Il est d'abord nécessaire que l'information soit sous une forme électronique pour pouvoir la manipuler, il faut donc en faire l'acquisition. Le symbole de l'application de GED est le scanner ou numériseur qui transforme le document papier en une image électronique manipulable et véhiculable sur ordinateur.

Certes, c'est un des moyens d'acquisition les plus répandus mais il faut rappeler que la GED a pour vocation de fédérer et de gérer tous les types de documents y compris ceux qui se trouvent déjà sous une forme électronique.

Le mode d'acquisition le plus répandu est le scanner, Il permet de numériser une page, une photo, un plan ...

II.3.2. LE TRAITEMENT DU DOCUMENT

A l'issue de la phase d'acquisition, l'image est souvent entachée de bruits, le traitement de celle-ci à l'aide de techniques appropriées, sert à faire ressortir la partie utile du fond de l'image.

Les images acquises ne peuvent échapper aux effets de dégradation dûs essentiellement aux phénomènes physiques tels que:

- Une diffraction du système optique.
- Un flou à cause du mouvement de l'image durant son acquisition.
- Des erreurs dues à l'environnement (manque d'éclairage, positionnement...

L'atténuation de ces dégradations pour une interprétation plus fiable de l'image se fait à l'aide de différentes méthodes parmi elles :

- Les opérations basées sur l'examen de l'histogramme. Elles servent à éliminer certains défauts de prise de vue agissant sur la dynamique de l'image.
- Le filtrage, qui a pour objet de diminuer les effets de bruits dans l'image.

II.3.3. L'INDEXATION

L'indexation des documents est l'opération qui consiste à décrire les documents en vue de leur exploitation ultérieure. Elle consiste très exactement à décrire le document avec des mots ou des expressions convenus préalablement.

Pour accéder à un document, il est nécessaire de disposer de sa description, celle-ci va, porter sur deux plans distincts mais complémentaires :

Une description externe contenant des informations sur le type de document, son origine, la date de sa prise en charge ou de sa création, pour les activités administratives ou techniques, le rattachement aux objets de base de l'entreprise (client, fournisseur, produit, etc., ...).

Une description du contenu, les enjeux de l'indexation et les difficultés se situent à ce niveau.

Cette description ainsi que certaines autres informations propres au système d'informations de l'organisme sont contenues dans une "Carte d'identité" appelée aussi fiche suiveuse, matérialisée par un enregistrement dans une base de données classique qui constitue ainsi un répertoire des documents présents dans le système [2].

La carte d'identité du document est considérée comme un document secondaire synthétisant les informations du document original, dit document primaire. De base, elle contient un certain nombre de champs obligatoires comme la date d'acquisition ou de production (respectivement par exemple pour un document scanné ou un spool), un numéro d'entrée dans le système, l'identification de l'expéditeur ou du créateur, du destinataire, etc. Elle peut être complétée par de nombreux champs contenant des mots clés, un résumé du document primaire, une date obligatoire de traitement, ...

On voit que certains champs qui apparaissent comme d'intérêt secondaire auront pour certains types de documents ou même pour certaines entreprises beaucoup d'importance alors que d'autres sont superflus.

Pour acquérir des indexes on a le choix entre plusieurs techniques :[4] :

- L'acquisition manuelle,
- L'extraction automatique des mots clés par programme,
- La reconnaissance optique des caractères « OCR »
- L'analyse « full-text »(texte intégral) des documents
- Les codes à barres

II.3.4. STOCKAGE ET CONSERVATION

La conservation est un autre élément essentiel d'un système de GED. Elle aussi doit être étudiée avec minutie avant que ne soit défini le type de périphérique de stockage (disque magnétique, disque optique réinscriptible ou non réinscriptible, CD-WORM juke-box) et que le stockage soit hiérarchisé en fonction des types de documents, de leur fréquence de consultation, de l'intérêt à les avoir en ligne, de la durée de la conservation et de l'importance stratégique de l'information. Chaque type de stockage possède des caractéristiques précises qui le rend plus ou moins apte à satisfaire tel besoin ou tel budget. Il faut souligner au passage que, bien que continuant à évoluer rapidement, la technologie est aujourd'hui parfaitement maîtrisée et les moyens de stockage des systèmes de GED sont désormais parfaitement fiables et performants [5].

Les périphériques de stockage utilisés pour la GED sont pour une partie communs à d'autres applications informatiques. Ainsi, il est fréquent d'utiliser des disques magnétiques en premier stockage. En revanche, certains périphériques, sans être exclusivement réservés à la GED, lui sont plus souvent associés qu'à d'autres applications. Ce sont les périphériques à base de disques optiques numériques ou DON.

II.3.5. EXPLOITATION

Une des finalités essentielles d'un système de GED est de rechercher des documents, soit pour les restituer à l'utilisateur qui les consultera, soit pour les communiquer à un autre utilisateur ou les diffuser à une ou plusieurs personnes. La qualité et la rapidité de la recherche dépendent de l'indexation choisie et de la qualité de l'acquisition.

La recherche de documents se fait par l'intermédiaire d'un langage qui peut être réduit à sa plus simple expression, par exemple une liste de mots clés combinés ou non, ou elle peut être très complexes et faire appel, par exemple, à des techniques d'interrogation en langage naturel [1].

Le premier objectif de ces langages, à la bonne fin duquel contribue la méthode d'indexation, est la pertinence des recherches qui doit conduire à limiter le "bruit", c'est-à-dire la sélection de documents ne répondant pas exactement à la requête, et le "silence", c'est à dire le défaut de sélection de documents disponibles qui répondent à la requête. Plusieurs techniques

existent et sont mises en œuvre dans les systèmes de GED, soit indépendamment et exclusivement les unes des autres, soit pour certaines d'entre elles de façon combinée. Les principales méthodes de recherche sont:

- La recherche booléenne qui consiste à présenter la requête sous la forme d'une expression booléenne de mots clés appartenant au dictionnaire ou au thesaurus, mots séparés par des opérateurs booléens (ET, OU, SAUF, ...).
- La recherche plein texte ou texte intégral, qui évite l'indexation préalable des documents; ils sont décrits par eux même, c'est-à-dire par la chaîne des caractères qui les constitue, Les recherches booléennes s'appliquent également à ce cas. Cependant, on associe souvent à cette technique des méthodes d'interrogation dites en langage naturel, c'est-à-dire reposant sur une analyse linguistique de la requête de l'utilisateur et conduisant à l'élaboration d'une requête 'interne' à partir de laquelle est réalisée la recherche.
- Recherche par navigation ou technique hypertextuelle; cette technique suppose une structuration préalable des documents avec la mise en œuvre des liens hypertextuels.

Une fois sélectionnés, les documents peuvent être :

- transmis pour traitement à des systèmes informatiques, communiqués par messagerie (par exemple en pièce jointe) ou encore envoyés par télécopie.
- Affichés sur des écrans de consultation.
- Imprimés localement ou à distance sur des imprimantes désignées.
- Envoyés par réseau vers d'autres postes de travail (fréquemment sous le contrôle d'un outil de workflow ou de groupware).

III. LES DIFFERENTES ARCHITECTURES D'UN SYSTEME DE GED

On rencontre deux types d'architecture dans les systèmes de gestion électronique de documents :

•La station autonome

Elle est bâtie autour d'un micro avec un scanner, une imprimante et une unité d'écriture-lecture

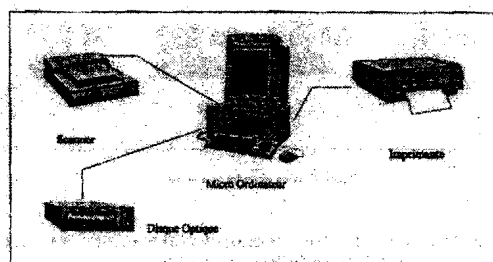


Fig. 2 : Station autonome

•La station répartie en réseau

Cette architecture est surtout construite autour de micro-ordinateurs. Un premier PC gère la base de données, un second se transforme en station d'archivage avec scanner, un troisième administre le réseau, d'autres pilotent des DON. Les postes de consultation sont chacun sur un micro-ordinateur. Des configurations réparties existent aussi autour de mini-ordinateurs.

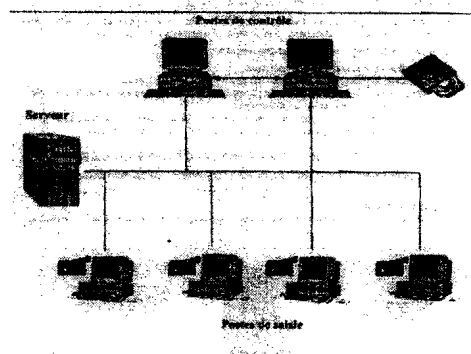


Fig.3 : Station répartie en réseau

IV. ETUDE D'UNE SOLUTION : L'ARCHIVAGE DES MANUSCRITS DE LA BIBLIOTHEQUE NATIONALE D'ALGERIE

IV. 1 LES CARACTERISTIQUES DES MANUSCRITS

Le manuscrit offre à la lecture un objet transformé au fil des siècles c'est ce qui fait sa particularité. Ses principales caractéristiques sont

1. L'écriture

Comme son nom l'indique, le manuscrit est écrit à la main. Il a été copié, recopié, corrigé et même parfois annoté par un ou plusieurs copistes, à des époques très lointaines qui remonteraient au troisième siècle hégire, ceci a permis essentiellement de mettre en circulation un exemplaire corrigé et fidèle d'un texte.

2. Le support

Le manuscrit contrairement à l'imprimé est écrit sur plusieurs supports de différentes matières :

•Le papyrus

Ce sont des feuilles formées de minces lamelles de roseau. Cette matière est très fragile, ceci explique la disparition presque totale des manuscrits écrits sur ce support.

•Le parchemin

Les découvertes récentes montrent que le parchemin est un très ancien support d'écriture. Il était fabriqué principalement à partir de peau de mouton, de chèvre ou de veau. Son utilisation soulevait un important problème économique.

•Le papler

Son utilisation était rare jusqu'au milieu du quatorzième siècle car le papier coûtait cher, d'ailleurs plusieurs manuscrits de différents thèmes, l'écrits par différents auteurs existent dans un même et seul livre appelé recueil. Le format de ses feuilles fournit de précieux renseignements sur l'origine du manuscrit et la date de sa copie.

3. Texte et illustration

Il existe certaines mutations qui concernent l'organisation et le repérage du texte lui même; d'autres modifient les supports qu'entretiennent le texte et son illustration enluminée.

Par une part l'ornementation a pur but de rendre mieux lisibles les articulations et la construction du texte : les initiales figurées et historiées en soulignent les divisions, les dessins des marges peuvent en proposer une hiérarchisation, les miniatures ont charge d'en transmettre, littéralement ou symboliquement, le sens.

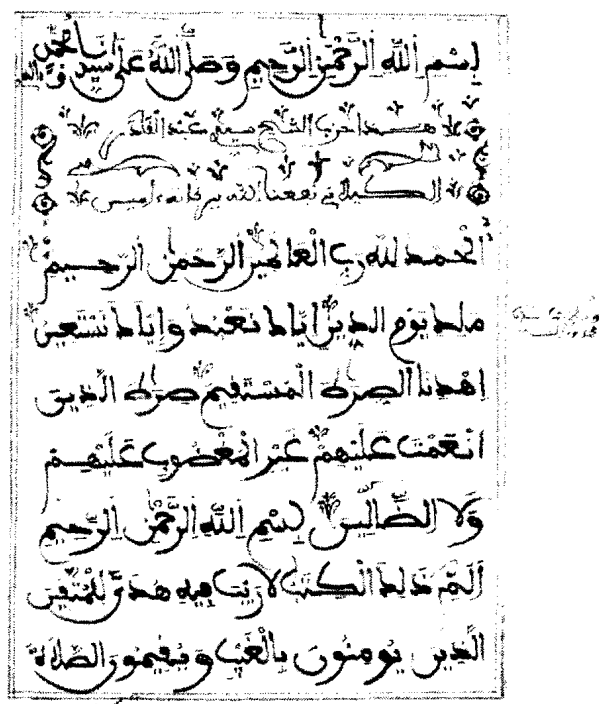


Fig VI.1 : Exemple de manuscrit

Description et format

Le manuscrit est décrit par l'encadrement du texte, la régularité des lignes, le tracé des lettres au calame, orné de miniatures géométriques florales et feuillures, il est enluminé de couleur noire de l'encre, bleue, rouge, jaune, vert, traversé de filets d'or, tous témoins d'une époque évoquant le beau.

Le texte est généralement écrit sur une grande partie de la page mais reste une partie appelée marge, réservée pour les corrections et les remarques explicatives sur le contenu, l'auteur, ... si ceci concerne une annotation.

Pour ce qui est de la pagination, une tomason est mise en bas de la page à gauche, indiquant le premier mot de la page suivante.

IV.2 LES MANUSCRITS DE LA BIBLIOTHEQUE NATIONALE D'ALGERIE

D'après les dernières statistiques de l'année 1997, la bibliothèque nationale contiendrait 3609 manuscrits variés relatifs à toutes les branches surtout à usage didactique.

On y trouve des collections diverses de Coran, de tous les formats, de tous les types d'écritures, de tous les grammages de papier, une grande quantité d'ouvrages relatifs aux sciences coraniques et d'autres ouvrages de grammaire, de lexicographie, de mystique, d'astronomie, de mathématiques, beaucoup de commentaires et de biographies de saints.

Le plus ancien manuscrit que possède notre Bibliothèque Nationale date du quatrième siècle hégire, c'est une partie du Coran écrite en coufique sur parchemin.

•Description du manuscrit

Les manuscrits sont classés par ordre alphabétique des titres, ils sont décrits par une fiche signalétique qui contient en général:

- La côte du manuscrit.
- Son numéro d'ordre.
- Le titre du manuscrit.
- Le nom de l'auteur et la date de sa mort (aide essentiellement à retrouver l'époque où a été écrit le manuscrit).
- Les références bibliographiques.
- Le type d'écriture.
- La date de la copie.
- Le nombre de pages.
- Le format en centimètres.
- Enfin, la mention d'une édition avec tous les détails qui s'y rapportent.

•Supports des manuscrits de la BN

Le support des manuscrits va du parchemin, le plus ancien -peaux de gazelle, de veau, de bouc, de chèvre - au papier passant par le papyrus.

•Les écritures

L'écriture de même est variée. On trouve tous les spécimens de calligraphies, de type coufique, oriental, andalou, africain, berbère et principalement l'arabe, le persan et le turc.

•Format des manuscrits de la BN

Les formats des manuscrits diffèrent aussi, on y trouve des rouleaux, des codex (assemblage de feuillets pliés, groupés en un ou plusieurs cahiers) et des livres de différentes dimensions.

Le format est A3 A4 et plus petit. Les manuscrits constituent le fond initial de la bibliothèque nationale, au nombre de 3609 ouvrages dont 2935 en format inférieur à A4.

•Etat des manuscrits de la BN

Le plus souvent, la qualité du manuscrit est mauvaise : les lettres sont effacées, les feuilles sont trouées, déchirées, jaunies, les pages sont arrachées. La raison en est, bien sûr, les intempéries –l'eau, le soleil et l'humidité –, les mauvais traitements, les microbes, le manque de soins et les déménagements. L'âge des manuscrits souvent très anciens, est aussi une des raisons de leur mauvais état : certains datent de la période almohade, soit du V^{ème} et VI^{ème} X^{ème} et XII^{ème} siècles.

Les manuscrits, bien que placés à l'abri dans le magasin qui leur est propre, sont traités pour être contre ces atteintes. La bibliothèque dispose d'un atelier de reliure., d'un atelier de restauration raccords, collage des feuilles - mais il est préférable de trouver une solution qui éliminera définitivement la manipulation des manuscrits pour mieux les protéger.

IV.3. PROPOSITION DUN SGED ADAPTE AUX MANUSCRITS DE LA BIBLIOTHEQUE NATIONALE

•Problématique

Beaucoup d'ouvrages ont été la proie des vers, des moisissures, de l'humidité et du temps. Certaines pages d'un ouvrage sont d'une écriture différente du reste de l'ouvrage, écrites à des périodes successives. Des pages sont noircies, jaunies, trouées, l'écriture est effacée. A beaucoup de volumes il manque des pages. On tombe ça et là sur des lambeaux de feuilles volantes, sur des

manuscrits mutilés, sans page initiale, ni page finale ou sur des liasses. Cependant, beaucoup de livres ont pu être reconstitués, à partir de feuillets dispersés mais on ne peut affirmer que tout est reconstitué, restauré et rétabli.

Le projet GED de la Bibliothèque Nationale rentre dans le cadre de l'organisation générale de cette dernière et l'organisation de l'information en particulier. Mais le but essentiel est de conserver et revaloriser le fond précieux des manuscrits.

•Les recommandations

Sur la base des informations présentées ci dessus, recueillies au niveau du service des manuscrits de la bibliothèque, nous avons pris les décisions suivantes :

- La quantité de documents à numériser étant très importante, il est préférable de procéder par tranches. Les documents dont le format est inférieur au format A4 seront numérisés en premier car ces documents :
 - Représentent plus des deux tiers du fond.
- Ils ne demandent pas un matériel très coûteux comparativement aux autres formats. .
- L'état des documents étant mauvais dans la plupart des cas et leur manipulation risque de les affecter encore plus, pour cela, l'utilisation de caméra est préférable aux scanners.
- Les feuillets présentent un état de dégradation apparent, l'image numérisée risque d'être de mauvaises qualité. Après l'étape de numérisation des manuscrits, des outils de traitement d'images peuvent être utilisés pour l'amélioration de la qualité des images numérisées.

•Solution

La Bibliothèque Nationale a adopté la solution GED pour diverses raisons. Le document électronique ne peut être consulté que si l'on a l'accès, la disposition , la connaissance du matériel et les autorisations d'accès.

•Objectifs

La solution proposée est l'archivage électronique connue sous l'appellation de Gestion Electronique de Documents. Elle est basée sur les informations et données propres à la Bibliothèque Nationale. Les objectifs visés pour établir cette solution sont :

- Approche modulaire et par étape,
- Solution ouverte, extensible et accessible par d'autres environnements.
- Coût minimum
- Souplesse et facilité de manipulation des données stockés en adoptant l'environnement Windows et les formats compatibles de sauvegarde d'images (tiff, jpeg).

•La mise en œuvre

Avant de mettre le projet en œuvre, il faut définir ce que l'on cherche à obtenir ou à éviter. L'enjeu est suffisamment important pour devoir procéder à une étude rigoureuse. Pour cela, la mise en œuvre de cette opération se fera en deux phases :

Phase 1 : Numériser et cataloguer les documents dont le format est inférieur ou égal au format A4 (nécessite un matériel peu coûteux et disponible au niveau de la bibliothèque nationale). La durée estimée pour la saisie des 2935 documents, l'équivalent de 1.500.000 pages est d'environ une année.

Phase 2 : Numériser et cataloguer les documents dont le format est supérieur à A4 en second lieu car ceci nécessite un matériel coûteux tel qu'un scanner pour format AO ou caméra à grande résolution.

V. LE SYSTEME PROJETE POUR LA BIBLIOTHEQUE NATIONALE

Un système de gestion électronique de documents "SGED" est une chaîne complète de traitement de l'information. Plusieurs types d'équipements informatiques, matériels et logiciels sont réunis dans un seul but : mettre à la disposition de tous, un document pertinent.

Le système projeté pour la bibliothèque est constitué de deux modules :

•Le module de saisie

➤La numérisation et le stockage

Avant que le document ne soit exploitable en environnement informatique, il faudrait au préalable le numériser, par la suite , procéder à son stockage. Cette opération est assurée dans un SGED par le module de saisie.

➤ **Le choix du support**

• **Le CD-WORM "CD-R"**

Pour stocker la masse de documents existants au sein de service de manuscrits de la Bibliothèque Nationale, nous avons choisi le support optique CD-R (CD-Recordable) qui est aujourd'hui considéré comme l'un des plus économiques et les plus fiables de tous les supports, vu sa durée de vie qui dépasse facilement le siècle. C'est un support très simple à utiliser, solide pratiquement inusable, de plus, il répond à des applications particulières tel que l'archivage.

• **Le Juke-boxe**

Cette solution convient lorsque plusieurs dizaines voir plusieurs centaines de CD-R existent. La solution réside dans le chargement automatique des supports par un automate qui est le juke-boxe. Ce dernier chargera le bon support optique dans le bon lecteur au bon moment.

• **Le module de consultation et de recherche**

Les CD-R créés par le module de saisie et de stockage peuvent être consultés. Une recherche paramètre suivant le titre, le nom d'auteur, le numéro de côte, ... etc. est possible. Le résultat peut être consulté ou imprimé.

VI. CONCLUSION

La Bibliothèque Nationale demeure un lieu de curiosité pour les bibliophiles à la recherche d'anciens livres et de précieux manuscrits. Elle doit les mettre à la disposition de tous, c'est ce qu'elle envisage en adoptant le système de gestion électronique de documents aux mesures des possibilités matérielles, financières et humaines.

L'adoption d'une telle technique au sein de la Bibliothèque Nationale lui apporte un complément de moyens, elle ne supprime pas le manuscrit "sur support papier, il restera conservé sur les rayons mais allégera sa gestion, mieux encore, elle facilitera l'accès pour sa consultation.

Pour répondre aux préoccupations les plus diverses en matière de circulation de l'information, prenant en compte les variables propres à chaque projet : volume des manuscrits, natures de leurs supports primaires et des supports de stockage les plus adaptés et surtout aisance d'accès et de consultation, l'intégration de l'outil SGED est nécessaire au sein de la Bibliothèque Nationale.

