

**République Algérienne Démocratique et Populaire**  
**Ministère de l'enseignement supérieur**  
**et de la recherche scientifique**

**Université du 08 mai 45, Guelma**



**Faculté des sciences économiques et commerciales**  
**et sciences de gestion**  
**Département des sciences de gestion**

Mémoire présenté pour l'obtention  
du diplôme de Master en sciences de gestion  
Option: Techniques d'information et de communication  
dans l'entreprise

**Conception d'un prototype d'un système**  
**d'archivage électronique documentaire**  
(au niveau du centre régional des archives, Wilaya de Guelma)

**Présenté par**  
Nebti Bilal  
Ben nacer Mohammed Yassine

**Sous la direction de** Kelaiaia Abdessalem

**Juin 2011**



# Résumé

L'archivage électronique devient le sujet d'actualité pour un bon nombre d'organisations tant publique que privées. Ceci s'explique certes par une augmentation extrêmement forte du volume de données électroniques gérées au quotidien, mais aussi par une évolution des technologies.

L'archivage électronique ne se limite pas à une simple dématérialisation des techniques d'archivage traditionnelles. Outre l'influence des nouvelles obligations, ce nouveau type d'archivage doit être considéré très en amont dans la chaîne de valeur de l'information.

En effet dès que l'information existe sous forme électronique et que son intégrité est garantie les seules modifications qu'elle puisse subir sont celles relatives à un changement de support, voir un système d'accès plus sophistiqué en matière d'indexation.

Notre travail se veut d'apporter un premier pas pour une solution aux divers problèmes liés à l'archivage au sein du centre régional des archives tant sur le niveau organisationnelle que sur le niveau opérationnel. Ainsi, une restructuration de l'organisation du service d'archivage actuel et la conception d'un prototype d'un système d'archivage électronique documentaire sont proposés pour rendre accessible les archives aux différents utilisateurs.

# Abstract

Electronic archiving becomes the hot topic for many organizations both public and private. This is explained by an increase in extremely high volume of electronic data managed on a daily basis, but also by evolving technologies.

Electronic archiving is not limited to mere computerization of traditional archiving techniques. Besides the influence of new bonds, this new type of archiving is to be considered very early in the value chain of information.

Indeed as soon as the information exists in electronic form and that its integrity is guaranteed that the only changes are those relating to undergo a change of medium, see a more sophisticated access system for indexing.

Our work is intended to provide a first step towards a solution to various problems related to archiving in the regional center of archives on the organizational level as on the operational level. Thus, a restructuring of the organization's existing archiving service and design a prototype of an electronic document filing system are available to make the archives accessible to different users.

# Remerciements

En préambule à ce mémoire, nous souhaitons adresser ici tous nos remerciements aux personnes qui nous ont apporté leur aide et qui ont ainsi contribué à l'élaboration de ce mémoire.

Tout d'abord Monsieur Kelaiaia Abdessalem, directeur de ce mémoire, pour l'aide et le temps qu'il a bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.

Nous exprimons notre gratitude à toutes les personnes qui ont accepté de répondre à nos questions avec gentillesse.

Nous remercions le chef de service d'archive au niveau du centre régional des archives pour la confiance qu'il nous a manifesté et pour le partage de sa riche expérience.

Enfin, Nous adressons nos plus sincères remerciements à tous nos proches et amis qui nous ont toujours soutenu et encouragé au cours de la réalisation de ce mémoire.

# Dédicace



*Ce travail est dédié à nos chers parents pour leurs amours et sacrifices, nos frères et sœurs pour leurs soutiens et encouragements, nos amis et à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce projet.*

# Table des matières

Introduction générale

## Chapitre I : Archivage, définition et rôle

1. Archivage.....	11
1.1. Définition du mot archive. ....	13
1.2. Autre sens du mot archive. ....	13
1.2.1. Service administratif.....	13
1.2.2. Les bâtiments.....	14
1.3. Archives informatiques.....	15
2. Les outils de l'archivage.....	16
3. Processus d'archivage.....	18
4. De bonnes raisons d'archiver.....	20
5. Cycle de vie des archives.....	21
5.1. Phase d'activité.....	22
5.2. Phase de semi-activité.....	22
5.3. Phase d'inactivité.....	23
6. Evolution de l'archivage.....	24
6.1. Multiplication des services d'archives.....	24
6.2. Un plus grand besoin d'archiver.....	24
6.3. <i>Records Management</i> : modernisation de l'archivage.....	24
6.4. L'apparition du numérique.....	25
7. Place de l'archivage dans l'entreprise.....	26
7.1. Enjeux de l'archivage.....	26
7.2. Perception de l'archivage.....	28
8. Archivage papier vs archivage électronique.....	29
8.1. Une valeur probante similaire.....	29
8.2. Des règles de gestion identiques.....	30
8.3. De profondes différences.....	31

## Chapitre II : Archivage électronique des documents État de l'art

1. Document électronique .....	35
1.1. Définition du document.....	35
1.2. Définition du document électronique.....	35
1.3. Traitement des documents numériques (Dématérialisation).....	35
1.3.1. Numérisation des documents .....	35
1.3.2. Reconnaissance Optique de Caractères (OCR).....	35
1.3.3. Détermination de la structure d'un document.....	36
1.3.4. Formats de représentation des documents.....	37
1.3.5. Indexation.....	39
1.3.6. Recherche d'information.....	39
2. La gestion électronique des documents (GED).....	40
2.1. Définition.....	40
2.2. Avantages de la Gestion Electronique de Documents.....	40
2.3. Les différents types de GED.....	41
2.3.1. La GED technique.....	41
2.3.2. La GED documentaire .....	41
2.3.3. La GED COLD .....	41
2.3.4. La GED administrative.....	41
3. De la GED au <i>Records Management</i> .....	41
3.1. Qu'est ce que le <i>record management</i> .....	41
3.2. Document management/GED et Records management/Archivage .....	41
4. Archivage électronique des documents à valeur probante .....	43
4.1. Les enjeux de la conservation sur le très long terme : l'obsolescence technologique..	43
4.1.1. Le vieillissement du support .....	43
4.1.2. L'obsolescence des techniques .....	43
4.2. La valeur probante d'un document .....	45
4.3. Présentation de l'archivage électronique.....	45
4.4. Objectifs de l'archivage électronique.....	46
4.4.1. Pérennité/réversibilité.....	46
4.4.2. Communication.....	46
4.4.3. Traçabilité.....	47
4.4.4. Protection des données et archivage légal.....	47



4.5. Exigences de l'archivage électronique.....	47
4.6. L'organisation type d'un système d'archivage électronique.....	48
4.6.1. Les acteurs du processus.....	48
4.6.2. Le contrat.....	49
4.6.3. La protection des données.....	49
4.6.4. Le processus type d'Archivage électronique.....	50
5. Cadre normatif.....	50
5.1. MOREQ2.....	51
5.2. DUBLIN CORE.....	51
5.3. NF Z 42-013.....	55
5.4. OAIS <i>Open Archival Information System</i> .....	55
5.5. ISAD (G).....	58
5.6. <i>EAD (Encoded Archival Description)</i> .....	64
5.7. La norme ISO 15489 en réponse aux besoins d'archivage.....	65
Conclusion.....	66
<b>Chapitre III : Réalisation d'un prototype d'un système d'archivage électronique</b>	
1. Présentation de l'environnement d'étude.....	67
1.1. Présentation.....	67
1.2. Mission.....	67
1.3. Organigramme.....	67
1.4. Fonctionnement.....	67
2. Résultats de l'étude menée.....	68
3. Solution proposée.....	69
3.1. Solution organisationnelle et fonctionnelle.....	69
3.2. Description de la solution.....	71
3.2.1. Bureau de Dématérialisation (numérisation).....	71
3.2.1.1. Processus de numérisation.....	71
3.2.1.2. Solution technique proposée pour la dématérialisation.....	74
3.2.2. Bureau d'indexation et de recherche.....	78
3.2.2.1. La recherche d'information.....	78
3.2.2.2. Processus de recherche d'information.....	78
3.2.2.3. Solution technique proposée pour l'indexation et la recherche de L'information.....	84
<b>Conclusion générale.....</b>	<b>87</b>

## Liste des figures

Figure 1.1 : Théorie des trois Ages

Figure 2.1 : Processus type d'Archivage électronique

Figure 3.1 : Organigramme du centre régional des archives

Figure 3.2 : Schéma organisationnel de la structure proposée

Figure 3.3 : Schéma fonctionnel de la structure proposée.

Figure 3.4 : Processus de dématérialisation

Figure 3.5 : Tâches rapides *ABBYY FineReader*

Figure 3.6 : Fenêtre principale de *ABBYY FineReader*

Figure 3.7: Processus général de recherche d'information

Figure 3.8 : Interface principal de *DocFetcher*

Figure 3.9: *DocFetcher* en mode minimaliste

## Liste des tableaux

Tableau 1 : Tableau des différences essentielles entre les deux systèmes GED/SAE

# Introduction Générale

L'informatique en tant qu'outil de travail a évolué d'une façon telle que son utilisation dans n'importe quel domaine, soulage l'individu dans sa communication avec son environnement. Elle permet, par définition l'accès rapide à l'information de n'importe ordre et cela, en la gérant automatiquement. Un service d'archives quel que soit le degré de son importance brasse quotidiennement des masses d'informations. Sa mission fondamentale, loin d'être axée uniquement sur "La conservation" est particulièrement orientée vers "La communication". En quelque sorte un service d'archives peut avoir droit de cité que s'il communique les informations qu'il conserve à l'intérieur des milliers de liasses disposées sur ses rayonnages au fond de ses magasins. Collecter, traiter, conserver et enfin, communiquer telles sont les missions initiales d'un service d'archives.

Chacune des ses tâches est intimement liée aux autres. En d'autres termes, aucune d'entre elles n'aurait de valeur si elle est séparée des autres tâches.

Il est certes bien établi, que jusqu'à présent l'archiviste a accompli sa mission en utilisant les moyens traditionnels mais, toujours est-il que l'accroissement considérable de l'offre et la demande en matière d'informations commence à le dépasser au point de le confiner dans un rôle passif alors qu'en réalité, il devrait jouer tout simplement un rôle dynamique.

A l'heure actuelle notre monde est divisé en deux groupes bien distincts : Un groupe rapidement informé capable de prendre les décisions opportunes face aux problèmes qu'il rencontre et un autre groupe non informé ou informé en retard, qui subit les décisions du premier. Le premier a su accorder une valeur à ses différentes structures d'information, alors que le second, les a tout bonnement ignorées.

En raison de sa fonction de gestionnaire de l'information, un service d'archives s'inscrit parmi ces secteurs. Et comme il a été signalé plus haut son rôle consiste à communiquer et de ce fait, l'informatique en tant qu'outil, doit nécessairement le soulager dans sa tâche, à condition de l'adapter aux besoins propres de sa tâche.

## **Problématique et objectif visé**

Aujourd'hui, avec la poussée technologique et la compétitivité accrue dans le domaine des nouvelles technologies de communication et de l'information, les organisations sont inévitablement, voire obligées à prendre en compte toute évolution informatique pour faciliter

la gestion de l'information. Les archives documentaires ont été toujours au premier plan de cet objectif.

L'archivage électronique ne doit pas être perçu comme une simple transformation de l'archivage traditionnel papier en électronique. Il correspond à une nouvelle organisation des données dans l'entreprise indispensable à prendre en compte.

L'objectif principal de ce mémoire est la conception d'un prototype d'un système d'archivage électronique (organisationnel et opérationnel) du centre régionale des archives, qui doit permettre aux service d'archivage de conserver les documents électronique et les récupérer de façon facile.

Expert PDF Evaluation

# Chapitre I

## Archivage, définition et rôle

Ce chapitre est consacré à une étude bibliographique qui précise l'exploitation de la notion d'archivage, qui est encore définie de multiples façons. Nous verrons que depuis plusieurs années, notamment avec l'arrivée de l'internet et des nouvelles technologies, l'archivage a pris de l'importance et est maintenant perçu d'une toute autre façon par les entreprises. Cette nouvelle aire est également caractérisée par l'expansion de l'archivage électronique, dont nous parlerons ensuite plus en détail.

### 1. Archivage

L'Archivage est un terme, à la fois récent dans la langue française, puisqu'il n'est utilisé que depuis quelques décennies seulement, et complexe par le fait qu'il n'existe pas de définition légale du terme « archivage », alors qu'il en existe une du mot « archives », dans la loi n°88-09 , qui définit ce terme comme « l'ensemble des documents contenant une information, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. » [jora, 1988a]

Plusieurs auteurs ont prêté à ce terme des définitions différentes. L'archivage est défini comme « l'ensemble des actions, outils et méthodes mises en œuvre pour conserver à moyen ou long terme des informations dans le but de les exploiter » [Sédallian, 2002] c'est-à-dire les actions qui conduisent à ce qu'un document soit capté, stocké, indexé de façon à le rendre accessible par la suite. Cette définition, aujourd'hui, est obsolète car l'archivage touche l'ensemble des documents, les trois âges.

Il se définit au si comme étant « la démarche d'organisation qui a pour objectif d'identifier, de mettre en sécurité et de maintenir disponibles l'ensemble des documents qui engagent une entreprise ou un organisme vis-à-vis de tiers ou de son activité future et dont le défaut représenterait un risque ». [Chabin. 2010a]

L'unité de base de l'archivage tourne autour de ce qu'on appelle un document, qui est considéré comme une entité physique constituée par un support individualisé sur lequel sont fixées des informations. Ces deux définitions montrent que l'archivage n'englobe pas tous les

documents, toutes les données ou toutes les informations, mais seulement les documents qui engagent leur détenteur. Ces documents, originaux ou copies, constitués ou non en dossiers, sont fixés sur des supports, qui peuvent être physiques ou électroniques, c'est ce dont nous parlerons dans une prochaine partie.

Aujourd'hui, l'archivage est perçu différemment, il est devenu la démarche managériale de contrôle du cycle de vie des documents à risque dans l'entreprise. Cette nouvelle perception de l'archivage a entraîné la création d'un système s'appuyant sur la gestion des documents utiles à un organisme depuis leur création jusqu'à la fin de leur vie administrative et qui assure leur fiabilité, leur intégrité et accessibilité. Il s'agit du « *Records Management* », expression anglo-saxonne ayant pour signification approximative en français « gestion de l'archivage », le Records management désigne « un champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liées à la forme des documents ». **[Records, 2008]**

Autrement dit, c'est un système qui accompagne et gère les documents de leur création à l'extinction de leur utilité par le producteur en passant par leur réception et leur conservation.

Le records management est en quelque sorte la modernisation de la fonction archivage, sauf qu'il ne prend en compte que la gestion des archives courantes et intermédiaires, les archives définitives n'étant pas prises en charge dans ce système. Nous verrons que les archives numériques prennent de plus en plus d'importance au sein de cette procédure. L'archivage électronique est devenu un réel défi pour le *records management*, le support numérique ayant autant de valeur que le support papier. Le *Records Management* symbolise d'une certaine façon l'évolution de l'archivage, thème qui sera abordé plus bas.

L'archivage, comme le précise la définition donnée, constitue un ensemble d'outils; en effet, il suppose la réalisation de tableaux de gestion des archives, de plans de classement et d'instruments de recherche, de règles de conservation, etc....

### **1.1. Définition du mot archive**

« Une archive est un document ancien, et essentiellement sous forme papier... voire parchemin ! C'est un document administratif, de la "paperasse" dont on n'a plus besoin, mais qu'on est obligé de garder... Alors on le stocke dans une boîte ou un classeur, à la cave ou dans un hangar... » **[STS, 2009]**. Cette définition simpliste est évidemment très réductrice et ne résiste pas à l'analyse :

La notion traditionnelle de document (càd une ou plusieurs pages de papier) a totalement évolué depuis l'avènement de l'informatique : Un fichier Microsoft Word ou un e-mail est-il un document ? Oui, si l'on se réfère à l'usage qu'on en fera, mais évidemment non dans sa forme !

Une archive n'est pas nécessairement ancienne. En fait dès lors qu'il n'est plus appelé à évoluer, un document est éligible au statut d'archive. Une facture émise, par exemple, n'évoluera plus jamais. Quand bien même elle comporterait des erreurs, on émettra un avoir, ou une facture complémentaire, mais la facture initiale ne sera pas modifiée pour autant et devra être conservée.

Si l'on admet le point précédent, l'idée selon laquelle une archive est un objet dont on n'a plus besoin, se révèle tout à fait inexacte !

S'il n'existe évidemment pas de définition « universelle » de la notion d'archive, on peut néanmoins, en prenant en considération ces différents points, en imaginer une plus rationnelle. A cet égard, la loi algérienne, en propose une, bien plus satisfaisante :

« ... Les archives sont, au sens de la présente loi, constituées par l'ensemble des documents produits ou reçu par le Parti, l'Etat, les collectivités locales, les personnes physiques ou morales de droit public ou de droit privé, dans l'exercice de leur activité, identifiés par leur intérêt et leur valeur et soit conservés par leur détenteur ou leur propriétaire soit transmis à l'institution d'archives compétents» [jora, 1988b]

## **1.2. Autre sens du mot archive [delmas, 2009]**

### **1.2.1. Service administratif**

Le mot archives désigne aussi des **institutions** ou des **services administratifs** (publics ou privés) qui ont pour mission de conserver les fonds d'archives d'une institution ou d'un groupe d'institutions (publiques ou privées).

Dans ce cas, on ajoute au mot archives l'adjectif qualificatif qui définit le statut du service d'archives, qui identifie habituellement la juridiction ou le type d'organisme de qui dépend le service.

#### **Exemple**

Les expressions : «Archives nationales» pour l'institution qui conserve les archives de l'État, «Archives régionales» pour le service qui conserve les archives des institutions régionales, «Archives municipales» pour le service responsable des archives des municipalités et des communes, «Archives hospitalières» pour celui des archives des hôpitaux etc.

Comme il s'agit de services ou d'institutions, le mot archives est dans le cas de ces expressions employé avec une majuscule.



### 1.2.2. Les bâtiments

Le mot archives désigne aussi les **locaux** ou les **bâtiments** qui abritent les services chargés et organisés pour conserver et gérer les fonds d'archives.

Dans des conditions idéales, on peut ainsi définir le bâtiment d'archives :



« ...il s'agit d'un édifice qui réunit des fonctions nombreuses, complexes... et contradictoires. Il doit assurer la protection des documents, tout en répondant aux besoins des utilisateurs (le personnel des archives, mais aussi les publics). On y distingue une zone de stockage qui, petit à petit, va devenir une zone de conservation (...) et qui est nettement séparée du reste de l'édifice (des zones administratives et de travail ouvertes au public). » [delmas, 2009]





Le plus souvent, il s'agit de locaux situés dans de grands immeubles administratifs. Ces locaux assurent la protection des documents d'archives dans des magasins (installés dans les étages et aussi dans les sous-sols) et l'hébergement du service d'archives et des archivistes en distinguant les espaces de bureaux pour le personnel et la salle de consultation pour les utilisateurs [delmas, 2009]

### **1.3. Archives informatiques**

Les archives électroniques sont des mails, des documents bureautiques, des dossiers numérisés, des données échangées par télé procédures, des bases de données... etc. L'AFNOR définit l'archivage électronique par : «L'ensemble des actions, outils, et méthodes mis en œuvre pour conserver à moyen et long terme des informations dans le but de les exploiter» [abbasi et al, 2009]

Au-delà du stockage, de la sauvegarde et de la gestion électronique des documents, l'archivage électronique peut être défini comme «l'ensemble des actions visant à identifier, recueillir, classer et conserver des informations, en vue de consultation ultérieure, sur un support adapté et sécurisé, pour la durée nécessaire à la satisfaction des obligations légales ou des besoins d'information» [abbasi et al, 2009]

## **2. Les outils de l'archivage**

Ces outils sont divers et permettent aux archivistes de mettre en place des procédures de gestion qui garantissent l'accessibilité et la pérennité des documents : le plan de classement, et le tableau de gestion des archives pour les services producteurs, établis selon la norme ISO 15489 [AAF, 2005b].

-Le plan de classement [Dronne, 2006d] est la colonne vertébrale du tableau de gestion. Il est employé pour référencer et classer les dossiers dans les bureaux. Il permet de faire l'analyse des activités d'un service. Généralement, il est organisé de cette façon, en trois niveaux : fonction, activité et sous-activité. Les activités et sous-activités représentent en fait les missions effectuées par le service. Cette structuration hiérarchique permet d'avoir une vue d'ensemble du fonds. Un plan de classement permet ainsi de ranger, classer et retrouver, aisément, les archives courantes et intermédiaires.

-Vient ensuite, un second outil, le tableau de gestion qui est établi en collaboration avec les services [Agnieszka , Bonard, 2010c] . Il a pour objectif de faciliter le tri des documents. Ce document permet de distinguer les documents à archiver et ceux qui ne le sont pas. Ce dernier permet de classer les documents les uns par rapport aux autres. Ainsi les documents présentant les mêmes caractéristiques et liés à la même activité sont regroupés dans une même classe de documents.

Le tableau de gestion concerne tous les supports : que ce soit papier ou numérique. Un tableau de gestion est établi en accord entre les services producteurs et les Archives départementales. Ce tableau décrit le type de documents produits et fixe pour chacun d'eux la durée d'utilité administrative et le sort final. La durée d'utilité administrative appelée aussi « D.U.A » désigne la durée légale ou pratique pendant laquelle un document est susceptible d'être utilisé par le service producteur ou son successeur, au terme de laquelle est appliquée la décision concernant son traitement final. Le document ne peut être détruit pendant cette période qui constitue sa durée minimale de conservation. Cette durée est généralement exprimée en année. A la fin de cette durée d'utilité administrative, l'archiviste applique ce que l'on appelle le sort final. Nous en distinguons trois : la conservation, l'élimination ou le tri, pour certains types de documents volumineux présentant un intérêt. Ce sort final s'effectue en fonction de l'utilité, de l'intérêt du document. Un tableau de gestion d'archives se présente sous la forme d'un tableau, de quatre colonnes obligatoires ou plus : [Agnieszka , Bonard, 2010c]

- ✓ **Le type ou catégorie de document** : désignation des catégories de documents ou dossiers produits par les services dans l'exercice de leurs missions. La description se limite à l'intitulé et peut être complétée par la composition détaillée du dossier (liste des pièces),
- ✓ **La durée d'utilité administrative (D.U.A)** : durée légale ou pratique pendant laquelle un document est susceptible d'être utilisé par le service producteur ou son successeur, au terme de laquelle est appliquée la décision concernant son traitement final,
- ✓ **Le sort final** : Il s'agit du sort final des archives au terme de la durée d'utilité administrative. Il existe trois traitements finaux des documents : tri, conservation et destruction,
- ✓ **Les observations** : informations complémentaires sur le contexte de production, le service producteur, la D.U.A (explication, justification), le sort final (explication, justification), sur les modalités de tri, etc....

Il est également possible d'ajouter les colonnes suivantes :

- ✓ **Le service** : désignation du ou des services qui détient ou détiennent le type de dossier,
- ✓ **La référence** permettant l'identification de chaque document ou dossier grâce à un numéro unique,
- ✓ **La référence-texte ou texte réglementaire**: Il s'agit des textes réglementaires fixant les règles de gestion des documents (lois, décrets, instructions interministérielles de tri, procédures internes, etc.),
- ✓ **La durée-service** : durée de conservation souhaitée par les services. Notons que cette durée ne peut être inférieure à la D.U.A.

Pour les archives dites intermédiaires, peut être réalisé un plan de récolement ou plus exactement un état des lieux. Cet outil permet de gérer l'arriéré et ainsi de décrire le contenu de chaque unité documentaire qui peut être un dossier, une boîte. Cela aide à l'identification des archives dont la durée d'utilité administrative est en cours, afin de les rendre plus accessibles et de les retrouver le plus facilement possible. Les dossiers seront recensés en suivant l'ordre dans lequel ils sont présumés avoir été rangés. Ce plan de récolement est réalisé, bien souvent, sous la forme d'un tableau, pour une meilleure compréhension.

Il va sans dire que ces outils doivent être mis à jour à chaque changement, et au fur et à mesure de l'évolution des compétences.

- charte d'archivage : Un service d'archives doit également penser à son projet d'archivage, représenté par une charte d'archivage, ce projet d'archivage représente « l'ensemble des études, décisions et actions qui permettent à une entreprise ou un organisme de mettre en place une politique, un référentiel, des procédures et des outils de sorte que tous ses documents engageants soient sous contrôle, c'est-à-dire identifiés, sécurisés, conservés, accessibles, gérés tout au long de leur cycle de vie et détruits ou transférés aux archives historiques à échéance de leur durée de conservation » [Chabin. 2010c]. La charte d'archivage est un document contractuel entre le service archives et le reste des services d'une organisation. Le terme « charte » désigne un document écrit qui formalise les engagements de plusieurs partenaires sur des principes et objectifs précis tendant, par exemple, à améliorer la qualité et l'efficacité du service ou incitant à respecter un socle de règles communes dans l'intérêt de tous. Une charte a vocation à être affichée et connue de tous. Son application fait l'objet d'évaluations régulières.

Une charte d'archivage énonce un ensemble de règles et de bonnes pratiques relatives à l'archivage. Elle est un vecteur incontournable pour formaliser, uniformiser et diffuser les

principes d'archivage au sein d'un organisme. Elle fait l'inventaire des obligations des différents partenaires impliqués dans la chaîne d'archivage. Charte d'archivage s'inscrit dans la tradition de l'archivage papier, et s'organise autour de trois éléments : **[Chabin, 2007b]** .

- ✓ la présentation des référents archives,
- ✓ les procédures d'archivage,
- ✓ le référentiel de classement et d'archivage.

### **3. Processus d'archivage**

Le projet d'archivage s'appuie sur le processus d'archivage, selon les principes du Records Management. Articulé avec les processus métiers, il se décompose en trois sous-processus chronologiques **[AAF, 2005a]**, qui suivent le cycle de vie du document engageant : versement, conservation et destruction, et un sous-processus transverse: la mise à disposition ou l'accès aux utilisateurs. Les outils de gestion d'archives, décrits dans la partie précédente, sont élaborés en fonction de chaque étape du processus d'archivage. Les principales étapes de ce processus sont :

- ✓ **Analyse et classement du document produit ou reçu**

Cette étape d'analyse permet d'identifier les types de documents produits ainsi que les activités des différents services. Le document est alors classé dans une rubrique du plan de classement des activités. Cette opération indique si ce dernier sera ou non enregistré dans le système d'archivage et précise les règles de conservation dans le cas où il est archivé.

- ✓ **Capture et enregistrement du document**

Cette étape montre le rattachement d'un document à un plan de classement. A ce document, sera ajouté des ajouts de description, afin que celui-ci puisse être facilement retrouvé.

- ✓ **Analyse et ajout de métadonnées**

Cette étape a pour but la description complète du document. Celle-ci se fait par l'intégration de métadonnées, qui sont « des données structurées ou semi-structurées qui permettent de qualifier et de gérer les documents archivés tout au long de leur cycle de vie » **[rietsch et al, 2006a]** . Trois types de métadonnées peuvent être distingués :

- les métadonnées descriptives : description du contenu intellectuel (ex. : titre, auteur, date, mots clés...),
- les métadonnées de gestion (ou de structure) : elles aident à organiser, à valider puis à archiver les ressources organisationnelles,

- les métadonnées de préservation (ou administratives) : métadonnées destinées à assurer la conservation à long terme de ressources électroniques. Elles incluent les données techniques telles que la gestion des droits, le contrôle d'accès, les conditions d'utilisation...).

Les métadonnées permettent ainsi de :

- gérer le cycle de vie (savoir combien de temps on doit conserver l'information, à quelles autres informations elle est rattachée, quand on peut la détruire),
- gérer les droits d'accès,
- gérer la recherche,
- gérer l'authenticité du document (valeur de preuve),
- assurer la traçabilité,
- exploiter le document dans son contexte.

✓ **Stockage sécurisé**

La sécurité des documents doit être garantie. Elle est synonyme d'identification, d'intégrité et de confidentialité. cette sécurité doit permettre et garantir [rietsch et al, 2006a] :

- l'identification,
- l'authentification,
- la sauvegarde,
- la lisibilité,
- la traçabilité, qui est «le fait de créer, d'enregistrer et de préserver les données relatives aux mouvements et à l'utilisation des documents».

✓ **Prise en compte des évolutions des documents**

Cela signifie que tous les changements liés au document concernant son statut, sa durée de conservation sont mémorisés.

✓ **Communication, mise à disposition, accès**

Cette étape a pour objectif la traçabilité des actions de communication, de localisation, des utilisateurs et des motifs d'utilisation du document.

✓ **Application du sort final**

Arrivé à la fin de la durée d'utilité administrative, le sort final est appliqué : il est décidé si le document doit être conservé pour être transféré aux archives définitives ou détruit.

#### 4. De bonnes raisons d'archiver

Dans une interrogation sur le pourquoi de l'archivage, sur les principales raisons qui amènent les entreprises à archiver leurs documents.

Tout d'abord, l'archivage permet aux entreprises, de garder une trace des activités, les informations archivées pourront ou pourraient être réutilisées pour expliquer ou justifier quelque chose. L'archivage n'est pas utilisé pour une unique raison, mais pour plusieurs telles que [Chabin, 1999] :

✓ **Faire de la place**

Les archives intermédiaires s'entassent vite dans les armoires de bureaux. On ne sait plus où les mettre. Cela ne fait qu'empirer, le volume de documents ne fait que s'accroître.

L'informatique n'a pas résolu le problème, il a clairement fait empirer les choses.

Désormais, les agents font photocopies sur photocopies, et ils se retrouvent donc avec une quantité impressionnante de documents, parfois en doublons voir en triplets. Il faut cependant se poser une autre question : Est-ce que tous ces documents ont une valeur quelconque pour l'entreprise, resserviront-ils un jour ? Un tri des documents doit être fait, afin d'éliminer ceux qui n'ont pas ou plus d'intérêt, et de conserver les documents qui seront encore consultés. L'archivage permet ainsi aux services de trier et stocker leurs archives dans un endroit sécurisé, dans l'unique but de les retrouver par la suite.

✓ **Pour retrouver**

Cette seconde raison d'archiver résulte de la difficulté d'accéder à l'information. Il va sans dire que pour être retrouvés, les documents stockés doivent être correctement décrits, et les données bien organisées. Les services producteurs doivent mettre en place un procédé de classement rigoureux, s'ils veulent se souvenir de la place de chaque document. Ce discours vaut autant pour les archives numériques que les archives physiques. Si ces documents ne sont pas classés ou inventoriés, ils perdront de leur existence, ils seront pour ainsi dire inconnus.

✓ **Pour sauvegarder et préserver la mémoire de l'entreprise**

En effet, la raison d'être de l'archivage est de conforter la mémoire et surtout de « créer une mémoire indépendante de la mémoire humaine, plus objective, plus large, plus pérenne ». Il est important de capitaliser les documents pour qu'ils soient ensuite ré exploités.

Une fois devenus archives historiques (ou définitives), les documents représentent en quelque sorte l'identité de l'entreprise avec ses activités et son histoire.

L'archivage peut alors être considéré comme le « contrefort du *knowledge management* », discipline dont l'objet est la maîtrise de l'information utile, la gestion des connaissances.

L'archivage et le *knowledge management* sont deux disciplines complémentaires qui s'assemblent pour transmettre aux générations futures, aux nouveaux arrivants dans une entreprise, le témoignage de ce qui a été fait. L'archivage intervient dans l'organisation des informations, qui seront ensuite retransmises par le biais du *knowledge management*.

C'est ici qu'interviennent toutes les raisons d'archiver : pour pouvoir un jour être partagées et transmises, les informations doivent être bien ordonnées, bien décrites pour être « retrouvées ». Toutes ces raisons sont reliées entre elles, car la bonne organisation de la mise en archive résulte de la sélection, la conservation et la mise à disposition de l'information.

## 5. Cycle de vie des archives

A la base, le cycle de vie est un concept assez général. Il s'applique à toutes les professions liées au documentaire. Il correspond à l'âge des archives, qui n'est exprimé ni en mois ni en années mais qui désigne la période de vie qu'elles sont en train de traverser.

Il existe une théorie fondée sur le cycle de vie des archives. Il s'agit de la théorie dite « des trois âges ». Ce concept est né dans la seconde moitié du XX<sup>e</sup> siècle aux Etats-Unis. Il a été utilisé pour la toute première fois en 1948 dans un rapport d'un groupe de travail de la Commission *Hoover* sur l'organisation et le fonctionnement de l'administration fédérale des Etats-Unis [Ducharme, 2011]. Dans les années 50, *Théodore Schellenberg* formule cette théorie des trois âges, en évoquant la distinction entre la valeur primaire (administrative) et la valeur secondaire (témoignage) de tout document. Cette théorie doit son apparition en France à *Yves Pérotin*, en 1961, dans un article où il parle des « archives courantes », des « archives intermédiaires » et des « archives archivées ». Il souhaitait ainsi intéresser les archivistes à la production des archives avant leur versement afin d'en contrôler l'évaluation.

### 5.1. Phase d'activité:

#### • archives courantes ou documents actifs

Période pendant laquelle les dossiers « actifs » sont indispensables au soutien des activités quotidiennes d'une organisation

- Appelés à être utilisés fréquemment, celles-ci doivent demeurer près des espaces administratifs et, par conséquent, être facilement et rapidement accessibles.

Exemple : le dossier d'un membre du personnel est actif tant et aussi longtemps que celui-ci reste à l'emploi de l'organisation. Les informations contenues dans ce dossier s'avèrent indispensables à la gestion courante de l'organisation.

## 5.2. Phase de semi-activité:

### • archives intermédiaires ou documents semi-actifs

Période pendant laquelle les dossiers semi-actifs doivent être conservés pour des raisons administratives ou légales ... mais n'ont pas à être utilisés pour soutenir les activités quotidiennes d'une organisation

- Même s'ils répondent toujours aux objets de leur création, la faible fréquence de leur utilisation ne justifie plus une conservation près des espaces administratifs.

- Ils font alors l'objet d'une gestion centralisée, soit au sein de l'organisation, soit dans des locaux situés à l'extérieur

### Avantages de la gestion des archives intermédiaires

- Réduire d'au moins de moitié les coûts de conservation
- Récupérer des équipements
- Diminuer la masse documentaire totale
- Accélérer le classement et le repérage
- Récupérer des espaces coûteux qui peuvent ainsi être affectés à d'autres fins

Exemple: lorsqu'un employé n'est plus à l'emploi de l'organisation, son dossier devient semi-actif et est conservé pendant toute la période de validité de ses droits.

Son dossier peut être *réactivé* s'il revient travailler pour l'organisation.

## 5.3. Phase d'inactivité:

### • archives définitives ou documents inactifs

Période à compter de laquelle les dossiers «inactifs» n'ont plus de valeur prévisible pour l'organisation qui les a produits.

- N'ayant plus à répondre aux objets de leur création, les dossiers sont ou éliminés ou conservés comme archives définitives.

Exemple: une fois terminée sa période de semi-activité (en règle générale, après le décès de la personne concernée), notre dossier du personnel, ou plutôt quelques-uns des pièces qui le composent, pourraient être conservés comme archives définitives afin de témoigner de son passage dans l'organisation.

La figure ci-dessous décrit le cycle de vies des archives [mouhoubi, 2002a] .



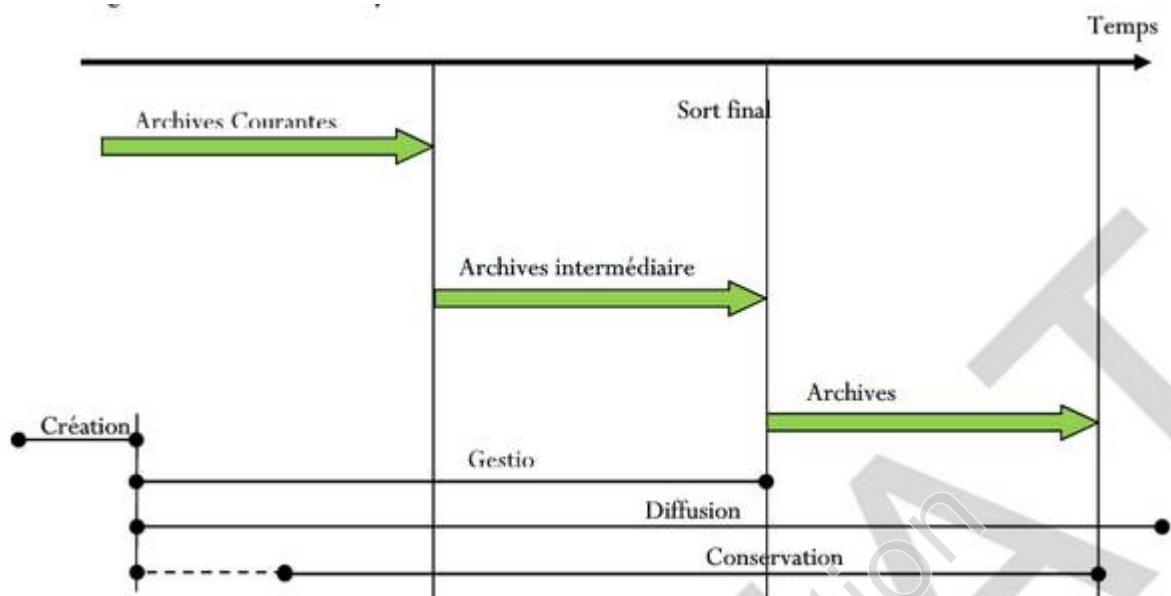


Figure1.1 : la théorie des trois Ages

## 6. Evolution de l'archivage

### 6.1. Multiplication des services d'archives

Depuis les années 1980 [Delsalle, 1998], les sociétés d'archivage se sont multipliées, proposant leurs services de stockage puis de gestion d'archives à des entreprises privées toujours plus nombreuses mais également au service public. C'est ainsi qu'aujourd'hui, dans un grand nombre d'entreprises, la fonction d'archivage est reconnue, cela signifie qu'une personne voir même un service au sein de l'entreprise est chargée de l'archivage, Ce service s'occupe désormais de la réception d'informations de diverses natures provenant de différents services dans le but d'être conservées durant les délais adaptés aux types de documents et à leurs finalités. Ces documents, une fois archivés sont restitués autant que nécessaire aux services producteurs.

Cette multiplication de création de services d'archives est liée au fait que les services producteurs émettent un plus grand besoin d'archiver pour classer et retrouver les documents, dont le volume s'accroît massivement.

### 6.2. Un plus grand besoin d'archiver

L'évolution de l'archivage correspond plus à un besoin apparu en amont de l'activité des entreprises qu'à une volonté de gérer une mémoire après arrêt de l'activité. Trois facteurs peuvent expliquer ce besoin d'archiver [Agnieszka, Bonard, 2010a] : l'inflation du volume

des documents et des données produits, l'arrivée et la progression des hautes technologies et leur application, et enfin l'accroissement de la réglementation caractérisé par la multiplication des normes, et les risques juridiques qui en découlent. L'archivage prend une orientation tout à fait nouvelle, dans la mesure où il devient partie intégrante et surtout active du cycle de vie du document.

Autrefois, nous évoquions la notion de documents vivants ou morts. Un document était considéré comme vivant si il était accessible et encore modifiable alors qu'au contraire, un document était dit mort si il était figé et archivé. Cette tendance a changé, car désormais ce même document archivé a comme particularité de ne plus être modifiable tout en demeurant accessible tant qu'il présente une quelconque utilité.

### **6.3. *Records Management* : modernisation de l'archivage**

De plus, nous pouvons constater que les archivistes s'impliquent plus qu'auparavant, en amont dans la production des dossiers, en relation avec les producteurs de documents afin de faciliter le traitement post-âge courant et de répondre aux demandes des services qui font face à une production croissante de documents papiers et numériques. L'archiviste s'est vu attribuer des missions supplémentaires avec la politique de *Records Management*. Désormais, il doit [Gagnon-Arguin et al. 2003] :

- recevoir les documents et les données devant être conservés au-delà de leur durée d'utilité administrative,
- compléter les descriptions de documents à l'aide de métadonnées pour optimiser la conservation à long terme,
- assurer les éliminations d'une partie de l'information archivée après l'accord des services producteurs,
- aider à la recherche des documents archivés
- assurer la communication des documents dans le respect de la législation et des exigences de l'organisme.

Cette nouvelle implication est due en partie à une modernisation de l'archivage par le *Records Management*. Le *Records Management* est en effet en plein cœur de l'évolution de l'archivage. Antérieurement, l'archivage s'occupait du traitement et classement des documents après leur production, alors que maintenant, le *Records Management* permet d'organiser le cycle de vie des documents en amont de la production. Toutefois, il faut rappeler qu'archivage et *records management* ne recouvrent pas exactement les mêmes éléments, il est vrai que tous

les deux traitent les archives papiers et numériques, mais contrairement à l'archivage, et d'après la norme ISO 15489, «La norme couvre les archives courantes et intermédiaires et ne prend pas en compte les archives historiques. Dans le cycle de vie des archives» [[Dronne, 2006c] . Néanmoins, il convient d'apporter une petite précision : le *records management* ne gère peut être pas les archives historiques mais il assure la gestion des documents appelés à terme à devenir archives définitives.

#### **6.4. L'apparition du numérique**

Nous savons tous que l'archivage connaît aujourd'hui une nouvelle extension avec l'épanouissement de l'internet. « Le numérique perturbe bien des choses dans la notion de mémoire et d'archivage» [Müller, 2011] . Il constitue un véritable défi dans le sens où si l'on veut préserver la mémoire de l'entreprise, il faut que la pérennité des informations soit assurée. Elles doivent pouvoir être lisibles et accessibles que ce soit dans dix ou cinquante ans. Faut-il alors privilégier l'archivage papier ou l'archivage numérique ? Pour en savoir plus et pour répondre le mieux possible à cette question, il nous faut nous pencher sur l'archivage électronique, ses avantages et ses inconvénients.

### **7. Place de l'archivage dans l'entreprise**

L'archivage a évolué et s'est modernisé ; sa place au sein de l'entreprise a changé. La fonction archivage est dès lors perçue différemment, et ces enjeux n'en sont que plus nombreux.

#### **7.1. Enjeux de l'archivage**

« L'archivage relève de la responsabilité de toute personne morale, comme conséquence logique de l'environnement réglementaire, des risques de non- disponibilité de l'information ou de sur-conservation, avec les exigences afférentes d'authenticité, d'intégrité et de fiabilité des documents dans le temps. » [Chabin, 2010b]

En effet, dans les entreprises, les enjeux induits par l'archivage, qu'il soit papier ou électronique, sont multiples et font face à de nombreuses exigences. Ils sont :

- ✓ **juridiques**

Le principal risque est de ne pas pouvoir retrouver les données requises dans un délai imparti lors d'un contrôle de l'état par exemple; non seulement les données doivent avoir été archivées, mais elles doivent présenter des caractéristiques d'authenticité, d'intégrité et de non-répudiation leur permettant d'être retenues comme élément de preuve valide. Les documents, ayant valeur probante, doivent respecter les lois en vigueur portant sur la conservation des différents types de documents.

✓ **réglementaires**

Cet enjeu est de taille, car les entreprises doivent veiller au bon respect des lois en vigueur. Un service archives doit sans cesse se tenir au courant de la réglementation et des exigences en matière de tri, conservation, et communication des archives. Il est, en effet, important de s'y conformer, et de bien connaître l'étendue des sanctions en cas de non respect de ces obligations. Un service archives doit également se rapporter aux normes qui sont établies, et qui définissent les principes et procédures d'archivage à respecter.

✓ **patrimoniaux et historiques**

il est important de conserver les documents, ayant un intérêt certain pour l'entreprise, et pour sa mémoire. Ces documents archivés peuvent faire l'objet d'une réutilisation pour de nouveaux projets. Cela évite également de « recréer l'information, opération qui peut coûter cher et faire perdre à l'entreprise un temps précieux » [rietsch et al, 2006b]

✓ **logistiques**

Les données doivent avoir été bien archivées techniquement, car sans cela il sera pratiquement impossible d'y accéder. Il faut pour cela avoir saisi correctement les métadonnées reliées au document. L'entreprise pourra alors le retrouver aisément, et aura la possibilité d'effectuer des recherches. Pour les documents mal décrits, mal classés, d'autres problèmes se posent : dans le cas d'un logiciel de gestion d'archives, les moteurs de recherche ne produiront aucun résultat à la requête; ou alors, si les données existent, elles ne seront pas intelligibles, car aucun moyen de les décoder et de les interpréter. Cet enjeu consiste donc à faciliter l'accès à l'information.

✓ **sécuritaires**

Si les données ne sont pas ou insuffisamment protégées, elles risquent d'être divulguées. Cela peut être grave si ces données sont confidentielles (données stratégiques, ou personnelles). Cet enjeu est sensiblement lié au précédent. Cet enjeu intervient que ce soit pour les documents numériques ou papiers. Lorsqu'on parle de sécurité, cela évoque le lieu de stockage : un magasin d'archives, un disque dur. Le local de stockage des archives doit être protégé contre l'humidité, les insectes, les incendies, et autres nuisibles tandis que les disques durs doivent protéger leurs archives par le biais d'identifiants, de mots de passe pour éviter les intrusions extérieures.

✓ **technologiques**

L'enjeu technique est double, à la fois dans l'espace, avec les problèmes **d'interopérabilité** entre systèmes et, dans le temps, avec le défi de pérennité des données sur le long terme, face

à l'obsolescence récurrente des formats, supports et outils de restitution. Le système mis en place doit être capable de garantir la lisibilité et l'intelligibilité des données sur le long terme. Toutefois, cet enjeu ne se résume pas qu'à cela, il peut également toucher un aspect plus technique concernant la mémoire du système. Celle-ci devra être en mesure de contenir le volume d'archives stockées. Comme nous avons pu le voir précédemment, cet enjeu est propre à l'archivage numérique.

✓ **financiers**

L'enjeu financier est double également, il peut être symbolisé par des coûts liés au risque encouru dans le cas où l'entreprise se trouve dans l'incapacité de retrouver les documents demandés. Nous retiendrons aussi un second aspect : celui du temps perdu à la recherche d'informations ou de l'investissement perdu dans des logiciels ou supports qui ne se maintiennent pas dans le temps.

Nous pouvons tout simplement résumer cela aux conséquences, pour l'entreprise, de ne pas pouvoir retrouver les informations qu'elle a produites à un moment de son activité, alors qu'elle a besoin de les communiquer ou de les réutiliser.

## **7.2. Perception de l'archivage**

Autrefois, lorsque le terme d'archivage était évoqué, nous avions plusieurs idées reçues, les archives étaient vues comme « un nid à poussière », en résumé un lieu où s'entassait de vieux papiers sans importance. Les archives étaient considérées par une grande majorité comme tout ce qui se trouve derrière la porte portant l'écriteau "Archives", généralement à la cave, sinon au bout d'un couloir, éventuellement au grenier.

Désormais, le regard sur l'archivage a changé, il n'est plus seulement citoyen mais il est aussi culturel. Dans la seconde moitié du XX<sup>ème</sup> siècle, l'idée d'archivage progresse encore avec la notion de patrimoine archivistique mondial. Les entreprises se rendent compte que ne pas archiver, « c'est laisser au hasard le devenir des informations que l'on a créées ». [sellal, 2007]

L'archivage des documents d'une entreprise est une action vitale, il aide à la préservation de la mémoire de l'entreprise. Les nouvelles technologies appellent à repenser l'archivage, c'est ainsi que du support papier englobant les textes, les dessins, etc..., nous sommes passés aux supports numériques.

Nous pouvons dire que l'archivage a un bel avenir devant lui, car malgré les outils perfectionnés mis à la disposition des entreprises, la masse de documents demeure difficile à maîtriser et nécessite une gestion des archives rigoureuse. Les locaux d'archives et serveurs informatiques sont saturés, et la question est : d'où provient cette saturation ? Plusieurs raisons

peuvent être évoquées : la démesure des besoins en mémoire ou information de la société, un tri mal effectué avec pour conséquence la redondance des documents, la conservation de documents n'ayant plus d'intérêt pour l'entreprise. Les entreprises se sont aperçues, au fil des années, que l'archivage présentait un triple intérêt :

- La gestion courante du service : disposer en permanence des informations utiles à son activité, aide à la prise de décision, assurer la continuité de l'action administrative,
- La justification des droits et des obligations des personnes physiques ou morales, publiques ou privées : conserver les preuves en cas de contestation,
- La sauvegarde de la mémoire : constituer les matériaux de l'histoire.

Il ne faut cependant pas tout généraliser, la perception de l'archivage a peut être changé et évolué, mais il demeure tout de même quelques idées reçues telles que : les archives ne nous concernent pas, les archives sont de vieux papiers ou encore les archives ne permettent que de faire le vide dans un bureau.

En effet, d'un côté, des organisations voient encore en l'archivage, une tâche longue et contraignante. Ils ne se sentent pas concernés par l'archivage, qu'ils considèrent comme une perte de temps. La gestion des archives ne fait en effet pas parti de leur priorité, ils ne prêtent guère attention aux pratiques d'archivage mises en place. Ces organisations ont bien souvent le sentiment de ne pas avoir d'archives. Or, il faut rappeler que l'ensemble des documents produits dans une administration sont des archives, quelque soit leur sort final. Ces organisations ont bien souvent des difficultés à estimer l'intérêt, la valeur historique de leurs archives, de ce fait, ils gardent généralement la totalité de leurs documents, ou ils éliminent des documents ayant de l'intérêt. Ils éliminent progressivement leurs documents car ils les considèrent comme inutiles ou sans intérêts, ou juste pour faire de la place dans leur bureau.

Il faut donc se dire, que malgré ces quelques préjugés, l'archivage occupe désormais un rôle prépondérant au sein de l'entreprise. Ce service doit avoir une place centrale car sans l'archivage, les informations risqueraient de disparaître, de se disperser et d'être inutilisables.

## **8. Archivage papier vs archivage électronique**

L'enjeu de l'archivage électronique est de traiter les archives électroniques selon les mêmes règles que les archives papier. Les principes restent identiques à ceux des archives papier, mais les moyens sont plus lourds et les risques plus grands pour assurer l'intégrité et la sécurité. Il existe encore entre les deux archivages des différences flagrantes dues en partie aux technologies informatiques.

### **8.1. Une valeur probante similaire**

La première similitude entre papier et numérique émane du cadre juridique, qui stipule que désormais numérique et papier auront la même valeur, le même intérêt. En effet, d'après la loi du 20 juin 2005 [Cod civil, 2007], « L'écrit sous forme électronique est admis en preuve au même titre que l'écrit sur support papier, sous réserve que puisse être dûment identifiée la personne dont il émane et qu'il soit établi et conservé dans des conditions de nature à en garantir l'intégrité ».

Les originaux numériques doivent donc être conservés. Désormais, voilà maintenant plus d'une décennie, l'informatique est utilisée pour produire des documents de façon autonome sans aucune intervention du papier.

Cependant, pour être reçu à titre de preuve, le document numérique doit être authentique, c'est à dire qu'il doit prouver qu'il est bien ce qu'il prétend être. Le seul moyen de garantir cette authenticité est la conservation et la traçabilité du document tout au long de son cycle de vie.

Le document électronique est identifié et authentifié par le biais de ce que l'on appelle « la signature électronique », dont la définition nous est donnée dans [Cod civil, 2007]. Une signature électronique est « la signature nécessaire à la perfection d'un acte juridique identifie celui qui l'appose. Elle manifeste le consentement des parties aux obligations qui découlent de cet acte. Quand elle est apposée par un officier public, elle confère l'authenticité à l'acte». Elle fait donc figure de procédé fiable d'identification garantissant la fiabilité et l'authenticité du document auquel elle s'attache. Cette signature permet, entre autres, de répondre aux exigences suivantes :

- ❖ être liée à son destinataire,
- ❖ identifier son signataire,
- ❖ être créée sous contrôle exclusif de son signataire,
- ❖ être liée aux données auxquelles elle se rapporte de telle sorte que toute modification ultérieure des données soit détectable.

### **8.2. Des règles de gestion identiques**

Nous pouvons également associer aux archivages papiers et numériques d'autres concordances, qui sont cette fois liées aux règles de gestion des archives. Ces similitudes sont dues, en partie, au fait que l'archivage électronique intègre, lui aussi, les règles de gestion du *records management*. L'archivage électronique traite un document dès sa validation, et respecte ainsi tout comme pour l'archivage papier, le cycle de vie des archives. Les caractéristiques nécessaires pour la maîtrise de l'archivage électronique sont décrites d'après

le manuel Moreq22 [Chabin, 2008b], qui reprend certains concepts fondamentaux en accord avec les principes du *Records Management* :

- ❖ D'une part, les archives sur support électronique ont les mêmes durées d'utilité administrative que celles sur support papier et sont ainsi soumises aux mêmes modalités de tri et d'élimination que celles-ci,
- ❖ D'autre part, les archives électroniques, pouvant être détruites, font l'objet de la même procédure d'élimination qui consiste, avant destruction matérielle, la rédaction d'un bordereau nécessitant impérativement l'approbation des. En cas de données confidentielles, la destruction doit être sécurisée pour éviter toute récupération de l'information,
- ❖ Comme pour les archives papier, les archives électroniques à conserver sont décrites et analysées dans des inventaires ou répertoires avec des données de description traditionnelle (objet, action, dates), de provenance (service producteur, service versant), de gestion (nombre de boîtes, sort final), de référencement et de localisation (cote, adresse de rangement), de communication (délais de communicabilité). Toutefois, au-delà de ces données, les archives électroniques seront complétées par des données d'intelligibilité (format, données techniques).

L'archivage électronique intègre donc, tout comme le papier, les notions de plan de classement, délais de conservation, sort final des documents, identification, recherche, et restitution. Il reprend en quelque sorte les étapes du processus du *Records Management*.

### 8.3. De profondes différences

Il ne faut cependant pas oublier d'évoquer les différences qui perdurent entre papier et numérique. [Chabin, 2007a]

il ya quatre caractéristiques fondamentales, qui différencient le papier et le numérique :

#### ❖ L'unité de mesure

Premièrement, papier et numérique n'ont pas la même unité de mesure. En effet, le papier se compte en pages, en mètres linéaires de rayonnages, tandis que le document numérique se compte en octets, qui est l'unité de base pour le codage d'un caractère d'écriture, à l'aide de valeurs binaires représentées par 0 ou 1.

#### ❖ La sensorialité

Deuxièmement, les documents numériques sont caractérisés par leur immatérialité.

Contrairement au papier, ils ne réunissent aucun des cinq sens. Ce ne sont pas des objets que l'on peut toucher ou encore moins feuilleter, ce sont des objets qui ne sont ni concrets ni tactiles. Les archives numériques n'ont également aucune odeur, aucun goût et n'émettent



aucun bruit « comparable à la page que l'on froisse ». Ce sont des objets virtuels, dépendant d'un système informatique.

#### ❖ **L'accessibilité**

Un problème subsiste avec le papier : l'éloignement des archives, qui se trouve parfois dans un lieu reculé et difficilement accessible. Pour consulter au plus rapide les archives papiers, il fallait penser à mettre en place un magasin archives se trouvant au plus près des services producteurs. Avec les archives numériques, les distances sont rompues. Les services peuvent ainsi accéder rapidement aux informations, par le biais d'une simple base de données, ou par l'accès à un logiciel de gestion électronique de documents.

#### ❖ **L'espérance de vie**

La différence fondamentale existante entre archivage papier et électronique est, sans nul doute, le problème des contraintes technologiques fortes. Il se pose toujours le souci de la pérennité, qui consiste à conserver le document et l'information qu'il contient dans son aspect physique comme dans son aspect intellectuel, sur le très long terme (30 ans et plus) et de manière à pouvoir le rendre accessible et compréhensible.

Dans le monde du papier, le support et le contenu sont indissociables, mais ce n'est évidemment plus vrai pour le document numérique. Sa pérennisation doit être envisagée d'une toute autre façon. Or, la plupart des fichiers informatiques de plus de 20 ans sont aujourd'hui illisibles, ce qui a pour conséquence plusieurs risques et contraintes inéluctables. L'illisibilité des documents est liée à plusieurs problèmes techniques tels que :

- l'obsolescence des formats d'encodage, à laquelle il faut ajouter la question de la disponibilité des outils permettant d'interpréter ces formats. Ce problème vient du fait que la plupart des protocoles et composants logiciels intervenant entre le support et la restitution des données sont en constante évolution. Ce sont : les normes d'encodage, les formats de fichiers, et les logiciels. L'évolution est rapide et n'offre pas toujours de compatibilité, c'est particulièrement vrai au-delà de quelques années. A ce jour, il n'existe pas de méthode simple garantissant l'accès aux documents électroniques sur le long terme.

La seule solution qui paraît la plus pertinente est de ne stocker l'information que dans des formats largement répandus, qui doivent être universels, ouverts c'est-à-dire complètement documentés et dont les spécifications sont publiées, non propriétaires (libre de droits), indépendants et dont la durée de vie est réputée longue.

- **la fragilité des supports, et leur durée de vie limitée**

Les principaux risques sont : la dégradation des supports, l'évolution trop rapide des environnements (matériels et logiciels), et l'obsolescence des systèmes. Tel ou tel support fonctionne sur un matériel ou système bien précis, si celui-ci a changé, le support ne sera plus forcément reconnu et lisible. Seconde difficulté : la durée de vie du support qui n'est pas toujours très longue. Prenons, par exemple, le cas d'un DVD, qui est un support optique.

Les supports magnétiques englobent les bandes magnétiques, les disques durs magnétiques.

Les supports optiques, quant à eux, sont les CD, les DVD. Les Mini Disc sont des supports magnéto-optique et les semi-conducteurs représentent la mémoire flash. Les supports du type disque dur ou disquette ne sont pas du tout adaptés au stockage moyen ou long terme. Les cartouches et les bandes magnétiques se maintiennent entre 15 et 30 ans. Certains supports optiques comme les CD-ROM, CD-WORM, DVD, DON WORM qui annoncent une durée intrinsèque de 20 à 50, voir 100 ans pour le Century disc en verre trempé. Ces supports optiques, appelés aussi WORM physique, sont recommandés par la norme NF Z 42-01326 [Dronne, 2006b].

Ces supports étant non-réinscriptibles, permettent la conservation du document sans modification. Leur durée de vie, malgré le fait qu'elle soit bonne, doit tenir compte de la durabilité plus faible des formats et des lecteurs. Dans le monde des archives physiques, les supports ont en général une durée de vie de plusieurs décennies voire plusieurs siècles.

Effectivement, la migration et l'émulation sont considérées comme les solutions les plus fiables. La mutation consiste, d'après la norme ISO 15489 [Dronne, 2006b] portant sur le *records management*, « à transférer des documents d'un système à un autre en préservant leurs caractéristiques, c'est à dire leur authenticité, intégrité, fiabilité et exploitabilité ». Cette migration peut être logique, correspondant à un changement de formats, ou physique pour le changement de supports.

D'après le manuel de spécifications Moreq2 [Chabin, 2008b], l'émulation représente « le transfert de l'information vers un nouveau matériel avec ajout d'un composant logiciel qui émule l'ancien matériel, permettant ainsi l'exécution de l'ancienne application ».

- **l'obsolescence des environnements matériels et logiciels dans lesquels**

L'information a été créée ou stockée. Les deux premiers problèmes découlent tout simplement de celui-ci. Les logiciels se développent très rapidement, et font ainsi l'objet de plusieurs

versions, qui posent le problème de reconnaissance des formats, d'où le fait que l'information ne peut être lue dans sa forme d'origine.

#### ❖ **La mise en archive**

Les documents papier sont normalement regroupés en dossiers dans des chemises cartonnées.

Les dossiers papier sont rattachés à une structure dite plan de classement.

Avec l'apparition de l'archivage électronique, la mise en archive peut être entièrement automatisée. Avec les nouvelles pratiques du *Records management*, qui prennent en compte l'archivage électronique, la désignation du document archivé intervient dès la production ou conception du document, et non à posteriori comme cela se pratique avec l'archivage simple.

Des logiciels de gestion d'archives physiques et électroniques ont été élaborés pour cela, suivre le document tout au long de son cycle de vie, de sa conception à l'application du sort final. Dans un logiciel de ce type, les documents électroniques sont gérés dans des dossiers et des répertoires électroniques; ils sont virtuels au sens où ces dossiers et répertoires ne « contiennent » rien ; en réalité ce sont les métadonnées attachées aux documents archivés qui créent les dossiers.

Nous pouvons donc en conclure que les archives électroniques sont de plus en plus nombreuses, et ne font que croître dans les entreprises. Les principes de l'archivage électronique restent les mêmes que ceux de l'archivage papier, mais ils divergent par le fait que pour l'archivage électronique, les moyens qui demeurent plus lourds, les risques plus grands pour assurer la pérennité, l'intégrité des documents.

# Chapitre II

## Archivage électronique des documents

### état de l'art

#### 1. Document électronique

##### 1.1. Définition du document

Plusieurs définitions sont proposées dans la littérature. Nous avons retenu les suivantes :

L'ISO fournit la définition du document : « l'ensemble d'un support d'information et des données enregistrées sur celui-ci sous une forme en général permanente et lisible par l'homme ou par une machine » [Aliouali, 2006b]

C'est un moyen de communication de la pensée, de la connaissance et du savoir.

##### 1.2. Définition du document électronique

« Un document électronique est un document existant sous format numérique de manière à être accessible par des outils informatiques de traitement de données ».

##### 1.3. Traitement des documents numériques (Dématérialisation)

###### 1.3.1. Numérisation des documents

Numériser un document c'est le traduire en bits ou plutôt transformer ses informations en chiffres binaires [ait ali yahia, 2008]. En d'autres termes, le convertir d'une information analogique en une information digitale exploitable par l'informatique. Ce procédé de numérisation peut s'effectuer par diverses technologies qui diffèrent selon l'utilisation et les buts prévus pour les documents numérisés et il peut s'appliquer à tout type d'information : texte, son, image fixe et image animée. La numérisation peut être réalisée par

- Traitement de texte : saisie manuelle des données ; opération coûteuse, et n'est utilisée généralement que par des sociétés privées ou des universités ;
- Un appareil photo numérique, un numériseur (scanner) à plat ou un numériseur à diapositives : ces appareils offrent une bonne qualité, mais la manipulation n'est pas aisée pour les documents reliés ;
- Un scanner de livres qui est bien adapté aux documents reliés.

###### 1.3.2. Reconnaissance Optique de Caractères (OCR)

La reconnaissance optique de caractères est une opération qui consiste à transformer le document résultant de la numérisation en format image en format texte, et donc à le rendre lisible à la machine. Cette opération est utile pour l'indexation automatique et par conséquent

pour la recherche d'information, mais elle n'est pas toujours conseillée, surtout pour les documents très anciens. Une fois faite, cette opération permet de fournir un document reflétant exactement le document original. Aujourd'hui, il existe plusieurs produits très performants qui permettent d'effectuer cette reconnaissance optique, tout en gardant la mise en page, la taille, les caractères... [ouarzeddini, 2008a].

### 1.3.3. Détermination de la structure d'un document

Une structure d'un document bien déterminée facilite beaucoup la tâche du lecteur, il peut ainsi accéder directement à la page souhaitée tout en ayant la possibilité de déceler tous les éléments d'identification du document (titre, auteur, paragraphes). Dans la détermination de la structure, il faut donc penser à bien présenter les éléments d'identification et le contenu des documents, aux liens hypertextes qui doivent être retrouvés facilement et donc, différents du texte normal, et au mode de consultation, c'est à dire linéaire ou hiérarchique...

« Il est bien évident que la consultation restera plus aisée avec une structure hiérarchique » [ait ali yahia, 2008] .

Il existe deux catégories de structure :

- Structure logique et structure physique ;
- Structure générique et structure individuelle.

Pour la première catégorie, la structure logique est mise en place par tout ce qui tente d'organiser le contenu intellectuel : chapitres, sections..., mais ces différents éléments ne peuvent pas se mettre en évidence que par une "apparence visuelle" : structure physique sans qu'il y ait vraiment une concordance logique entre les deux structures : « disposer le texte sur deux colonnes plutôt que sur une, le composer en drapeau plutôt que justifié, peut ne répondre à aucun impératif logique »

Pour la deuxième catégorie, elle peut être expliquée par le fait que la structure logique d'un document peut être à la fois individuelle (dans deux articles, on ne trouve pas le même nombre de chapitres, ni le même nombre de sous chapitres dans chaque chapitre...) et générique, si on essaie de définir une structure plus générale pour n'importe quel article scientifique, en disant par exemple qu'un article contient toujours dans l'ordre au moins une introduction, une partie (dont il y a au moins un chapitre, lequel possède au moins un paragraphe) et une conclusion.

### 1.3.4. Formats de représentation des documents

Il existe plusieurs formats permettant de stocker un document numérique ; la liste ci-dessous ne prétend pas être exhaustive mais vise à dresser un panorama des formats les plus fréquemment utilisés [ouarzeddini, 2008a]:

#### a. Formats textuels

- **TXT (*text*)**

C'est le format ASCII (*American Standard Code for Information Interchange*) :

Il comporte l'alphabet latin non étendu, sans aucune convention typographique (gras, italique, souligné), ni aucune mise en page (tabulations, tableaux, paragraphes, justification, centrage, etc.). Il peut être créé simplement, et constitue l'ensemble des formats des documents en mode caractère (Word Perfect, SGML, HTML, ...).

- **RTF (*Rich Text Format*)**

RTF est un format propriétaire de *Microsoft* devenu standard de fait. Il peut être considéré comme un véritable format d'échange de texte enrichi puisque la présentation et les conventions typographiques sont conservées.

- **PDF (*Portable Document Format*)**

PDF est un format propriétaire développé par Adobe, il n'est pas modifiable, mais permet d'être visualisé et imprimé de manière très simple avec un *Reader* qui est téléchargeable gratuitement d'Internet. Le format PDF offre une meilleure qualité d'affichage et assure la recherche en texte intégral dans des pages digitalisées, il respecte la mise en forme des documents originaux.

- **SGML (*Standard Generalized Markup Language*) . [ait ali yahia, 2008]**

C'est un langage de description de documents autorisant une structuration de l'information à l'aide de balises. Une balise désigne une marque particulière ajoutée à un texte afin de déduire sa structure ou le format dans lequel il sera édité. SGML a été adopté officiellement en octobre

1986 comme norme internationale.

La structure globale d'un document SGML est la suivante :

- un ensemble de déclarations où sont précisées les caractéristiques SGML utilisées telles que la version de la norme, le jeu de caractères utilisés, etc.
- une Définition de Type de Document (DTD) qui décrit la structure logique et exprime la manière dont les différents éléments d'information sont organisés et articulés entre eux,

- un contenu.
- **HTML (HyperText Mark-Up Language)**

C'est un langage dit de « marquage » (de « structuration » ou de « balisage ») dont le rôle est de formaliser l'écriture d'un document avec des balises de formatage. Les balises permettent d'indiquer la façon dont doit être présenté le document et les liens qu'il établit avec d'autres documents.

Le langage HTML permet notamment la lecture de documents sur Internet à partir de machines différentes, grâce au protocole HTTP, permettant d'accéder via le réseau à des documents repérés par une adresse unique, appelée URL.

- **XML : (eXtensible Markup Language). [ouarzeddini, 2008a]**

C'est en quelque sorte un langage HTML amélioré permettant de définir de nouvelles balises. Il s'agit effectivement d'un langage permettant de mettre en forme des documents grâce à des balises (markup). Contrairement à HTML, qui est à considérer comme un langage défini et figé (avec un nombre de balises limité), XML peut être considéré comme un métalangage permettant de définir d'autres langages, c'est-à-dire définir de nouvelles balises permettant de décrire la présentation d'un texte.

La force de XML réside dans sa capacité à pouvoir décrire n'importe quel domaine de données grâce à son extensibilité. Il va permettre de structurer, poser le vocabulaire et la syntaxe des données qu'il va contenir.

En réalité les balises XML décrivent le contenu plutôt que la présentation (contrairement à HTML). Ainsi, XML permet de séparer le contenu de la présentation ce qui permet par exemple d'afficher un même document sur des applications ou des périphériques différents sans pour autant nécessiter de créer autant de versions du document que l'on nécessite de représentations.

## **b. Formats graphiques**

- **BMP (Microsoft Windows Bitmap)**

C'est le format le plus simple développé par Microsoft et IBM, c'est pourquoi il est particulièrement répandu sur les plateformes Windows et OS/2. C'est un fichier d'images graphiques. C'est un format qui est commode à utiliser pour des fins, bureautiques, mais ce n'est pas un format de stockage.

- **TIFF (*Tagged Image File Format*)**

Conçu par Aldus et Microsoft, Ce format est devenu le format standard pour tout ce qui concerne les images en noir et blanc acquises numériquement par scanner, Il gère toutes les nuances de couleurs et intègre des informations de correction gamma. Les fichiers TIFF peuvent être lus avec n'importe quel logiciel de traitement d'image.

- **JPEG (*Joint Picture Expert Group*)**

Ce format est supporté sur le web. Le format JPEG permet de diminuer considérablement la taille des documents. Il est utilisé pour la compression des photos et permet de choisir le taux de compression.

- **GIF (*graphique interchange format*)**

Ce format est devenu le format de référence de l'image sur les grands serveurs d'information en réseau. Il est assez simple et donc très facilement lisible par les logiciels, d'où son succès sur le web. Le format GIF diminue sensiblement la taille du document de (10 à 30%), il est utilisé pour la compression des dessins.

### **1.3.5 Indexation**

L'indexation des documents numériques est soumise au même principe que celle des documents traditionnels qui consiste à extraire les concepts les plus représentatifs du document et de les représenter sous forme d'un index ou d'un modèle appelé métadonnées dans les systèmes documentaires [ouarzeddini, 2008a]. Pour les documents numériques, cette opération consiste donc à «fournir des informations utiles sur chaque document et ce dans le but d'en faciliter l'accès par les moteurs de recherche[ait ali yahia, 2008] .Ces informations s'appellent des métadonnées qui ont pour rôle de décrire formellement le document, et à partir desquelles, les moteurs de recherche vont effectuer leurs requêtes. Ces métadonnées comportent en général le titre de document, le nom de l'auteur, le type de document (revue, livre...), une liste de mots clés et le résumé du document.

### **1.3.6. Recherche d'information**

Comme dans tout système de recherche d'information, la recherche est l'opération la plus importante qui consiste à mettre en interaction l'utilisateur qui exprime son besoin à travers une requête et le système qui traduit ce besoin en un langage système, calcule la correspondance entre cette requête et les index des documents, puis fournit à l'utilisateur une liste des documents qu'il juge pertinents [dahak, 2006a]



## 2. La gestion électronique des documents (GED)

### 2.1. Définition

Elle est définie comme « la gestion, par des moyens informatiques, du cycle de vie d'un document électronique (textuel, graphique, sonore, vidéo, ...), allant de sa création à sa destruction, en passant par sa modification, sa publication, sa diffusion, ... de manière à optimiser l'accès à ce document, à l'information qu'il contient et à l'information qu'il concerne . [ponsot, 1995] .

### 2.2. Avantages de la Gestion Electronique de Documents

Au-delà de la simple conservation de la copie des documents sous forme d'archives électronique, les solutions de GED sont conçues pour permettre le partage d'informations en facilitant la recherche et la consultation des documents sur un réseau d'entreprise ou un intranet. Il s'agit de mutualiser les documents et de permettre une consultation rapide en accélérant les temps de recherche. Une grande partie de l'utilisation des dossiers en carton, souvent constitués de plusieurs documents d'origines différentes, peut se résumer dans son équivalent logiciel à une simple recherche à partir d'un PC; le dossier étant visible sur écran de manière instantanée. Les logiciels GED répondent ainsi aux besoins de conservation, de partage et de recherche en facilitant le rapprochement des documents connexes générés ou reçus par les organisations. Et, par la magie de l'informatique, la recherche de documents devient quasi immédiate avec tout le bénéfice d'une réactivité sans faille. Pouvoir réunir, à la demande un dossier complet en rapport avec une problématique particulière (suivi de commande, SAV, litige, relance etc.) est un avantage stratégique.

Un logiciel de gestion électronique de documents doit permettre de faciliter les tâches courantes au travers de fonctionnalités que l'on souhaitera ergonomiques :[Livre blanc, 2011]

- Collecter des documents par des moyens simples (moteur de recherche, sélection interactive des résultats) ;
- Offrir une visibilité d'ensemble sur les documents collectés ;
- Respecter les règles de confidentialité ;
- Mémoriser les requêtes associées aux recherches routinières ;
- Faciliter et accélérer l'indexation manuelle par un système de vidéo-codage qui offre une bonne visibilité ;
- Permettre de transmettre des notes et des informations en rapport avec les documents ;
- Suivre les actions et les décisions liées à un type de documents avec un *workflow* intégré ;

- Alerter les utilisateurs quand un nouveau document doit être pris en compte (indexation/classement manuel, décision dans le *workflow*, etc.) ;

### **2.3. Les différents types de GED**

Il existe différents type de GED, parmi les quelles citons [ait ali yahia, 2008] :

#### **2.3.1. La GED technique**

Appelée encore GED métier concerne toutes les applications directement liées à une profession. Elle se caractérise par la manipulation de documents dont le format est propre à un métier. Exemples (les plans, la documentation technique).

#### **2.3.2. La GED documentaire**

Elle est directement issue de la recherche documentaire particulièrement utilisée dans des applications de type bibliothèque, documentation scientifique et éventuellement revue de presse.

#### **2.3.3. La GED COLD**

Le terme COLD (*Computer Output on Laser Disk*) est utilisé pour l'archivage électronique des données informatiques. C'est la substitution des traditionnelles microfiches COM (*Computer Output Microfilm*) ou le listing.

Cette solution GED regroupe des programmes et des applications conçus pour gérer des fichiers spools (fichiers générés par des ordinateurs centraux ou des systèmes transactionnels : facture, bon de livraison, états comptables, relevés de compte, fiches de paie, etc.).

#### **2.3.4. La GED administrative**

Consiste à gérer les documents bureautiques d'une administration dans leur format d'origine (Word, Excel, etc.), de centraliser leur classement sur un ou plusieurs serveurs, d'échanger ces documents ou de les distribuer aux autres collaborateurs.

### **3. De la GED au *Records Management***

#### **3.1. Qu'est ce que le *record management***

Champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liées à la forme des documents. [Mouette, 2007b]

#### **3.2. Document management/GED et Records management/Archivage**

Les besoins de GED ne correspondent pas à des besoins de conservation. La GED sert à l'exploitation des données tandis que les documents sont conservés sur support papier.

En revanche, de nombreuses fonctionnalités des GED peuvent recouvrir celles des systèmes d'archivage électronique. « un système de GED, typiquement, inclut l'indexation des documents, la gestion du stockage, le contrôle des versions, l'intégration des applications bureautiques et des outils de recherche pour accéder aux documents. Certains systèmes d'archivage électronique fournissent des fonctions de GED complètes, d'autres une partie seulement. Inversement, certains systèmes de GED ont incorporé des fonctions caractéristiques du *records management* et de l'archivage ». [Chabin, 2006]

Pour clarifier les choses, le tableau ci-dessous met en évidence les différences essentielles entre les deux systèmes : [Chabin, 2006]

Un système de GED	Un SAE
Permet la modification des documents et la production de plusieurs versions	Interdit la modification des documents
Peut permettre la destruction des documents par leurs auteurs	Interdit la destruction de documents en dehors d'un contrôle strict
Peut comporter la gestion de durées de conservation	Comprend obligatoirement un contrôle rigoureux des durées de conservation
Peut comprendre une structure organisée de stockage sous le contrôle des utilisateurs	Comprend obligatoirement une structure rigoureuse de classement, gérer et contrôler par l'administrateur
Est a priori dédié à la gestion quotidienne des documents pour la conduite des affaires	Peut faciliter les tâches quotidiennes mais est aussi destiné à la constitution d'un fond sécurisé des documents probants de l'entreprise

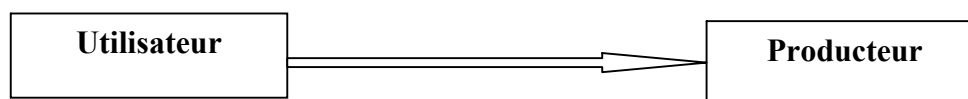


Tableau 2.1 : Document management/GED et Records management/Archivage

## **4. Archivage électronique des documents à valeur probante**

### **4.1. Les enjeux de la conservation sur le très long terme : l'obsolescence technologique**

#### **4.1.1. Le vieillissement du support [aliouali, 2006a]**

La problématique de la conservation des supports des documents numériques tourne autour de leur fragilité et de leur durée de vie; C'est une question qui est très ancienne et classique au niveau des bibliothèques mais d'actualité quand il s'agit de la conservation des sites web.

Trois types de support de stockage sont répertoriés : Les magnétiques (disquette), les optiques (cd-rom) et les magnéto-optiques (minidisque). Ce sont des supports d'une capacité de stockage différente allant de 1,2 Go pour les disquettes et de 4,7 Go pour les DVD et d'une durée de vie qui va de 75 à 200 ans pour un CD. A savoir qu'un CD gravé est plus fragile qu'un CD pressé. Certains laboratoires de recherche annonce une durée de vie de 5 ans pour les CD enregistrables contre 10 à 25 ans pour les CD pressés mais aucune données n'est stable jusqu'ici. Et comme tous types de support les conditions de leur stockage sont très importantes et selon la Digitale préservation Coalition la température de l'environnement de conservation des supports est très importante pour assurer une longue longévité.

#### **4.1.2. L'obsolescence des techniques**

Un document numérique est constitué d'un ensemble de couches, comme on l'a vu précédemment, et le traitement informatique de chaque couche nécessite un ensemble d'outils dont: les systèmes d'exploitation, les logiciels, les formats de données...etc. Donc l'utilisateur est finalement dépendant d'une certaine technologie qui est en fait un problème majeur dans tout processus de préservation des documents numérique. La durée de validité des programmes et périphériques est estimée entre 2 à 5 ans [aliouali, 2006a] outre cette période les données ne seront plus accessibles. Mais cela n'est que l'aspect matériel du problème. Un autre problème se pose aussi et qui en fait n'est que la partie logicielle et application qui constitue la partie la plus abstraite pour la lecture et la manipulation des documents numériques. Ces derniers se développent très rapidement d'une version à l'autre, posant en cour de route un autre problème qui est la compatibilité entre les versions au niveau ascendant. Le cas de Microsoft Word qui apparaît tous les 3 à 5 ans or que chaque version n'est compatible qu'avec la toute précédente. Ajouter à cela le problème de formats de données qui peuvent être à la fois :

- Propriétaire mais les caractéristiques sont connus et diffusés
- Propriétaire mais les caractéristiques sont tenue secrètes
- Formats standard produit par des organismes de normalisation tel que L'AFNOR et l'ISO

Donc, ce qui est ouvert et connu ne peut le rester dans quelques années, et pour pouvoir tracer une politique de conservation et de préservation, il faut prendre en considération tous les aspects techniques liés à l'environnement du document numérique

Il importe avant tout d'identifier les multiples aspects de l'objet que l'on veut conserver. Ensuite, pour réussir à le conserver, il faut prendre des mesures qui assurent la conservation à long terme d'autant d'aspects de ce dernier que possible. Donc pour conserver un document numérique, il faut [aliouali, 2006a]:

- **Fixer l'objet :**

L'archivage de chaque changement apporté sur une notice ou de fixer périodiquement un état de la base.

- **Conserver le contenu :**

La préservation à long terme des objets numériques vise essentiellement à préserver la valeur intellectuelle de l'information qu'ils contiennent. Ce qui n'est pas toujours évident, car définir ce qu'est le contenu intellectuel d'un document numérique est très complexe, et l'analyse des contenus peut se faire sur plusieurs niveaux d'abstraction pour définir le contenu à préserver. Ces niveaux correspondent aux différentes couches dont le document numérique est structuré. Donc parmi les exigences de la conservation, assurer l'accès au niveau le plus bas du document qui correspond à cette succession de bits allant de 0 à 1 tel un texte ASCII, préserver son contenu sans les variantes de polices pour l'embellir et sans les caractéristiques de la mise en page.

- **La référence :**

Pour qu'un objet d'information numérique soit non seulement préservé, mais aussi accessible, il faut pouvoir y faire référence dans des outils de recherche. La référence doit être un lien actif qui permette, par son activation d'avoir directement accès à la ressource, ou au minimum à des explications sur les conditions d'y accéder. Il est indispensable que ce lien reste persistant pour identifier la localisation

- **Conserver la source :**

C'est un des aspects les plus importants dans une chaîne de traitement pour la préservation. Elle consiste en l'enregistrement de l'origine de l'objet ainsi que tout l'historique de sa conservation et des environnements techniques qui ont accompagné sa vie. Le fait de conserver l'historique de la chaîne de conservation de la ressource depuis sa création sert la préservation de deux manières. Il permet :

- de s'assurer de l'authenticité de l'objet ;
- de connaître le contexte qui est un autre point important de la garantie de l'intégrité.

- **Conserver le contexte :**

Les objets numériques sont partiellement définis par le matériel et les logiciels sur lesquels leur diffusion repose, par leur mode de distribution et les liens qui les unissent à d'autres objets numériques. Conserver des objets numériques peut vouloir dire les affranchir d'une dépendance technique, changer leur mode de distribution et désactiver leurs liens. Dans ces circonstances, la conservation du contexte pose un défi exceptionnel.

- **Conserver la présentation :**

Le contenu est habituellement organisé selon une présentation, un format ou une mise en page, qui comprennent diverses polices et de tailles différentes, des colonnes, des marges, des en-têtes et des bas de page, de la pagination et ainsi de suite. Pour ce qui est du genre de document numérique, il se présente sous différents formats de données (SGML, XML et certaines formes de format PDF), les spécifications de mise en page sont séparées du contenu. Pour garder l'apparence d'origine d'un document, on doit aussi conserver ces spécifications de mise en page, particulièrement lorsque celles-ci contribuent de façon inhérente à la compréhension et à l'interprétation du contenu.

- **Conserver la fonctionnalité :**

Les objets numériques comportent une fonctionnalité qui dépasse de loin celle du document classique sur papier. Ces objets peuvent contenir des composantes multimédias (du texte, des graphiques, de l'audio et de la vidéo), ils peuvent se présenter en format hypertexte, contenir un contenu dynamique produit automatiquement à partir de réserves de données ou posséder des fonctions de navigation, telles les barres d'outils, les recherches par mots-clés ou les tables des matières interactives. Il faut avoir l'habileté de conserver toutes les fonctionnalités de l'objet conservé.

## **4.2. La valeur probante d'un document**

La valeur probante ou preuve d'un document désigne la faculté d'un document à être utilisé par celui qui réclame l'exécution d'une obligation ou qui se prévaut d'un fait à qui produit l'extinction de son obligation pour un document électronique [Hercek, 2010] .

## **4.3. Présentation de l'archivage électronique**

Avec l'introduction des nouvelles technologies dans les collectivités et dans les administrations, beaucoup de documents autrefois tenus sous forme papier sont aujourd'hui gérés sous forme électronique.

A l'instar de l'archive papier, l'archivage électronique est l'ensemble des procédures définies par une collectivité pour assurer la conservation de son patrimoine documentaire

conformément à un référentiel défini par l'autorité chargée du stockage et de la mise à disposition future des documents effectivement versés. Il représente, au sens général, « l'ensemble des actions, outils et méthodes visant à identifier, recueillir, classer et conserver des informations électroniques, qui sont mis en oeuvre pour conserver à moyen et long terme ces informations dans le but de les exploiter». [mouhoubi, 2002b]. En comparaison avec l'archivage papier, l'archivage électronique gère les documents numériques, qui sont des « ensembles composés d'un contenu, d'une structure logique, d'attributs de présentation permettant leur représentation, dotés d'une signification intelligible par l'homme ou lisible par une machine» Il peut être créé à l'état natif ou obtenu par un processus de transformation d'un document physique, par exemple par numérisation. Les documents bureautiques, les bases de données, les messages électroniques, les dossiers numérisés sont considérés comme des documents numériques.

#### **4.4. Objectifs de l'archivage électronique**

L'archivage électronique des documents/flux à valeur probante doit donc poursuivre 4 objectifs. [Midi-Pyrénées, 2008] ;

##### **4.4.1. Pérennité/réversibilité**

Les documents créés par la collectivité et conservés soit par elle-même soit par un tiers sont déposés sur des supports garantissant une consultation à minima égale à leurs DUA (duré d'utilité administratif), voire pour une période plus longue si ce sont des archives définitives. Les formats ouverts acceptés pour tous les types de documents (textes, plans, images, etc...) et la périodicité de la procédure de conversion sont des composantes essentielles à définir en début de projet.

##### **4.4.2. Communication**

L'indexation du document par le système choisi doit permettre une consultation partielle ou totale du document. Ainsi le document/flux est décrit dans un index hébergé sur un serveur distinct du serveur de stockage. Il comporte outre la nature du document (délibération, arrêté, fiche de paye, dossier de travaux, déclaration fiscale ou sociale, notation, etc...) et la date de création, les éléments suivants :

- description libre,
- les dates de début et de fin d'archivage,
- la langue du contenu d'information (obligatoire),
- la taille (nombre d'octets, nombre d'enregistrements, nombre de boîtes, etc.),
- l'historique de la conservation (interventions pratiquées sur le contenu depuis sa création jusqu'à son transfert dans un service d'archives),

- la position du contenu d'information dans le plan de classement du service producteur,
- des informations de format,
- des références à des objets liés,
- d'autres données descriptives,
- le service producteur du contenu d'information et le service d'archives qui en est responsable,
- des règles d'accès et de conservation,
- des mots-clés.

Le contenu du document est stocké dans une base de données **distincte et redondante**.

Il est consultable dans l'index selon les droits d'accès définis dans la convention d'archivage par le service d'archives ou par le demandeur d'archives.

#### **4.4.3. Traçabilité**

Pour un document à valeur probante comme pour des documents placés en archive définitive, il est important de pouvoir suivre les étapes de conservation. L'index propose : l'historique de la conservation (interventions pratiquées sur le contenu depuis sa création jusqu'à son transfert dans un service d'archives) et les références de l'horodatage générés lors de la transmission aux serveurs pour le contrôle de légalité (Actes).

#### **4.4.4. Protection des données et archivage légal**

Pour les documents/flux dont la valeur probante doit être certifiée en raison de leur exigence de conformité à une réglementation pour être exécutoire, le système d'archivage électronique doit garantir la protection des données et l'archivage légal des documents versés (garantie d'intégrité, force probatoire et copies certifiées conformes).

### **4.5. Exigences de l'archivage électronique**

L'archivage électronique doit tout d'abord répondre aux exigences suivantes :

- L'intégrité des documents

Cela consiste en la protection du document. L'intégrité d'un document prouve que celui-ci n'a subi aucun ajout, ni aucune modification qui pourrait provoquer l'altération ou la destruction des données. Elles doivent être conservées dans un format qui sera lisible. Il existe deux types d'intégrité : « l'intégrité physique qui porte sur le support et la présentation formelle du document, et l'intégrité diplomatique qui porte sur ce qui reste de l'information après altération du support ». [Agnieszka, Bonard, 2010b]



- la pérennité des données

La pérennité consiste à maintenir dans le temps l'intégrité des données. En effet, un document doit être lisible à n'importe quel moment. Nous verrons, un peu plus bas, que les supports et formats de stockage deviennent rapidement obsolètes. Pour assurer cette pérennité, il est alors important de mettre en place un processus de contrôle des données qui vérifie leur bonne lisibilité. Une autre solution consiste en l'utilisation de formats standards normalisés et pérennes tels que XML ou PDF/A et de supports fiables comme les disques optiques ou les cassettes magnétiques.

- la sécurité des documents

Il s'agit d'assurer la sécurité physique des locaux et des données. Ce se fait par le biais de règles de sécurité pour les bâtiments (l'anti-intrusion, lutte contre l'incendie, etc.), la sauvegarde ou duplication des documents.

- l'authenticité des documents

ils'agit de prouver qu'un document est bien original. Pour cela, il faut prouver que : le document est bien ce qu'il prétend être, qu'il a été créé ou envoyé par la personne qui prétend l'avoir créé ou envoyé, et qu'il a été créé ou envoyé à la date prétendue.

Ces exigences permettent ainsi :

- de faciliter l'accès à l'information,
- de répondre aux exigences légales de conservation et de communication,
- de relever le défi de l'obsolescence technologique récurrente.

#### **4.6. L'organisation type d'un système d'archivage électronique [Midi-Pyrénées, 2008]**

##### **4.6.1. Les acteurs du processus :**

La production d'une archive par une collectivité et son versement vers un service idoine s'organise autour d'une chaîne d'acteurs dont les caractéristiques sont décrites ci-après.

- **Service producteur :**

Les agents des collectivités produisent des documents qui sont soit imprimés, signés et numérisés soit entièrement dématérialisés au travers des applications métiers. Ce service transmet le document via un tiers de confiance au service de l'Etat concerné. L'horodatage de la transmission et l'épuisement des délais légaux de recours confère la force probante.

- **Service versant :**

Chaque service producteur transfère les archives courantes périodiquement et selon une procédure définie préalablement vers une plate-forme d'archivage électronique soit vers le service Archives de la collectivité.

- **Service de contrôle :**

Il vérifie l'intégrité des paquets de documents transmis vers la plateforme.

- **Service d'Archives :**

Il peut effectuer le versement des archives. Ses missions principales sont l'indexation et la conversion des documents. Ce service est responsable du traitement des requêtes de consultation et des éditions.

- **Demandeur d'Archives (professionnels et grand public)**

Il consulte les documents archivés, effectue des requêtes et demande le cas échéant des copies conformes.

#### **4.6.2. Le contrat**

Un engagement entre la collectivité et le service ou le tiers chargé de l'archivage définit les modalités de transfert d'archives et d'élimination, les modalités de communication entre un système producteur, un système versant et un système d'archivage. Sur le plan technique, il définit les formats de documents acceptés, le type et la structure des archives versées et la périodicité des versements. Le contrat permet à la plate-forme de gérer les versements et au service de contrôle de réaliser des audits du système.

#### **4.6.3. La protection des données**

Dans le cas de documents à valeur probante ou de documents signés électroniquement, la réglementation liée au protocole standard d'échange impose la génération de fichiers d'identification et de vérification :

- l'empreinte : associée à un fichier ou à une partie de fichier pour en contrôler l'intégrité
- la signature : le standard permet d'insérer une signature électronique. Les signatures apposées à l'origine sur les données versées, demeurent avec les données dans leur format d'origine

Dans le cas particulier des documents signés électroniquement :

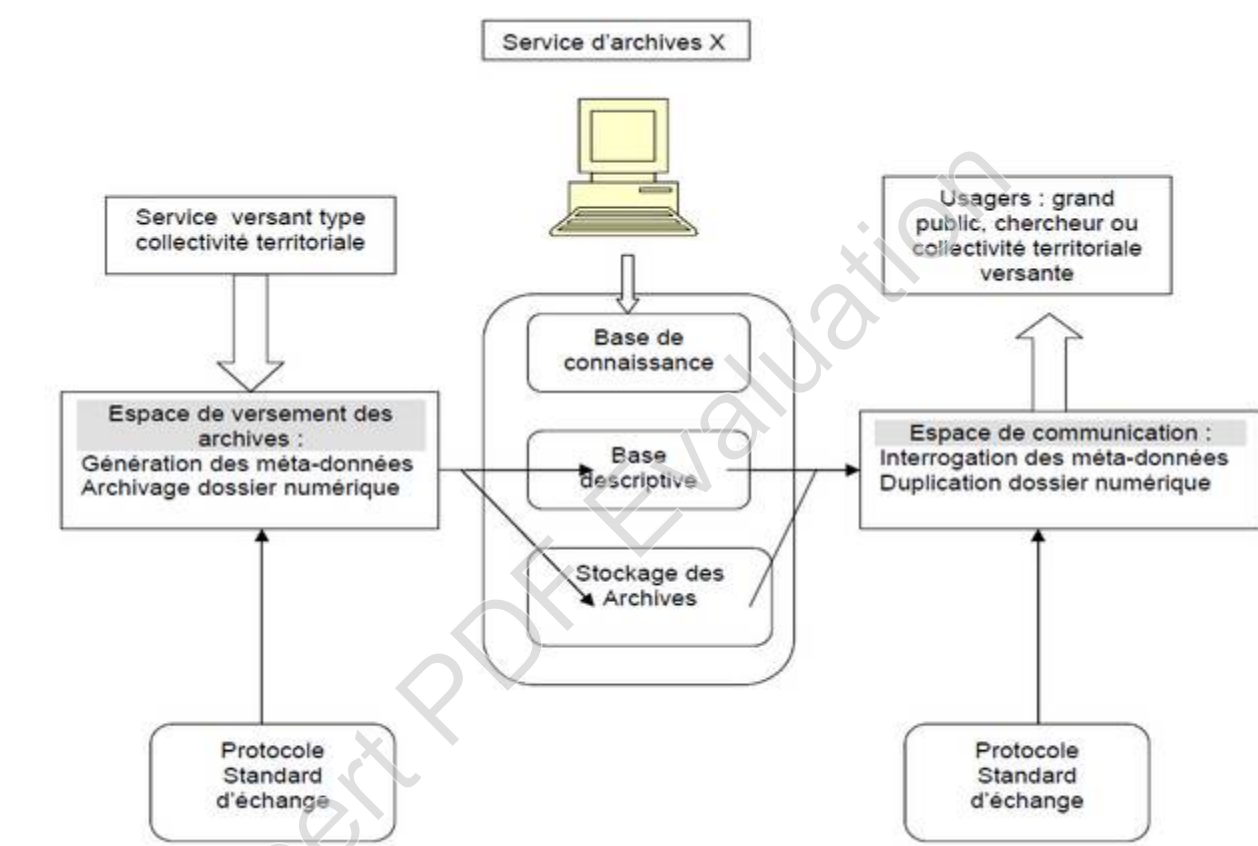
- il est recommandé de faire valider la signature peu de temps après son établissement (dans les limites de la durée de vie des certificats) et de faire figurer le résultat de cette vérification dans les métadonnées des documents qui seront ensuite archivés,
- il est préférable que soient enregistrés et figurés dans le document l'identification en clair du signataire ainsi que la date et heure de la signature.
- l'intégrité dans le temps est assurée par une enveloppe sécurisée d'un niveau organisationnel supérieur (ex. signature du service d'archives)

#### 4.6.4. Le processus type d'Archivage électronique

##### L'indexation et les métadonnées

Le document/flux est décrit dans un index hébergé sur un serveur distinct, la base descriptive. Celle-ci contient les métadonnées destinées à identifier et référencer le document.

Les renseignements contenus dans cet index sont constitutives de la base de méta données qui serviront, dans l'espace de communication, à rechercher des documents à partir d'un moteur de recherche.



Figur2.1 : Le processus type d'Archivage électronique [Midi-Pyrénées, 2008]

## 5. Cadre normatif

### Introduction

Une partie des documents et des données qui sont produits au quotidien dans les entreprises doit être archivée pour être conservée pendant quelques années ou quelques décennies, afin de se conformer aux contraintes réglementaires ou pour constituer une documentation interne pour les équipes. C'était vrai dans l'environnement papier et c'est encore plus vrai dans l'environnement électronique. En effet, si les papiers savent plus ou moins se conserver tous

seuls, les fichiers électroniques livrés à eux-mêmes sont rapidement inexploitable et illisibles. L'archivage électronique doit répondre aux problèmes posés par l'obsolescence des supports et des formats de codage, ainsi qu'à la nécessité d'identifier et de décrire les documents et les données numériques.

Pour l'entreprise, il s'agit d'anticiper le risque de non-disponibilité des données en cas de problème juridique ou technique, et de le faire au meilleur coût.

L'archivage électronique est aujourd'hui encadré par plusieurs normes qui traitent chacune un ou plusieurs aspects de la question : la norme internationale ISO 15489 propose des principes directeurs, des procédures et une démarche pour maîtrise de l'archivage ; le modèle européen MoReq2 détaille les spécifications fonctionnelles d'un système d'archivage électronique (*Electronic Records Management System*) ; la norme française NF 42-013 définit le cadre et les procédures de conservation des documents numériques natifs ou numérisés ; la norme OAIS définit un modèle d'information et un modèle fonctionnel pour la pérennisation de l'information numérique.

### **5.1. MOREQ2 [Chabin, 2008a]**

Moreq2 propose des spécifications génériques pour un système d'archivage électronique (SAE) ou un système électronique Records Management (ERMS).

Il s'agit d'une version mise à jour et enrichie du premier modèle d'exigences pour l'organisation de l'archivage électronique. Publié en 2001.

Moreq2 est destiné à être employé par un large éventail d'organisations des secteurs public et privé qui souhaite développer un système d'archivage électronique. Il peut être utilisé pour la rédaction des cahiers de charges, l'évaluation des systèmes existants, la préparation de supports de formation à la gestion documentaire, etc.

Moreq2 a été développé par la commission européenne à la demande du FORUM DLM (*document life cycle management*), une organisation européenne indépendante regroupant des services d'archives et d'autres organisations publiques et privées intéressées par l'archivage électronique, le *records management* et la gestion de l'information.

### **5.2. DUBLIN CORE [yahiaoui, 2008a]**

En Mars 1995 s'est tenu un *workshop* sur les Métadonnées, parrainé par *Online Computer Library Center (OCLC)* et le *National Center for Supercomputing Applications (NCSA)*, rassemblant 52 chercheurs et professionnels des bibliothèques, de l'informatique, et des spécialités connexes, pour faire avancer l'état de l'art dans le développement des descriptions de ressources électroniques. Les buts de ce *workshop* incluaient une compréhension commune des besoins, des points forts, des défauts, et des solutions à proposer; et l'atteinte d'un

consensus sur un ensemble d'éléments de métadonnées pour décrire des ressources d'informations.

Depuis qu'Internet contient plus d'information complète que de simples résumés professionnels, les indexeurs et catalogueurs tentent de gérer cela en utilisant les méthodes et systèmes existants: Il était évident qu'une alternative pour obtenir des métadonnées utilisables pour des ressources électroniques doit donner aux auteurs et aux fournisseurs d'information un moyen permettant de décrire les ressources eux-mêmes, sans formation intensive et spécifique préalable.

C'est pour atteindre ce but, que la tâche majeure du workshop sur les métadonnées, était d'identifier et de définir un ensemble simple d'éléments pour décrire des ressources électroniques diffusées (en réseau).

La première version de la synthèse des travaux du *workshop* établissait un ensemble minimal de treize éléments de métadonnées qui fut nommé : ensemble d'éléments de *core* de métadonnées de *Dublin* (ou plus simplement Dublin Core)..

Aujourd'hui, le Dublin Core est maintenu par le ***Dublin Core Metadata Initiative (DCMI)*** et donne lieu à une conférence annuelle au contenu riche et varié. Le DCMI comporte de nombreux groupes de travail, comme le groupe « éducation » qui tient une conférence commune avec le groupe LOM.

### **5.2.1. Intérêts et limites**

Dublin Core fait l'objet d'un large consensus et d'une large utilisation aujourd'hui grâce aux atouts suivants :

- Sa création dans un contexte international et multidisciplinaire
- Sa sémantique simple et "commune", facilement compréhensible, particulièrement pour les éléments de base
- Son extensibilité (compatible avec d'autres jeux d'éléments, évolutivité) et sa flexibilité (grande souplesse d'implémentation)
- Son adoption dans différents domaines, métiers et pays, et dans des applications non prévues initialement ou des domaines industriels connexes
- Son évolutivité au travers de groupes de travail ouverts
- La volonté du DCMI, de diffuser et faire adopter ce modèle
- La normalisation des 15 éléments de base à partir de 2003 par l'ISO Cependant, on peut résumer au passif du Dublin Core :
- Son côté généraliste et incomplet, nécessitant souvent des extensions

- Sa relative jeunesse et le fait qu'il évolue encore (bien que les éléments de base semblent être « gravés dans le marbre »).

### 5.2.2. Quelques principes du Dublin Core

Les auteurs du Dublin Core ont établi une liste de principes qui devaient guider davantage le développement de l'ensemble des éléments de métadonnées. Ces principes sont : propriété intrinsèque, extensibilité, indépendance de syntaxe, optionalité, répétabilité et modifiabilité .

- **Propriété intrinsèque**

Le Dublin Core a été dirigé, dès le départ pour décrire des propriétés intrinsèques de l'objet. Par exemple, l'élément "Sujet" est une donnée intrinsèque, tandis que des informations de transaction telle que coût et droit d'accès sont des données extrinsèques.

- **Extensibilité**

En plus de son emploi en traitant de la propriété intrinsèque des données, le mécanisme d'extension permettra l'inclusion de données intrinsèques pour des objets qui ne peuvent pas être décrits suffisamment par un petit ensemble d'éléments.

L'extensibilité est importante parce que les utilisateurs peuvent vouloir ajouter des descriptions supplémentaires à des champs, même ayant des buts spécifiques. En outre, la spécification du Dublin Core lui-même peut changer dans le temps, et le mécanisme d'extension permet des révisions.

- **Indépendance de Syntaxe**

Les constructions syntaxiques sont évitées parce qu'il est trop tôt de proposer des définitions formelles et parce que le Dublin Core est destiné à être employé tôt ou tard dans une large gamme de programmes d'application et de disciplines.

- **Optionalité**

Tous les éléments sont optionnels, pour deux raisons :

-La première est que le Dublin Core peut tôt ou tard être appliqué aux objets pour lesquels certains éléments n'ont pas de signification : Qui est l'auteur d'une image satellitaire?

-La seconde est qu'il semble futile de donner des descriptions complexes lorsque les auteurs du contenu prévoient de fournir la matière descriptive

- **Répétabilité**

Tous les éléments du Dublin Core sont répétables. Par exemple, plusieurs éléments 'auteur' seraient employés quand une ressource a plusieurs auteurs.

- **Modifiabilité**

Chaque élément dans le Dublin Core a une définition qui est sensée être évidente. Cependant, il est aussi nécessaire que les définitions des éléments satisfassent aux besoins de communautés différentes. Ce but est accompli en permettant à chaque élément d'être modifié par un qualificateur optionnel. Si aucun qualificateur n'est proposé, l'élément prend son sens commun.

Les qualificateurs seront typiquement dérivés de conventions connues dans la communauté des bibliothèques ou du domaine des connaissances propres à la ressource. Les qualificateurs sont importants parce qu'ils donnent au Dublin core, un mécanisme pour supprimer ou amoindrir l'écart entre des spécialistes et des utilisateurs quelconques.

Par exemple, les données dans l'élément 'sujet' consistent en mots ou expressions qui décrivent le contenu de l'objet. Cependant, un professionnel du catalogage peut vouloir se référer à la source d'où les termes du sujet sont pris. Dans un tel cas, l'élément peut être écrit ainsi : **Sujet (scheme = LCSH)**, indiquant que les termes de sujet sont pris de la liste: *Library of Congress Subject Headings*.

### 5.2.3. Liste des éléments du Dublin Core

Les éléments du DC peuvent être classés en trois groupes qui indiquent la classe ou le type d'information correspondante

- Des éléments qui concernent principalement le contenu de la ressource : Titre, sujet, description, source, langue, relation et couverture
- Des éléments liés à la propriété intellectuelle de la ressource : Créateur, éditeur, contributeur et droit
- Des éléments correspondants surtout à la matérialisation de la ressource : Date, type, format et identifiant.

### 5.2.4. Dublin Core qualifié

Les quinze éléments de base sont considérés comme un dénominateur commun et constitue ce que l'on appelle « Dublin Core simple ». Dans la plupart des cas, ces éléments sont insuffisants pour décrire une ressource spécifique à un domaine particulier avec précision.

C'est pour cela que dès le départ, les éléments de base ont été étendus (ou précisés) par un ensemble d'autres termes, parfois appelés « qualifieurs ». Deux classes de « qualifieurs » sont reconnues :

- Les « raffinements d'éléments » qui rendent plus spécifique le sens d'un élément
- Les « schémas d'encodage » ou vocabulaires contrôlés.

### 5.2.5. Les éléments supplémentaires et « raffinements d'éléments »

Le DCMI a défini trois éléments supplémentaires et une trentaine de raffinements d'éléments.

- Les éléments supplémentaires : **Audience, Provenance et RightsHolder** étendent

l'ensemble des quinze éléments de base.

- Les raffinements d'éléments précisent le sens d'un élément existant

### 5.3. NF Z 42-013 [Dronne, 2006a]

La norme **NFZ42013** énonce un ensemble de spécifications de l'ANFOR concernant les mesures techniques et organisationnelles à mettre en œuvre pour l'enregistrement, le stockage et la restitution de documents électroniques afin d'assurer la conservation et l'intégrité.

Le titre de la norme révisée en mars 2009 a évolué : (spécifications relatives à la conception et l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes)

La norme révisée est destinée :

- Aux organismes ou entreprises qui souhaitent mettre en œuvre des systèmes informatiques dans lesquels ils pourront stocker des images des documents électroniques de telle façon que leur fidélité et leur intégrité par rapport aux documents d'origine soient assurées
- Aux éditeurs de système d'archivage électronique
- Aux sociétés de services informatiques qui désirent concevoir ces systèmes
- Aux entreprises de services assurant le stockage de documents électroniques

### 5.4. OAIS *Open Archival Information System*. [micaelli, 2010]

Le modèle de référence OAIS (*Référence Model for an Open Archival Information System*) est le résultat d'un travail du CCSDS (*Consultative Committee for Space Data Systems*, organisme international de normalisation des agences spatiales) réalisé à la demande l'ISO.

Ce travail, auquel ont été associés des représentants de bibliothèques et d'archives institutionnelles, a abouti en 2002 à la rédaction d'un document qui spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage.

OAIS ne préjuge de la nature des objets à archiver, mais le modèle est particulièrement bien adapté aux objets numériques. Le modèle de référence OAIS est aujourd'hui une norme internationale ISO, (ISO 14721). Sa traduction en français, projet de norme ISO, est disponible depuis le printemps 2005. OAIS est un modèle abstrait. Il définit une terminologie et des concepts. Il identifie les acteurs, décrit les fonctions et les flux d'information, et



propose un modèle d'information adapté à la problématique de l'archivage numérique. OAIS n'est pas un recueil de spécifications techniques destinées à être directement implémentées. C'est un guide qui cerne la problématique dans son ensemble et qui contraint à se poser toutes les bonnes questions.

En tant que modèle conceptuel de référence, OAIS s'est aujourd'hui largement imposé, au niveau international, auprès de tous les acteurs institutionnels de l'archivage numérique.

#### **5.4.1. Les acteurs de l'archivage**

Le modèle OAIS identifie quatre rôles principaux dans un système d'archivage :

un acteur interne, l'« archive », c'est-à-dire l'opérateur du système d'archivage, et trois acteurs externes : le « management », les « producteurs » et les « utilisateurs ». [micaelli, 2010] :

- **Le « management » :**

Assure la fonction de décideur politique. Pour le management, le système d'archivage s'insère dans un plan stratégique global dont il n'est qu'un élément parmi d'autres. C'est évidemment au management qu'il revient de soutenir le dispositif, politiquement, financièrement, et sur le très long terme

- **Les « producteurs » :**

sont les personnes, ou plus vraisemblablement les organismes, qui fournissent les objets à archiver. Les objets numériques sur lesquels travaillent les producteurs avant archivage sont les SIP (*submission information package*). Une fois archivés, ils deviennent les AIP (*archival information package*) objets internes à l'Archive.

- **Les « utilisateurs » :**

Quant à eux, sont les organismes et les personnes qui ont accès aux objets archivés. Les objets numériques mis à la disposition d'utilisateurs sont les DIP (*dissemination information package*). OAIS identifie une classe particulière d'utilisateurs - la « communauté d'utilisateurs cible » - comme étant la population prioritairement bénéficiaire du service d'archivage. Les services à fournir seront distincts selon que la communauté d'utilisateurs cible est nombreuse ou non, experte ou « grand public », etc.

#### **5.4.2. Les fonctions de l'archivage :**

OAIS cartographie un système d'archivage selon 6 grands domaines fonctionnels [micaelli, 2010] :

-L'entité « **entrées** » reçoit, contrôle et valide les objets à archiver. Les objets eux-mêmes sont transmis à l'entité « stockage », tandis que les informations nécessaires à leur description et à leur gestion dans le temps sont transmises à l'entité « gestion des données ».

-L'entité « **stockage** » assure la conservation physique des objets archivés. Elle tient les objets archivés à la disposition de l'entité « accès ». Conformément aux règles établies par l'entité « administration », elle prend en charge la réalisation des copies multiples et le renouvellement des supports anciens.

-L'entité « **gestion des données** » prend en charge la tenue à jour de toutes les informations internes - base de données- nécessaires au système d'archivage. Elle fournit aux autres entités du système les informations descriptives des objets archivés (notamment à l'entité « accès ») et toutes les informations de gestion techniques et archivistiques nécessaires.

-L'entité « **administration** » assure la coordination générale du système. Elle en établit les règles internes. Elle veille à la qualité globale du service rendu et à son amélioration. Elle rend compte au management externe.

-L'entité « **planification de la pérennisation** » est la cellule de veille et de planification du système. Elle « écoute » l'environnement extérieur et émet des recommandations en vue de procéder aux évolutions nécessaires, notamment aux évolutions technologiques. Elle prépare et planifie ces évolutions. Elle est également responsable du suivi des changements qui peuvent s'opérer dans la communauté d'utilisateurs cible en vue de garantir que le service d'accès reste conforme aux attentes nouvelles des utilisateurs.

L'entité « **accès** » regroupe tous les services qui sont en interface directe avec les utilisateurs. Outre les fonctions de contrôle d'accès, il s'agit principalement de permettre aux utilisateurs de rechercher dans le « catalogue » des objets archivés, et de leur fournir les objets dont ils passent commande.

#### **Les migrations numériques :**

La préservation sur le très long terme d'un objet numérique ne peut se garantir que grâce à une succession d'opérations de migration. Ces migrations peuvent s'opérer à plusieurs niveaux. Elles peuvent être de simples recopies, ou bien des opérations plus complexes de reformatage. Le modèle OAIS distingue quatre types de migrations numériques :

#### **Le rafraîchissement :**

Copie bit à bit de l'information, dans laquelle un support généralement ancien est remplacé par un support identique généralement neuf.

#### **La duplication :**

recopie des objets archivés vers un nouveau type de support, sans changement de

l'organisation logique du stockage. Un exemple typique de duplication est le transfert des objets archivés vers des supports similaires plus capacitifs.

**Le ré-empaquetage :**

Recopie des objets archivés vers un nouveau type de support, nécessitant une nouvelle organisation logique du stockage. Typiquement, il peut s'agir d'une recopie d'objets archivés sur des supports disques on-line, vers des supports bandes off-line.

**La transformation :**

il ne s'agit plus d'une copie, mais d'une réelle modification du contenu de l'information, portant notamment sur sa forme. La transformation peut ou non être réversible. Un exemple de transformation pourrait être la conversion d'un document au format Word de MSOffice vers le format PDF.

### **5.5. ISAD (G) [CIA, 1999]**

La norme générale et internationale de description archivistique (en anglais *Central International Standard Archival Description*), communément désigné sous le sigle ISAD(G), a été publiée officiellement par le conseil international des archives en 1994. La dernière version date en 1999, son objectif est de permettre la diffusion d'informations sur le contenu et l'intérêt des fonds d'archive tant entre services d'archives qu'à destination des chercheurs, grâce à descriptions homogènes, des fichiers d'autorité communs et des accords sur les formats d'échange.

Elle repose sur trois principes :

- Validité du concept de fonds et donc du principe de respect des fonds
- Correspondance entre les niveaux de classement et les niveaux de description
- liaison hiérarchique des descriptions d'un niveau à l'autre

Elle comporte 26 éléments descriptifs dans la combinaison permet de décrire n'importe quelle unité archivistique.

Ces éléments sont répartis en 6 zones : identifications, contexte, contenu, conditions d'accès, sources complémentaires et notes.

La première zone qui est la plus importante comprend la cote de l'unité décrite, l'intitulé qui en résume le contenu, les dates extrêmes, le niveau de description choisi et des indications sur le volume de l'unité décrite (métrage linéaire, nombre de pièces, etc.) .

ISAD (G) ne traite pas de la gestion matérielle, de l'indexation et des règles d'écriture et de vocabulaire.

Voici la liste des descripteurs ISAD(G) : [CIA, 1999]

### 5.5.1. ZONE D'IDENTIFICATION

#### a) Référence

**-Objectif :** Identifier de manière unique l'unité de description et établir un lien avec la description qui la représente.

**-Règle :** Inscrire, dans la mesure où ils sont nécessaires pour une parfaite identification, les éléments suivants:

Le code du pays conformément à la dernière version de l'ISO 3166, Codes pour la représentation des noms de pays

Le code du service d'archives conformément à la norme nationale de codification des services d'archives, ou tout autre élément d'identification du lieu de conservation

la cote ou tout autre élément d'identification spécifique.

Chacun de ces éléments est nécessaire lorsqu'on envisage l'échange d'informations au niveau international.

#### b) Intitulé / analyse

**-Objectif :** Dénommer l'unité de description.

**-Règles :** Indiquer soit le titre que porte le document, soit un titre forgé ou une analyse concise, conformément aux règles de la description à plusieurs niveaux et aux usages nationaux.

Si nécessaire, abréger un titre trop long dans la mesure où on ne perd pas d'information essentielle.

Pour le titre forgé ou l'analyse, au niveau de description supérieur, introduire le nom du producteur. Aux niveaux inférieurs on peut introduire, par exemple, le nom de l'auteur du document, un terme indiquant le type de documents composant l'unité de description et, dans les cas appropriés, une expression faisant mention de l'action, de l'objet ou du lieu de conservation.

Différencier intitulé d'origine et titre forgé suivant les usages nationaux ou linguistiques.

#### c) Dates

**-Objectif :** Identifier et donner la (les) date(s) de l'unité de description.

**-Règle :** Donner au moins l'une des dates suivantes pour l'unité de description, en fonction des documents décrits et du niveau de description :

1. Date(s) indiquant la période au cours de laquelle les documents ont été rassemblés dans le cadre de l'activité du producteur
2. Date(s) à laquelle les documents ont été créés. Cette notion inclut naturellement la date de création des documents originaux rassemblés après la constitution organique de l'unité de

description, mais aussi celle des copies, des éditions, des différentes versions, des pièces jointes et même des originaux créés avant la constitution organique des dossiers.

Préciser la nature des dates indiquées. D'autres dates peuvent être fournies et identifiées selon les usages nationaux

Indiquer une seule date ou plusieurs dates selon le cas. Lorsque deux dates sont précisées, elles doivent être entendues inclusivement.

#### **d) Niveau de description**

**Objectif :** Identifier le niveau de classement auquel se situe l'unité de description.

**Règles :** Donner le niveau de cette unité de description.

#### **e) Importance matérielle et support de l'unité de description (quantité, volume ou dimension)**

**-Objectif :** Déterminer et donner

- a. l'importance matérielle de l'unité de description ;
- b. le support de l'unité de description.

**-Règles :** Indiquer l'importance matérielle de l'unité de description, c'est-à-dire le nombre d'unités matérielles (en chiffres arabes) et l'unité de mesure. Donner le(s) support(s) spécifique(s) de l'unité de description.

Ou bien, préciser le métrage linéaire ou le volume de l'unité de description. Si la mesure d'une unité de description est donnée en mètres linéaires et que l'on juge utile d'ajouter des renseignements supplémentaires, les inscrire entre parenthèses.

### **5.5.2. ZONE DU CONTEXTE**

#### **a) Nom du producteur**

**-Objectif :** Identifier le ou les producteurs de l'unité de description.

**-Règle :** Donner le nom de l'organisme (ou des organismes) ou de l'individu (ou des individus) responsable(s) de la création, de l'accroissement ou de la conservation de l'unité de description.

Le nom du producteur doit être donné selon une forme normalisée, conformément aux usages nationaux ou internationaux et aux principes de la norme ISAAR(CPF) .

#### **b) Histoire administrative/Notice biographique**

**-Objectif :** Fournir un historique ou une biographie du ou des producteurs de l'unité de description afin de situer les documents dans leur contexte et de les rendre plus intelligibles.

**-Règles :** Donner succinctement tout renseignement important sur l'origine, l'évolution et les fonctions du ou des organismes, ou sur la vie et les activités du ou des individus qui ont

produit l'unité de description. Si des renseignements supplémentaires figurent dans une source publiée, mentionner cette source.

Pour les personnes physiques ou les familles, donner leurs nom et prénom, leurs titres, dates de naissance et de décès, lieu de naissance, lieux de résidence, activités, profession ou postes occupés, nom d'origine et tout autre nom, réalisations importantes et lieu de décès.

#### **c) Historique de la conservation**

**-Objectif :** Fournir sur l'histoire de l'unité de description les renseignements importants qui contribuent à son interprétation et qui prouvent son authenticité et son intégrité.

**-Règles :** Enumérer les changements successifs de propriété, de responsabilité et/ou et de conservation de l'unité de description et indiquer les événements tels que l'histoire du classement, la production contemporaine d'instruments de recherche, la réutilisation des documents à d'autres fins ou le changement de logiciel, qui ont contribué à sa structure et à son classement actuels. Donner les dates de ces événements, dans la mesure où ils sont connus de façon certaine. Si l'historique de la conservation est inconnu, le signaler.

#### **d) Modalités d'entrée**

**-Objectif :** Identifier les circonstances de l'entrée dans les archives.

**-Règle :** Mentionner l'origine de l'entrée, ainsi que sa date et ses modalités, si ces renseignements ne sont pas confidentiels. Si l'origine est inconnue, le signaler. Eventuellement, inscrire le numéro d'inventaire.

### **5.5.3. ZONE DU CONTENU ET DE LA STRUCTURE**

#### **a) Présentation du contenu**

**-Objectif :** Permettre aux usagers d'apprécier l'intérêt de l'unité de description.

**-Règle :** Donner une brève présentation de la portée de l'unité de description (par exemple, la période chronologique ou la zone géographique concernées) et de son contenu (par exemple, la typologie générale, l'objet général, les procédures administratives). Cette présentation doit être adaptée au niveau de description.

#### **b) Évaluation, tris et éliminations, sort final :**

**-Objectif :** Fournir des renseignements sur l'évaluation, les opérations de tri ou d'élimination et le sort final des documents.

**-Règles :** Mentionner toute opération de tri et d'élimination effectuée ou programmée sur l'unité de description, surtout si elle est susceptible de modifier l'interprétation des documents. Le cas échéant, indiquer le responsable de l'opération.

**c) Accroissements :**

**-Objectif :** Informer l'utilisateur de possibles entrées complétant l'unité de description.

**-Règle :** Indiquer si des accroissements sont prévues. Le cas échéant, en évaluer la quantité et la fréquence.

**d) Mode de classement :**

**-Objectif :** Fournir des renseignements sur la structure, l'ordre ou le plan de classement de l'unité de description.

**-Règles:** Préciser la structure, l'ordre ou le plan de classement de l'unité de description. Indiquer les modalités du traitement effectué par les archivistes. Pour les documents électroniques, donner les renseignements de référence sur l'architecture du système.

On peut, si on le souhaite et en fonction des usages nationaux, placer plutôt toutes ces informations dans l'élément "Présentation du contenu". [CIA, 1999]

**5.5.4. ZONE DES CONDITIONS D'ACCÈS ET D'UTILISATION :**

**a) Conditions d'accès**

**-Objectif :** Fournir des renseignements sur le statut juridique et sur les autres dispositions de toute nature qui restreignent ou affectent l'accès à l'unité de description.

**-Règle :** Préciser en les distinguant les dispositions juridiques, les conventions particulières, les règlements intérieurs ou les considérations de toute nature qui ont des conséquences sur l'accès à l'unité de description. Lorsque c'est possible, indiquer la durée pendant laquelle l'accès restera réservé, et la date à laquelle les documents pourront être consultés.

**b) Conditions de reproduction**

**-Objectif :** Signaler toute restriction relative à la reproduction de l'unité de description.

**-Règle :** Donner des renseignements sur les conditions de reproduction de l'unité de description (en particulier les dispositions relatives au droit d'auteur), une fois autorisé l'accès aux documents. Si l'existence de telles conditions n'est pas connue, le préciser. S'il n'y a pas de restriction, aucun énoncé n'est nécessaire.

**c) Langue et écriture des documents**

**-Objectif :** Identifier la ou les langues, les systèmes d'écriture et de symboles utilisés dans l'unité de description.

**-Règle :** Indiquer la ou les principales langues utilisées dans les documents qui constituent l'unité de description. Mentionner tout alphabet ou écriture particulière et tout système de symboles ou d'abréviations utilisé.

**d) Caractéristiques matérielles et contraintes techniques**

**-Objectif :** Fournir des renseignements sur toute caractéristique matérielle ou contrainte technique importante qui a des conséquences sur l'utilisation de l'unité de description.

**-Règle :** Indiquer toute caractéristique matérielle importante qui limite l'utilisation de l'unité de description. Préciser en particulier les contraintes nécessitées par le souci de la préservation des documents. Indiquer tout logiciel ou matériel requis pour accéder à l'unité de description.

**e) Instruments de recherche :**

**-Objectif :** Signaler tout instrument de recherche intéressant l'unité de description.

**-Règle :** Donner des renseignements sur tout instrument de recherche que peuvent détenir le service d'archives ou le producteur et qui est susceptible d'informer l'utilisateur sur le contexte et le contenu de l'unité de description. Le cas échéant, indiquer où on peut en obtenir copie.

### **5.5.5. ZONE DES SOURCES COMPLÉMENTAIRES**

**a) Existence et lieu de conservation des originaux**

**-Objectif :** Quand l'unité de description est une reproduction, indiquer si les documents originaux existent encore, et s'il y a lieu, leur lieu de conservation et leurs conditions d'accès.

**-Règle :** Si les documents originaux sont encore conservés (que ce soit dans la même institution ou ailleurs), préciser leur lieu de conservation, leur référence, et les conditions d'accès. Il convient également de préciser si les documents originaux n'existent plus, ou si l'on ignore l'endroit où ils sont conservés.

**b) Existence et lieu de conservation de copies**

**-Objectif :** Signaler l'existence, le lieu de conservation et la disponibilité de copies de l'unité de description.

**-Règle :** Si l'unité de description existe sous une autre forme (soit dans le service d'archives soit ailleurs), le mentionner, en donner les cotes et le lieu de consultation.

**c) Sources complémentaires**

**-Objectif :** Signaler les sources complémentaires.

**-Règle :** Donner les informations sur les unités de description conservées, soit dans le même service d'archives, soit ailleurs, et qui ont un lien avec l'unité décrite soit en raison de leur provenance, soit pour un autre motif. Rédiger un paragraphe d'introduction approprié et expliquer la nature du lien. Si l'unité de description complémentaire est un instrument de recherche, utiliser l'élément de description pour faire cette mention.

**d) Bibliographie**

**-Objectif :** Identifier toutes les publications qui sont relatives à l'unité de description, qui l'utilisent ou qui l'étudient.



**-Règle :** Donner la référence, complétée le cas échéant par d'autres informations, de toutes les publications qui sont relatives à l'unité de description, qui l'utilisent ou qui l'étudient. Citer également les publications de fac-similés, ou les transcriptions.

#### **5.5.6. ZONE DES NOTES**

##### **a) Notes**

**-Objectif :** Fournir des renseignements qui ne peuvent être donnés dans aucune des autres zones.

**-Règle :** Donner les renseignements particuliers et/ou importants qui n'ont pas été mentionnés dans d'autres zones

#### **5.5.7. ZONE DE CONTRÔLE DE LA DESCRIPTION**

##### **a) Notes de l'archiviste**

**-Objectif:** Expliquer comment et par qui la description a été établie.

**-Règle:** Mentionner les sources consultées pour établir la description, et l'auteur de celle-ci.

##### **b) Règles ou conventions**

**-Objectif:** Identifier les règles ou les conventions sur lesquelles est fondée la description

**-Règle:** Citer les règles ou les conventions internationales, nationales ou locales suivies pour établir la description.

##### **c) Date(s) de la description**

**-Objectif:** Indiquer à quelle date la description a été préparée et/ou révisée.

**-Règle:** Donner la date à laquelle la description a été effectuée ou révisée

#### **5.6. EAD (*Encoded Archival Description*)**

L'EAD (*Encoded Archival Description*) [Ou Rabah, 2008a]

est un format basé sur le langage XML qui permet de structurer des descriptions des manuscrits ou de documents d'archives

Destiné à la description d'une production documentaire (métadonnées descriptives), l'EAD permet de structurer les instruments de recherche de type archivistique : inventaire, répertoire, catalogue de collection .....

L'EAD est utilisé en particulier par les services d'archives, bibliothèques et musées pour décrire des fonds d'archives, des collections de manuscrit, et des collections hiérarchisés de documents ou d'objets (photographies, microfilms, pièces de musée.).

Ce format a été initialement conçu pour permettre le traitement rétrospectif des instruments de recherche existants, documents imprimés ou produits avec outils de traitement de texte dans la perspective d'une édition papier. Ceci explique que :

- La structure générale du modèle reste proche de celle des documents imprimés
- Les règles d'utilisation sont peu contraignantes, afin de pouvoir être adaptées à des contextes divers
- Le modèle combine des éléments de description structurée et des éléments d'encodage de texte

### **Objectifs de L'EAD**

Ce format a été développé dans les années 1990 pour objectifs

- De développer un modèle permettant de traiter les instruments de recherche existants, dans leur diversité de forme et de structure
- De restituer l'organisation hiérarchisée des instruments de recherche et les interrelations entre les composants
- De conserver le principe d'héritage des informations entre les niveaux

### **5.7. La norme ISO 15489 en réponse aux besoins d'archivage [Mouette, 2007a]**

En 1996, une norme nationale australienne sur le Records management remporte un réel succès qui amène à sa présentation devant l'International Standard Organization (Organisation internationale de normalisation). En 2001, après divers débats sur les problèmes de compatibilité du projet de la norme avec les droits nationaux des pays, la norme ISO 15489 est publiée. Elle est reprise en 2002 par l'AFNOR (Association française pour la normalisation), sous le même titre NF ISO 15489. La définition du Records management de la norme est : « champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liées à la forme des documents ». Elle est à ce jour la seule norme sur l'organisation et la gestion des archives courantes et intermédiaires.

Le Records management a pour objet l'ensemble des documents à archiver, c'est-à-dire les documents que l'organisme a décidé de préserver à titre de preuve ou en raison de leur valeur informationnelle : il s'agit des documents essentiels ou utiles dans la conduite de ses activités. Le périmètre du Records management est donc rarement exhaustif, il couvre une partie des documents produits ou reçus. Ceux-ci sont pris en compte dans leur version définitive (non modifiable). La norme précise que les archives définitives ne rentrent pas dans le champ du Records management.

Cependant, au regard de la réglementation archivistique française qui oblige les organismes du secteur public à gérer correctement leurs archives sur tout leur cycle de vie, les documents

appelés à terme à devenir des archives historiques sont nécessairement couverts dès leur création par le Records management.

La norme souligne le rôle et l'importance d'une politique de Records management au sein d'un organisme en identifiant les responsabilités de chaque catégorie de personnel. Elle conforte l'idée que l'ensemble des collaborateurs a une part de responsabilité dans la gestion des records mais que certains ont des rôles plus spécifiques. Toutefois, cette norme s'adresse principalement aux dirigeants, professionnels de l'information et au records manager.

## **Conclusion**

L'archivage électronique a connu une avancée décisive, sur la signature électronique et la production de normes, tant pour l'organisation et la structuration de l'information à archiver que pour la conservation numérique pérenne, et avec les développements technologiques au service de la sécurité, de la traçabilité et de migration de l'information.

Toutefois, on compte plus de projets que de réalisations et de retours d'expérience, notamment en matière de conservation à long terme. Le contexte réglementaire, le nombre d'acteurs et de compétences impliquées dans un projet global et la grande variété des types d'information à gérer s'ajoutent aux défis proprement technologiques d'obsolescence des supports et des formats ou d'interopérabilité des systèmes.

Les années à venir verront nécessairement d'autres progrès. Les plus attendus concernent l'harmonisation des normes, la maîtrise des formats (tant leur nombre que les formats spécifiques ou complexes comme les fichiers 3D), une meilleure visibilité des risques et des coûts, avec une évolution de l'offre logicielle pour y répondre.

# Chapitre III

## Réalisation d'un prototype d'un système d'archivage électronique

### 1. Présentation de l'environnement d'étude

Avant de présenter en détail l'objectif visé par notre étude, nous devons parler de l'environnement dans lequel nous avons effectué le stage et qui est le centre régional des archives de la wilaya de Guelma.

#### 1.1. Présentation

Le centre régional des archives est encore une jeune institution (création administrative en cours). Jusqu'au moment de la rédaction de ce mémoire il joue le rôle du service des archives de la wilaya de Guelma. Il a été construit dans le but de rendre accessible les documents administratifs de la wilaya de Guelma aux administrations, aux chercheurs et aux citoyens ainsi que d'autres.

#### 1.2. Mission

Actuellement, le centre régional d'archives est placé directement sous l'égide du secrétariat général de la wilaya de Guelma et a pour mission de conseiller et d'assister les services d'administration concernant les documents qu'ils produisent et qu'ils reçoivent. Pour cela un calendrier de conservation a été mis en place pour ce service avec des archivistes assurant la relation entre les archives de la wilaya et les entités concernées.

#### 1.3. Organigramme

La figure 3.1 décrit l'organigramme hiérarchique actuel du centre régional des archives.

#### 1.4. Fonctionnement

Comme toute autre organisme algérien (à notre connaissance), le centre régional des archives continu à utiliser la méthode ancienne pour la gestion et l'exploitation des archives documentaires. Cette gestion présente beaucoup d'anomalies qui vont être décrites dans la deuxième section.

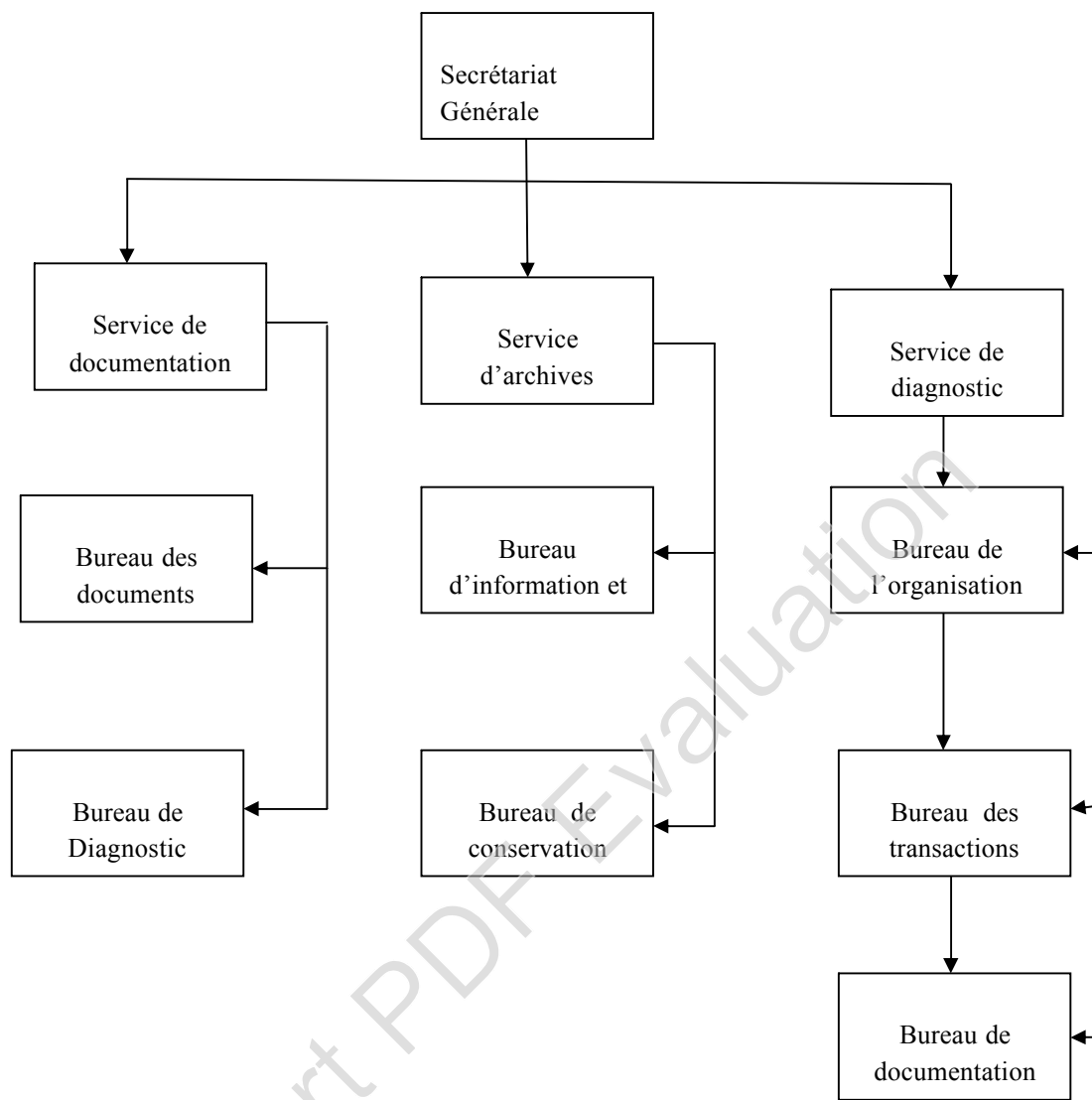


Figure 3.1 : Organigramme du centre régional des archives

## 2. Résultats de l'étude menée

A l'issue de notre étude, nous avons relevé certaines anomalies inhérentes à la gestion des archives.

L'analyse a mis en évidence que les archives se présente sous forme papier (la méthode classique), le papier étant fortement ressenti comme essentiel. La quantité de papier conservée est donc importante, ce qui complique la tâche de recherche d'une information, ajouter à cela l'inexistence d'une politique d'archivage et absence de moyens de gestion

efficaces telle qu'une application informatique ou une base de données ou un système d'archivage électronique. Les techniques d'archivage classiques ne sont même pas respectées. On utilise des registres d'archives et des boîtes d'archives avec un minimum de description, qui seront ensuite stocker d'une manière anarchique dans un des lieux de stockage du centre, ce qui mobilise une grande superficie et un personnel important pour accomplir les tâches de classement.

La qualité de l'archivage étant fortement dépendante des facteurs humains des erreurs de classement et des destructions accidentelles ne peuvent jamais être totalement évitées.

En ce qui concerne le côté équipement informatique, le centre régional des archives contient deux micro-ordinateurs, une imprimante et une photocopieuse, ce qui représente un sous équipement flagrant eu égard à l'ampleur de cette structure.

Pour remédier à ces anomalies, nous avons discuté avec les responsables du centre la possibilité de l'application des normes mondiales (étalées dans le premier chapitre) qui permettra de moderniser considérablement le centre régional d'archives. Ces responsables nous ont répondu qu'ils ont conscience de cette proposition et son application dépend fortement de la volonté de la tutelle.

De notre point de vue et eu égard à la masse importante des informations, une solution radicale serait de développer un système d'archivage électronique qui prendra en charge une grande partie du suivi de la gestion administrative (dématérialisation, indexation et recherche d'informations). La solution préconisée sera décrite dans ce qui suit.

### **3. Solution proposée**

#### **3.1. Solution organisationnelle et fonctionnelle**

Une solution organisationnelle sera de créer une nouvelle structure, intitulée service d'archivage électronique, rattachée au service des archives qui sera composé des bureaux suivants :

- Bureau de dématérialisation
- Bureau d'indexation et de recherche

Le bureau d'information et de l'assistance et le bureau de conservation seront rattachées à une nouvelle structure intitulée service d'archivage classique.

La figure 3.2 montre le schéma organisationnel de la structure proposée. La figure 3.3 montre le schéma fonctionnel de la structure proposée.

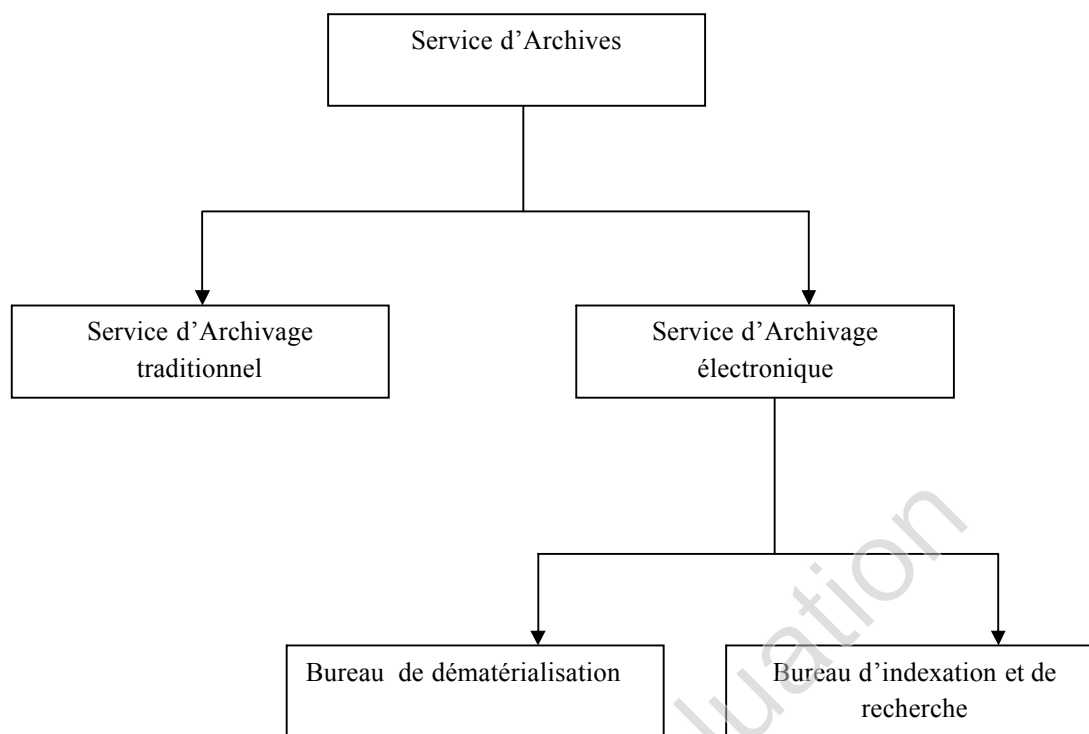


Figure 3.2 : Schéma organisationnel de la structure proposée

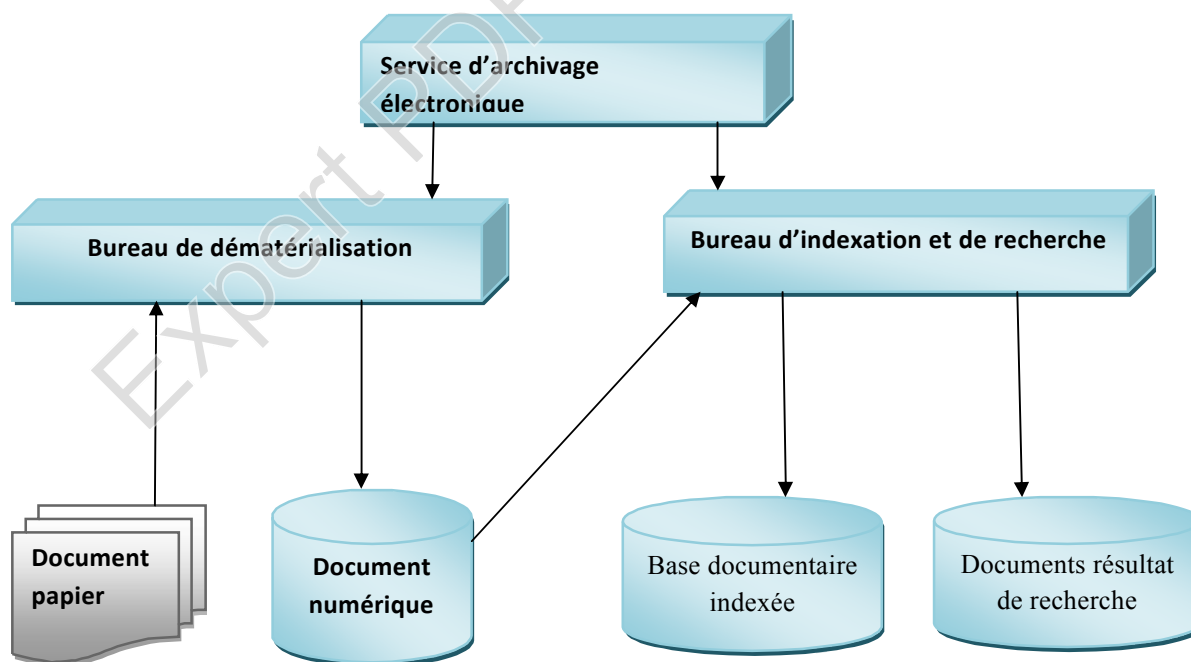


Figure 3.3 : Schéma fonctionnel de la structure proposée.

## 3.2. Description de la solution

### 3.2.1. Bureau de Dématérialisation (numérisation)

Les deux principales tâches au niveau de ce bureau sont :

- Numérisation des documents papier
- Conversion des documents numérisés au format standard adopté (dans un premier temps nous avons opté pour le format PDF).

#### 3.2.1.1. Processus de numérisation

##### a. numérisation

La numérisation de documents est une étape importante dans la mise en place d'un système d'archivage électronique. Le choix de la solution de numérisation doit prendre en compte toutes les étapes de traitement des documents depuis l'acquisition, la conversion du contenu jusqu'à la correction et la mise en exploitation du document final. Le but de cette numérisation est une utilisation du contenu converti, par exemple pour effectuer une recherche d'information.

La partie centrale de la numérisation concerne la reconnaissance de caractères et la structuration du contenu.

##### b. Chaîne de numérisation

Les principales étapes d'une chaîne de numérisation sont : [Belaïd et al, 2004]

**Acquisition** permettant la conversion du document papier sous la forme d'une image numérique (bitmap). Cette étape est importante car elle se préoccupe de la préparation des documents à saisir, du choix et du paramétrage du matériel de saisie (scanner), ainsi que du format de stockage des images.

**Prétraitement** dont le rôle est de préparer l'image du document au traitement. Les opérations de prétraitement sont relatives au redressement de l'image, à la suppression du bruit et de l'information redondante, et enfin à la sélection des zones de traitement utiles.

**Reconnaissance du contenu** qui conduit le plus souvent à la reconnaissance du texte et à l'extraction de la structure logique. Ces traitements s'accompagnent le plus souvent des opérations préparatoires de segmentation en blocs et de classification des médias (graphiques, tableaux, images, etc.).

**Correction des résultats** de la reconnaissance en vue de valider l'opération de numérisation. Cette opération peut se faire soit automatiquement par l'utilisation de dictionnaires et de méthodes de correction linguistiques, ou manuellement au travers d'interfaces dédiées.



Seules l'acquisition et la reconnaissance du contenu seront détaillées dans la suite. Le prétraitement est aujourd'hui intégré dans les OCR et est considéré comme suffisant en première approximation. Quant à la reconnaissance des résultats, c'est un problème en soi qui mériterait d'être détaillé dans un autre contexte.

### **c. Architecture d'intégration des OCR**

Quand on souhaite réaliser une plate-forme de numérisation, on doit se rendre le plus indépendant possible des logiciels d'OCR proprement dit afin essentiellement de pouvoir changer facilement de logiciel suivant les types de documents à traiter et/ou la sortie de nouveaux logiciels plus performants.

L'indépendance vis à vis des logiciels d'OCR s'obtient suivant trois critères: le format d'image en entrée, le format de sortie et l'activation du logiciel (paramétrage et appel) :

- **Format d'entrée**

Il faut ramener toutes les images au même format. On choisira en général TIFF qui est le format donné par la majorité des scanners et accepté par la plupart des OCR. Si cela n'est pas le cas, pour un OCR particulier, il faudra développer la conversion du format TIFF (**Tagged Image File Format**) est un format de fichier graphique bitmap (raster) au format demandé par l'OCR.

Cela devra être transparent dans la chaîne de numérisation; L'indépendance du format d'entrée permet d'utiliser des logiciels de prétraitement d'images, pour enlever le bruit et redresser l'image si nécessaire, qui peuvent ne pas être fournis par le logiciel d'OCR.

- **Format de sortie**

Ce format doit être normalisé pour faciliter tous les traitements ultérieurs. Si l'on souhaite juste s'occuper du texte, un format texte ISO peut suffire.

Dans le cas de documents structurés, un format plus évolué, type XML, reprenant les informations physiques du document (position du texte, police, structure des tableaux) est sans doute plus approprié. Le format de sortie servira à évaluer et à qualifier des OCR, à combiner le résultat de plusieurs OCR, à corriger les erreurs des OCR, et à effectuer des conversions vers des formats cibles plus proches de la structure logique des documents.

On voit la difficulté de la tâche s'il fallait reprogrammer tous ces traitements à chaque fois que l'on veut utiliser un nouveau logiciel d'OCR. Beaucoup de traitements se font en aval de l'OCR et nécessitent donc une indépendance vis à vis de ces logiciels.

Il y a toujours un effort de programmation pour normaliser le format de sortie de l'OCR (chaque OCR a son propre format de sortie) vers le format de sortie désiré. Cela peut se faire

soit par l'API (*Application programming interface*) de Microsoft Windows, par exemple, soit par un programme spécifique.

- **Activation et Paramétrage**

Il faut proposer une interface générale pour appeler les OCR afin de faciliter la tâche de l'opérateur et l'intégration de nouveaux OCR.

L'effort de programmation se situe dans l'adaptation entre la méthode d'appel de l'OCR et l'interface générale.

Les logiciels de SAE, qui intègrent un module d'OCR (type *Ascent* de *Kofax*), utilisent ce type d'approche indépendante de l'OCR, ce qui leur permet d'adapter leur logiciel à la demande du client avec tel ou tel OCR.

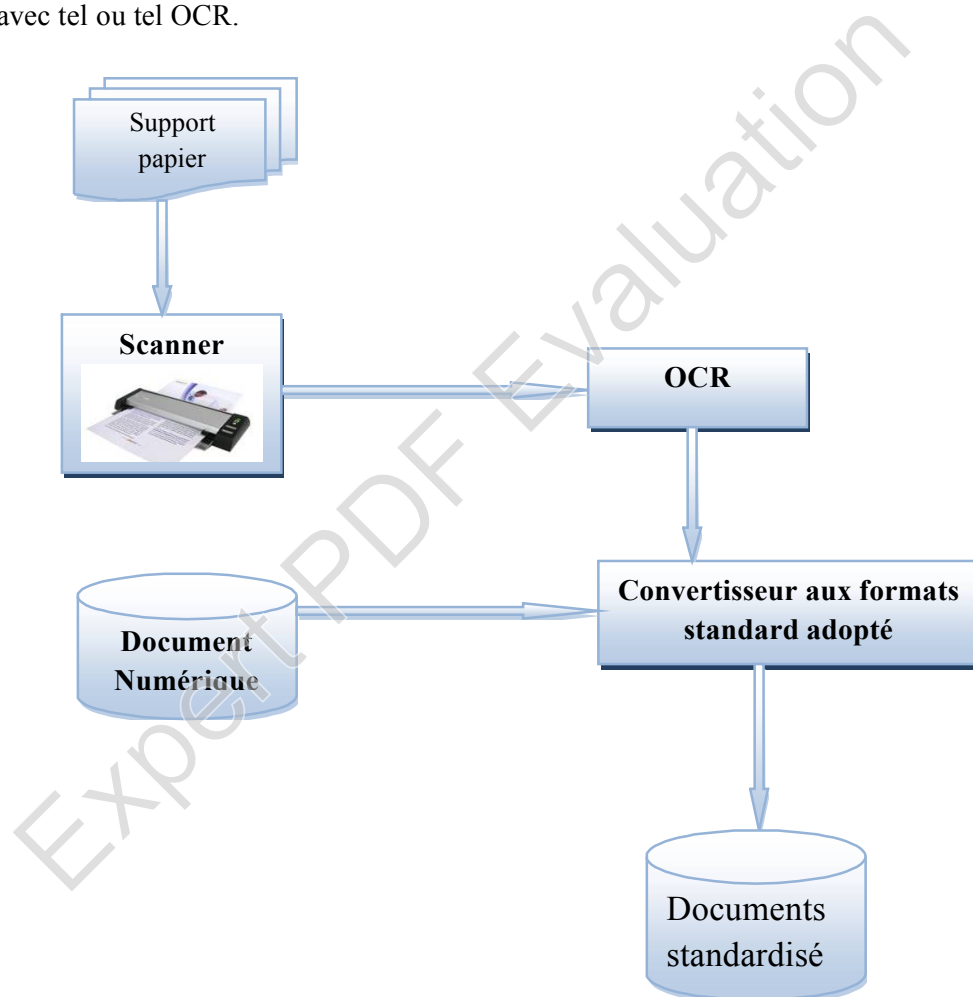


Figure 3.4 : Processus de dématérialisation

### 3.2.1.2. Solution technique proposée pour la dématérialisation

La figure 3.4 présente le processus de dématérialisation proposé. Pour le concrétiser chaque étape de ce processus, nous avons essayé de trouver des applications les plus simples d'utilisation en prenant compte, bien sûr, la qualité et les performances de ces applications.

La solution préconisée s'articule sur :

#### a. Matériel

Il existe aujourd'hui une panoplie de matériels adaptés à la numérisation de documents dans toutes ses formes avec des possibilités variées de résolution et de définition.

Concernant le matériel, le scanner est aujourd'hui l'appareil dédié à la saisie de documents. Muni de capteurs sensibles à la lumière, le scanner restitue la couleur diffusée par le document en tout point. On distingue plusieurs types de scanner adaptés aux différents usages :

- bureautique, dédiés à la saisie à plat, offrant des résolutions importantes mais sont souvent limités à des formats réduits de documents,
- de livres, dédiés à la numérisation de livres ouverts ayant des formats importants et pouvant donc s'adapter à l'inclinaison et au bombage des feuilles,
- de microfilm ou microfiche, proposant une numérisation en mode bitonal, plus rarement en niveaux de gris, s'échelonnant entre 200 et 400 ppp, d) de diapositive ou de transparent.

#### b. Logiciel de reconnaissance de caractères performants

Les techniques de lecture automatique de documents ont beaucoup évolué et mûri cette dernière décennie et l'on voit fleurir sur le marché des logiciels de moins en moins chers, de plus en plus complets et fiables offrant des solutions de numérisation très fidèles.

Ces logiciels, appartenant à la famille des OCR sont aujourd'hui capables de distinguer les différents médias dans le document (texte, graphique et photographie), d'identifier les structures linéaires et tabulaires, de faire face à une variation importante de la typographie, d'interpréter et de restituer plusieurs styles éditoriaux. Des bancs d'essais sont effectués couramment sur les dernières versions de ces logiciels montrant les nouvelles possibilités offertes et donnant une évaluation correcte et précise de leur capacité de reconnaissance en termes de taux de confiance, de précision et de vitesse d'exécution, par type de document et de typographie utilisée.

Pour notre solution nous avons choisi un logiciel, qui à notre avis et selon des statistiques, a fait ses épreuves dans le domaine à savoir *ABBYY FineReader*.

### **b.1. Qu'est-ce qu'*ABBYY FineReader*?**

*ABBYY FineReader* est un système de reconnaissance optique de caractères (OCR) conçu par ABBYY Software [ABBYY est une société internationale avec plus de 900 employés à travers le monde (de Janvier 2010)]. Il permet de convertir des documents numérisés, des documents PDF et des fichiers d'images, y compris les photos numériques, en formats modifiables. [Web2. 11]

### **b.2. Avantages d'*ABBYY FineReader***

- **Reconnaissance rapide et précise**

Le système OCR utilisé par *ABBYY FineReader* offre aux utilisateurs une reconnaissance rapide et précise et préserve le format source de tout document (y compris le texte sur images d'arrière plan, le texte en couleur sur des arrières plans colorés et le texte autour d'une image, etc.).

Grâce à la technologie de reconnaissance de document adaptative d'*ABBYY (ADRT®)*, *ABBYY FineReader* peut analyser et traiter un document en entier, au lieu de procéder page par page. Cette approche conserve la structure du document source, y compris le format, les hyperliens, les courriels, les en-têtes et les pieds de page, les légendes d'images et de tableaux, les numéros de page et les notes de bas page.

*ABBYY FineReader* peut reconnaître des documents écrits dans une ou plusieurs des 186 langues prises en charge, y compris le coréen, le chinois, le japonais, le thaï et l'hébreu. En ce qui concerne la langue arabe, cette dernière n'est pas encore prise en charge. *ABBYY FineReader* propose aussi une détection automatique des langues du document traité.

De plus, *ABBYY FineReader* est quasiment infaillible quant aux défauts d'impression et peut reconnaître des textes imprimés dans pratiquement toutes les polices.

Ce programme contient également une large gamme d'options pour les données de sortie : les documents peuvent être enregistrés sous une grande variété de formats, envoyés par courriel ou transférés vers d'autres applications pour un traitement complémentaire.

- **Simplicité d'utilisation**

L'interface simple et intuitive d'*ABBYY FineReader* pilotée par les résultats permet d'utiliser le programme sans formation complémentaire. Les nouveaux utilisateurs peuvent maîtriser les fonctions principales rapidement. Les utilisateurs peuvent modifier la langue de l'interface directement dans le programme.

Les tâches rapides proposent une liste des tâches utilisateur les plus fréquemment utilisées pour la conversion des documents numérisés, des PDF et fichiers d'images en formats modifiables et permet de récupérer un document électronique en un clic de souris.

L'intégration transparente à Microsoft Office et à l'explorateur Windows permet de reconnaître des documents directement à partir de Microsoft Outlook, Microsoft Word, Microsoft Excel et de l'explorateur Windows.

*ABBYY FineReader* contient également un menu d'aide avec des exemples qui illustrent comment utiliser le programme pour exécuter un certain nombre de tâches complexes relatives à la conversion.

### **b.3. Interface *ABBYY FineReader 10***

*ABBYY FineReader* possède une interface pilotée par les résultats, conviviale et intuitive qui permet d'utiliser l'application sans formation complémentaire. Les fonctions principales sont maîtrisées par les nouveaux utilisateurs rapidement. En plus son interface peut être personnalisée.

Nous pouvons personnaliser notre espace de travail pour répondre aux besoins particuliers, ce qui est très important dans le cas du centre régional des archives. Par exemple nous pouvons :

- Modifier la position et la taille des fenêtres
- Définir la barre d'accès rapide pour l'accès aux commandes les plus fréquemment utilisées
- Définir les raccourcis clavier ; nous pouvons modifier les combinaisons de touches prédéfinies et ajouter nos propres combinaisons pour exécuter des commandes spécifiques du programme.
- Sélectionner la langue d'interface favorite, etc.

Au démarrage d'*ABBYY FineReader* affiche la fenêtre Nouvelle tâche (figure 3.5), qui s'ouvre par défaut après le chargement de l'application.



Figure 3.5 : Tâches rapides *ABBYY FineReader*

La figure 3.6 présente la fenêtre principale d'ABBYY FineReader qui affiche le document ABBYY FineReader actuellement ouvert.

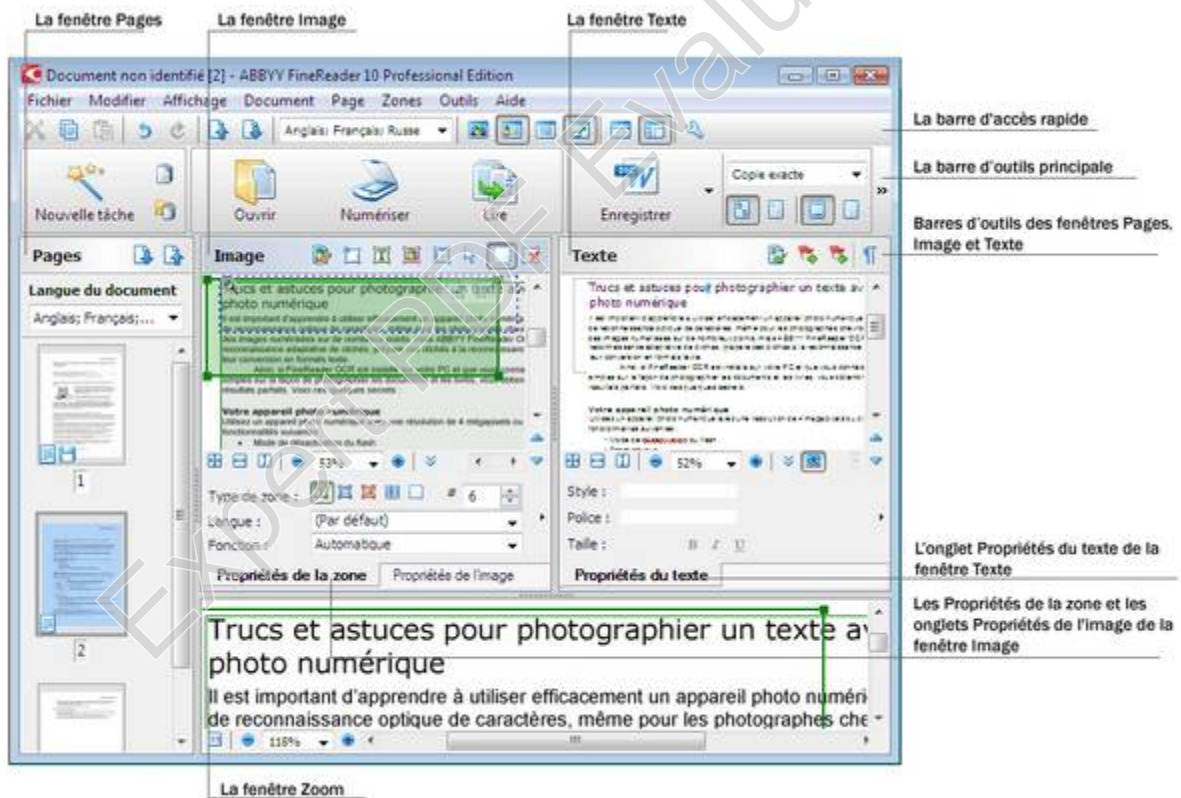


Figure 3.6 : La fenêtre principale de ABBYY FineReader

#### **b.4. Tâches rapides ABBYY FineReader**

Le traitement de documents avec ABBYY FineReader implique souvent la même séquence d'exécution, à savoir numérisation et reconnaissance d'un document suivis de l'enregistrement des résultats dans un format en particulier. Afin d'exécuter les tâches les plus fréquemment utilisées, ABBYY FineReader inclut l'option Tâches rapides afin d'obtenir un texte reconnu en un seul clic.

#### **3.2.2. Bureau d'indexation et de recherche**

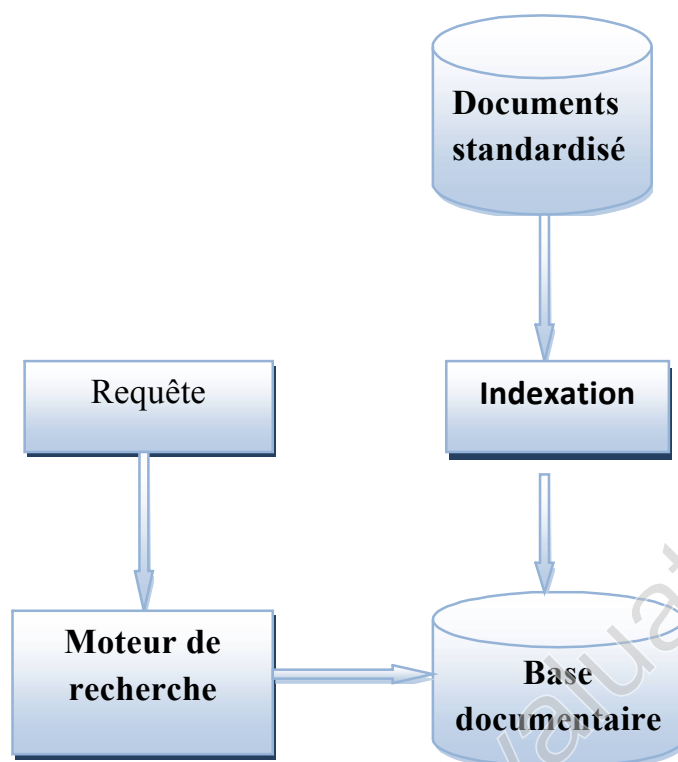
##### **3.2.2.1. La recherche d'information**

La Recherche d'Information (RI) est l'ensemble des techniques permettant de sélectionner à partir d'une collection de documents, ceux qui sont susceptibles de répondre au besoin de l'utilisateur exprimé via une requête [dahak, 2006b] .

Les systèmes de recherche d'information ont pour fonction de permettre à l'utilisateur d'accéder à des documents qui contribuent à résoudre le problème d'information qui motive sa recherche. Ainsi le système peut être vu comme un instrument de prédiction de la pertinence des documents du corpus pour l'utilisateur. Les indices traditionnellement utilisés par le système pour estimer cette pertinence sont de nature thématique, et sont fournis par l'utilisateur sous la forme d'un ensemble de mots-clés : la requête. Ainsi le système implémente une fonction de correspondance entre document et requête qui modélise la dimension thématique de la pertinence.

##### **3.2.2.2. Processus de recherche d'information**

Un Système de Recherche d'Information (SRI) intègre trois fonctions principales. La figure 3.7 illustre l'architecture générale d'un système de recherche d'information. D'un côté, on a l'information accessible dans le système c'est-à-dire la collection de documents. D'un autre côté, on dispose du besoin en information exprimé par l'utilisateur sous forme de requête. Ensuite, afin d'être exploitables par le système, l'information aussi bien que le besoin en information passent par des étapes de traitement en vue de sélectionner des informations pertinentes en réponses à la requête utilisateur. Il s'agit principalement du processus de *représentation* et de celui de *recherche*.



**Figure 3.7: Processus général de recherche d'information**

Le processus de représentation a pour rôle d'extraire d'un document ou d'une requête, une représentation paramétrée qui couvre au mieux son contenu sémantique. C'est le processus d'indexation.

Le processus de recherche représente le processus du noyau d'un SRI. Il comprend la fonction de décision fondamentale qui permet d'associer à une requête, l'ensemble des documents pertinents à restituer.

#### **a. Processus d'Indexation**

L'indexation est une étape cruciale, elle s'effectue en amont de l'étape de recherche effective de l'information. Son but est d'assurer la recherche dans des conditions acceptables en termes de coût et d'efficacité. L'indexation consiste à analyser le document lors de l'organisation du fond documentaire afin de produire un ensemble de mots clés, appelés aussi descripteurs utilisés ultérieurement dans le processus de recherche. [ouarzeddini, 2008]



### **a.1. Modes d'indexation**

L'indexation peut être effectuée selon trois modes différents : **[Hadj henni, 2009]**

#### **Mode manuel**

Chaque document est analysé par un spécialiste du domaine ou un documentaliste. Ce mode garantit une bonne précision mais exige un gros effort intellectuel.

L'indexation manuelle est caractérisée par un haut degré de subjectivité du fait qu'elle soit réalisée par un être humain. En conséquence, pour un même document, des termes différents peuvent être affectés par des indexeurs différents.

#### **Mode automatique**

A l'aide d'un processus entièrement informatisé. C'est le mode le plus utilisé, il comprend les étapes suivantes : extraction automatique des descripteurs, utilisation d'un anti-dictionnaire pour éliminer les mots outils, lemmatisation, repérage de groupes de mots, pondération des mots et enfin, création de l'index.

L'indexation automatique repose sur deux idées très simples, la première consiste à considérer les mots qui apparaissent dans les granules et compter leurs occurrences.

La seconde stipule qu'il est important de prendre en compte la distribution de ces mots dans les documents car les fréquences d'apparition des co-occurents permettent de déterminer les thématiques et leur signification dans les textes.

#### **Mode semi-automatique**

Dans ce mode, un premier processus automatique permet d'extraire les termes du document. Cependant, le choix final reste au spécialiste du domaine ou au documentaliste pour établir les relations entre les mots clés et choisir les termes significatifs.

Quelque soit le mode utilisé, le résultat de l'indexation est un ensemble de termes définissant ce qui est appelé le langage d'indexation.

### **a.2. Types de langages d'indexation**

On peut distinguer deux types de langage d'indexation : le langage contrôlé et le langage libre **[Hadj henni, 2009]**.

#### **Langage contrôlé**

Il s'agit d'un lexique figé de descripteurs. L'indexation est alors le plus souvent manuelle, parfois semi-automatique. Un professionnel choisit un ou plusieurs descripteurs, à partir d'un thésaurus ou de bases terminologiques, pour représenter le document.

### **Langage libre**

Les descripteurs sont extraits automatiquement des documents, ou de la requête de l'utilisateur. Il faut souligner que ces deux types de langage d'indexation peuvent coexister dans certains SRI.

#### **a.3. Pondération des termes**

La pondération est l'une des fonctions fondamentales en RI. Elle repose sur l'idée que le poids d'un terme dans un document traduit son importance dans ce document. Les méthodes statistiques furent exploitées dès le début de la RI.

La plupart des formules de pondération, proposées dans la littérature, se basent sur deux facteurs : la pondération locale et la pondération globale. La première quantifie la représentativité locale d'un terme dans le document (*Term Frequency Tf*), et la deuxième quantifie la représentativité du terme vis-à-vis de la collection des documents (*Idf* : *Inverse of Document Frequency*). [SAID L'HADJ, 2009]

**Term Frequency (Tf)** : cette mesure indique que l'importance du terme dans le document est proportionnelle à la fréquence du terme dans le même document.

**Inverse of Document Frequency (Idf)** : cette mesure calcule l'importance d'un terme dans toute la collection. Elle traduit l'impact d'un terme selon son nombre d'apparition dans la base documentaire.

La combinaison des deux mesures ( $Tf * Idf$ ) donne une bonne approximation de l'importance du terme dans le document, particulièrement dans les corpus de documents de tailles homogènes.

La construction de l'index ainsi que la pondération sont au cœur du problème de la recherche d'information. Elles résultent d'une dualité entre deux objectifs contradictoires qui sont *la spécificité* et *l'exhaustivité*. Le premier, tend à ce que la description du document soit la plus complète possible, alors que le deuxième consiste en une meilleure différenciation (discrimination) entre les descriptions des documents pour mieux les distinguer.

Une bonne méthode d'indexation doit établir *un compromis* entre ces deux aspects.

Cet équilibre se retrouve notamment dans le schéma de pondération  $tf*idf$ .

#### **a.4. Extraction des descripteurs**

Les descripteurs représentent l'information atomique d'un index. Ils sont censés représenter le contenu du document. Ils sont aussi désignés par "unités élémentaires" (tokens). Le but de l'extraction des descripteurs est de les choisir de manière à ce que l'index perde le moins d'information sémantique possible [yahiaoui, 2008b]

A travers les travaux de la RI, nous pouvons recenser les différents types de descripteurs utilisés [Ou Rabah. 2008b] :

- Les *mots simples* non vides.
- Les *lemmes* ou les *racines* des mots simples,
- Les *n*-grammes qui sont une représentation originale d'un texte en séquence de N caractères consécutifs.
- Les *groupes de mots* : un groupe de mots est souvent plus riche sémantiquement que les mots qui le composent pris isolément. Cet argument a conduit à considérer les groupes de mots comme unités de base dans le langage d'indexation.
- Les *contextes* c'est-à-dire les termes n'apparaissant pas explicitement dans le texte du document mais ayant un lien sémantique et/ou de cooccurrence avec les mots du document.
- Les *concepts* qui sont des expressions contenant un ou plusieurs mots. Ils peuvent être écrits de manière libre par l'utilisateur ou, ce qui est souvent le cas, choisis dans une liste de concepts (vocabulaire contrôlé : thésaurus, une ontologie, une hiérarchie de concepts, etc).

L'ajout de termes reliés sémantiquement aux concepts décrits dans le document est déterminé par un calcul de proximité ou de similarité sémantique [SAID L'HADJ, 2009] Ceci est rendu possible notamment grâce à la disponibilité de plus en plus croissante de ressources sémantiques.

#### **a.5. Approches d'extraction des groupes de mots**

##### **Approches statistiques**

Essentiellement basées sur les cooccurrences. Les groupes de mots sont extraits en utilisant la cooccurrence, en partant de l'hypothèse que des termes (souvent réduits à deux ou trois mots) qui apparaissent ensemble dans le texte peuvent représenter un concept. De la même manière que pour les mots simples, le principe de pondération locale et globale ( $tf*idf$ ) est appliqué pour éliminer les "groupes de mots outils" car, ils ne contribuent pas à la discrimination des documents.

### **Approches linguistiques**

Ces approches se basent sur une analyse syntaxique partielle ou sur l'utilisation de patrons syntaxiques pour détecter les termes composés. Le plus souvent, un ensemble de patrons syntaxiques comme (NOM NOM) ou (NOM PREP NOM) est utilisé pour l'identification des groupes de mots.

### **Approches mixtes**

Ces approches combinent les deux informations (syntaxe et cooccurrence) pour améliorer la précision dans la détection des termes composés.

Malgré les nombreuses études consacrées à l'extraction des groupes de mots, il n'existe pas encore, d'après une méthode effective qui permet de distinguer les termes des non termes d'un point de vue syntaxique.

#### **b. Reformulation de Requêtes**

Parfois l'utilisateur est confronté à une situation difficile, il est incapable de trouver les mots précis pour exprimer son besoin en information. Alors, certains pour ne pas dire la majorité des documents qui lui sont retournés l'intéressent moins que d'autres.

Outre cela, certaines requêtes sont courtes, du coup, elles ne sont pas sémantiquement riches, pour que le SRI puisse retourner des documents pertinents. Afin de pallier ces problèmes, les chercheurs en RI se sont orientés vers l'intégration d'une étape supplémentaire dans le processus de recherche : la reformulation ou expansion de requêtes.

Elle consiste à modifier la requête initiale de l'utilisateur par l'ajout de termes significatifs et/ou la ré-estimation de leur poids.

Si les termes rajoutés proviennent des documents de la collection, on parle de réinjection de pertinence (relevance feedback). En revanche, s'ils sont issus d'une ressource conceptuelle externe (ontologie, thésaurus ou dictionnaire) on parle, dans ce cas de reformulation de requêtes directe. Donnent un état de l'art complet de l'expansion des requêtes ainsi que des différentes approches suivies dans cet axe.

#### **c. Appariement document-requête**

A la réception d'une requête, le SRI lui crée une représentation similaire à celles des documents, et calcule un score de correspondance, entre chacune de ces représentation set celle de la requête, à partir d'une fonction de similarité (ou de probabilité) RSV (Q,d) (*Retrieval Status Value*) où Q est une requête et D un document de la base. Ce score traduit la pertinence du document par rapport à la requête, autrement dit la pertinence système. Ce processus est appelé alors Appariement Document-Requête.

Le processus d'appariement est intimement lié au processus d'indexation et de pondération des termes des requêtes et des documents. En effet, leurs représentations ainsi que leur appariement permettent de définir un modèle de recherche d'information.

### **3.2.2.3. Solution technique proposée pour l'indexation et la recherche de l'information**

#### **a. Matériel**

Le sous équipement que nous avons constaté lors de notre stage au niveau du centre régional des archives nous oblige à dire qu'une acquisition d'équipement informatique évolué est incontournable. En effet, l'évolution de la base documentaire et son traitement nécessiteront un matériel puissant et de marque.

#### **b. Logiciel**

On propose un logiciel efficace d'indexation et de recherche qui s'appelle *DocFetcher*.

*DocFetcher* est un logiciel libre de recherche de bureau ; [Web1. 11] il permet de rechercher des informations à l'intérieur des documents enregistrés sur une base documentaire. Autrement dit, c'est un moteur de recherche comme Google, mais pour une base documentaire locale. Le logiciel est actuellement disponible pour Windows et Linux.

#### **b.1. Fonctionnement de DocFetcher**

*DocFetcher* ; Figure 3.8, crée des index à partir des fichiers. La recherche s'effectue ensuite en une fraction de seconde à partir de ces index. Nous pouvons soit créer des index permanents pour les dossiers comprenant un grand nombre de fichiers rarement modifiés ou des index temporaires pour les petits dossiers qui contiennent des fichiers qui changent souvent.

Les index temporaires peuvent être créés en cliquant droit sur un dossier et en choisissant "Rechercher avec *DocFetcher*". Ils sont automatiquement supprimés lorsque le programme est fermé. Nous pouvons aussi demander à *DocFetcher* de conserver ces index.

Le processus d'indexation peut prendre quelques minutes pour grande base documentaire. Sur un ordinateur moyen, 200 documents peuvent être indexés par minute.

Une fois que l'index est créé, nous pouvons taper des mots clés dans le champ de recherche de *DocFetcher*, par exemple "Wilaya de Guelma", puis taper sur la touche Entrée. *DocFetcher* affiche alors tous les documents dans la base documentaire qui contiennent les mots recherchés, le plus souvent en moins d'une seconde.

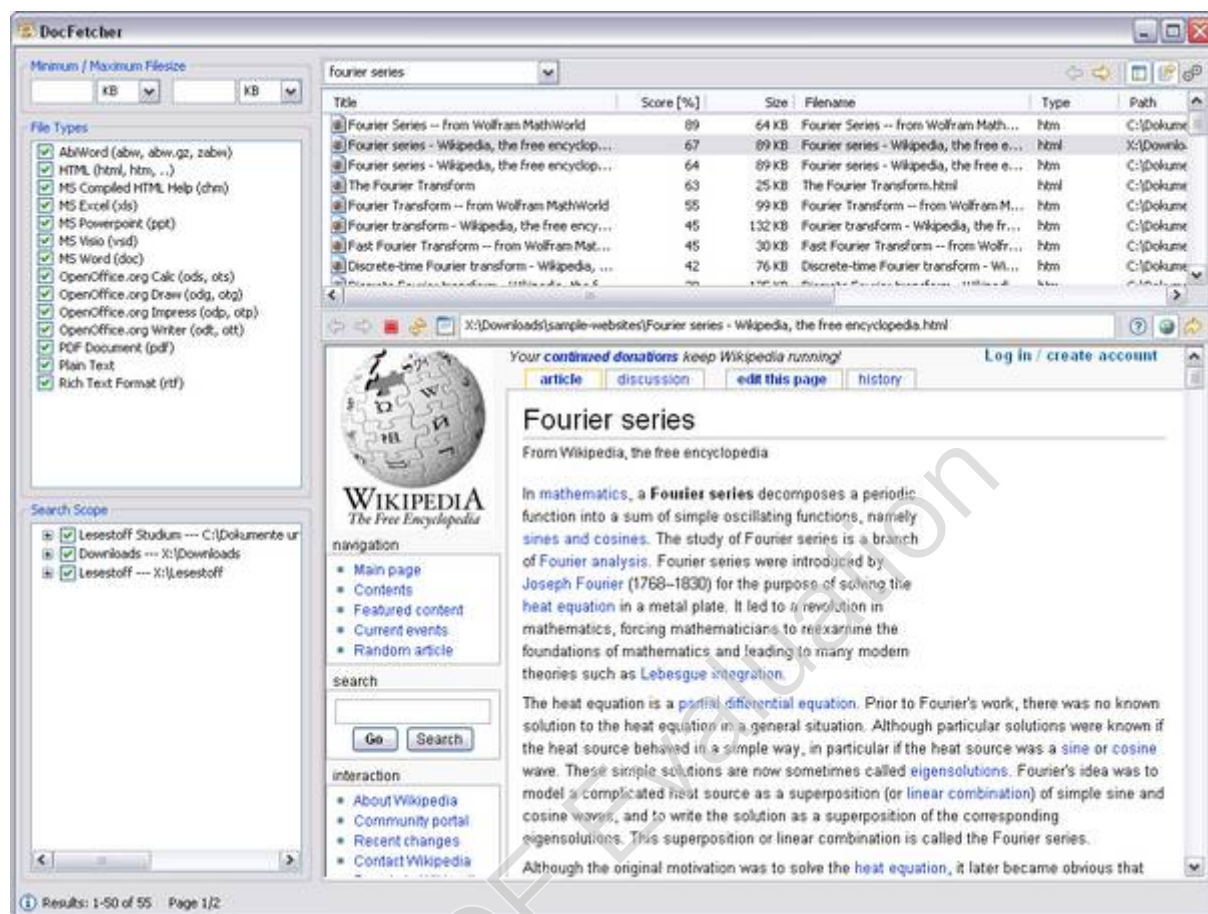


Figure 3.8 : Interface principal de DocFetcher

## b.2. Autres fonctions de DocFetcher

### Mise à jour automatique des index

Les index sont mis à jours automatiquement lorsque les fichiers sont modifiés, même lorsque DocFetcher n'est pas en fonctionnement. Cela est réalisé grâce à un démon (service) qui surveille les modifications en arrière-plan. Ce démon utilise le CPU de manière minimale : il ne réindexe pas les fichiers modifiés, mais signale simplement à DocFetcher les fichiers qui devront être réindexés la prochaine fois que le logiciel sera utilisé.

### Une version portable

La version portable, figure 3.9, fonctionne à la fois sur Windows et Linux. Nous pouvons placer les documents à indexer dans un dossier et ensuite déplacer librement le dossier complet (c'est-à-dire DocFetcher + index + documents) vers un autre ordinateur, un volume crypté (TrueCrypt ou FreeOTFE), un lecteur de CD-Rom ou un lecteur USB. La version

portable peut même être utilisée pour partager un dossier de documents indexés sur un réseau local ou entre plusieurs systèmes d'exploitation sur un ordinateur multi-boot.

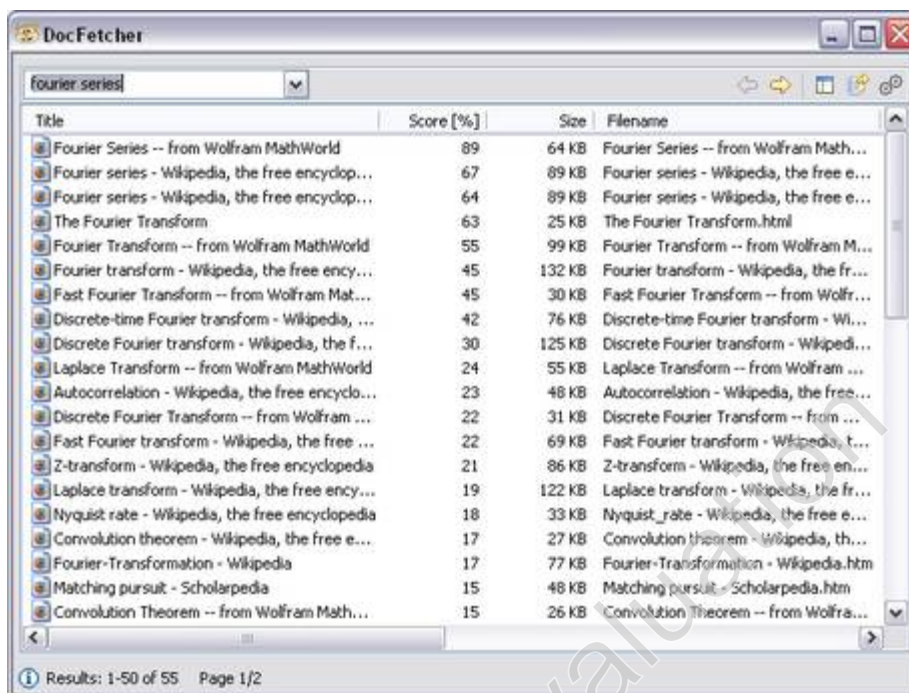


Figure 3.9: DocFetcher en mode minimaliste

### Recherche dans le code source des fichiers

Les extensions de fichiers reconnues par DocFetcher peuvent être personnalisées. Nous pouvons donc utiliser le logiciel pour toutes sortes de fichiers texte.

En plus sur le panneau de prévisualisation il y a un surlignage des mots cherchés et navigateur web intégré. Les résultats de recherche pouvant être triés et filtrés selon différents critères (types de fichier, taille, chemin, etc.).

## Conclusion Générale

Avant de commencer ce mémoire, le domaine des archives et l'archivage électronique on particulier était inconnu pour nous ainsi une premier étape de recherche documentaire et d'appropriation du sujet a été nécessaire.

Ce stage a été pour nous une expérience exceptionnelle tant au plan professionnel que personnel. Nous avons fait la connaissance de personnes accueillantes, disponibles et généreuses qui ont fait de ce stage une grande réussite pour nous.

Nous avons pu apercevoir, au cours de notre étude, que très peu d'éditeurs de logiciel, proposent des outils liés à la gestion d'archives physiques tout en associant les archives électroniques.

Cette étude nous a amenée aux conclusions suivantes: pour réussir une première mise en place d'un système d'archivage électronique, dans ce cas présent, d'une solution de gestion d'archives physiques, il faut tout d'abord que le service archives en sein de centre régional d'archive soit opérationnel et qu'il soit doté d'une bonne organisation (en prenant en compte l'organisation que nous avons proposé). La mise en place d'un outil perfectionné, pour être réussie, doit être associée à des pratiques d'archivage bien pensées, à une collaboration entre le service Archives et Informatique

Enfin l'archivage électronique présente de nombreux avantages mais s'avère difficile à réaliser. Il faut donc envisager la possibilité de faire appel à un prestataire externe spécialisé qui connaît les contraintes techniques et les normes en vigueur. Dès lors, il est indispensable de porter une attention toute particulière à la rédaction du contrat qui lie le client avec le tiers archive pour éviter toute déconvenue. Il convient de prévoir notamment le niveau de services souhaité et notamment les conséquences d'une perte des archives.



## Bibliographie

**[AAF, 2005a]**

Groupe métier AAF - ADBS « Records management ». Comprendre et Pratiquer le records management. Version 2.2005.p.17.

**[AAF, 2005b]**

Groupe métier AAF - ADBS « Records management ». Comprendre et Pratiquer le records management. Version 2.2005.p.11.

**[abbasi et al, 2009]**

Rafika abbasi (Tunisie) .Christine ABOMO AKONO (Cameroun) .Sif Daoula , BOUZIDI (Algérie) .Lucienne DINOMAS (France) .Stéphane LONGUET (France) ,Violeta RAKIPI (Albanie) .Danilo SARENAC (Serbie). Stage technique international, des archives.2009.

**[Agnieszka, Bonard, 2010a]**

Agnieszka Tona, responsable universitaire. Anne Emmanuelle Bonard, responsable professionnelle. Mémoire Master 2.Comment choisir une solution de gestion d'archives .Université Charles De Gaulle .Lille. 2010. P.21.

**[Agnieszka, Bonard, 2010b]**

Agnieszka Tona, responsable universitaire. Anne Emmanuelle Bonard, responsable professionnelle. Mémoire Master 2.Comment choisir une solution de gestion d'archives .Université Charles De Gaulle Lille. 2010p.24.

**[Agnieszka , Bonard, 2010c]**

Agnieszka Tona, responsable universitaire. Anne Emmanuelle Bonard, responsable professionnelle. Mémoire Master 2. Comment choisir une solution de gestion d'archives Université Charles De Gaulle Lille. 2010. P.12.

**[ait ali yahia, 2008]**

Dahbia ait ali yahia. «Une mémoire documentaire a base de XML». Mémoire de magister. Alger.2007/2008. P. 32 .

**[Aliouali, 2006a]**

Aliouali Nadia. «Problématique de la préservation des documents numériques : Cas du dépôt légal». Mémoire de Magister. Alger. 2005/2006. P.19.

**[Aliouali, 2006b]**

Aliouali Nadia. «Problématique de la préservation des documents numériques: Cas du dépôt légal». Mémoire de Magister. Institut National de Formation en Informatique 2005/2006.p.16

**[Belaïd et al , 2004]**

Abdel Belaïd (LORIA, UMR 7503 du CNRS, Nancy). Hubert Emptoz (RFV-ISDN, INSA Lyon). Georges Vignaux (UPR 36 du CNRS, Paris). RTP 33 : « Document et contenu : création, indexation, navigation» Action spécifique 96 : « Numérisation et valorisation des collections ». Février 2004.p.5.

**[Chabin, 1999] :**

Chabin Marie-Anne. Je pense donc j'archive: l'archive dans la société de l'information. [en ligne].1999.p.24.consulter sur <<http://books.google.fr/books?id=ZgPXPkD0o5MC&pg>>

**[Chabin, 2006] :**

Marie-Anne, «Archive 17. CSTIC –Afnor –15/09/06.Extrait de MoReq».p.7 .

**[Chabin, 2007a] :**

Chabin Marie-Anne. «Archiver, et après ?» Paris : Djakarta, 2007. p.66-69.

**[Chabin, 2007b] :**

Chabin .Marie-Anne. «Records management et gestion des archives ISO 15489 et MoReq». 2007. p.6.

**[Chabin, 2008a] :**

Chabin .Marie-Anne. «Exigences types pour la maîtrise de l'archivage électronique». Mise à jour et extension 2008.p.12.

**[Chabin, 2008b]**

Chabin .Marie-Anne. «Exigences types pour la maîtrise de l'archivage électronique» Mise à jour et extension 2008 – Spécifications MoReq2. P. 16

**[Chabin. 2010a] :**

Chabin .Marie-Anne. «Nouveau glossaire de l'archivage». février 2010. p.4.

**[Chabin, 2010b] :**

Chabin .Marie-Anne. «Nouveau glossaire de l'archivage». février 2010. p.8.

**[Chabin. 2010c] :**

CHABIN .Marie-Anne. «Nouveau glossaire de l'archivage». 2010 .p. 17.

**[CIA, 1999]**

Conseil international des archives. ISAD(G):«Norme générale et internationale de description archivistique». Adoptée par Comité sur les normes de description Stockholm, Suède, 19-22 septembre 1999.

**[Cod civil, 2007]:**

Cod civil. Art 223 ter .2007. p. 50.

**[dahak, 2006a]**

Fouad dahak. «Indexation des documents Semi-Structurés». Mémoire de magister. Alger. 2005/2006. p.7.

**[dahak, 2006b]**

Fouad dahak. «Indexation des documents Semi-Structurés». Mémoire de Magister.alger.2005/2006. p.7.

**[delmas, 2009]**

Bruno delmas .professeur d'archivistique. Cours archivistique enligne.sur <http://www.piaf-archives.org/espace-formation/mod/resource/view.php?r=81>.  
31/07/09

**[Delsalle, 1998]**

Paul Delsalle. Une histoire de l'archivistique.1998.p.212.enligne sur [<http://books.google.fr/books?id=>](http://books.google.fr/books?id=>)

**[Dronne, 2006a]**

Marie Mathias Dronne. «Conduire un projet d'archivage électronique : réflexions et mises en œuvre». INTD. Le 25 octobre 2006.p.17.

**[Dronne, 2006b]**

Marie Mathias Dronne. «Conduire un projet d'archivage électronique : réflexions et mises en œuvre» .INTD.2006.p.17.

**[Dronne, 2006c]**

Marie Mathias Dronne. «Conduire un projet d'archivage électronique : réflexions et mises en œuvre» .INTD.2006.p.22.

**[Dronne, 2006d]**

Marie Mathias Dronne. «Conduire un projet d'archivage électronique : Réflexions et mises en œuvre».INTD. 2006. p 27.

**[Ducharme, 2011]**

Daniel Ducharme. Introduction à l'archivistique. Leçon 2.1.2011. p3-7.

**[Gagnon-Arguin et al. 2003]**

Louise Gagnon-Arguin, Jacques Grimard, Robert Garon. La gestion d'un centre d'archives. 2003. p.77. En ligne sur <http://books.google.fr/books?id=>>.

**[Hadj henni, 2009]:**

Hadj henni M'hamed. Approche ontologique pour la modélisation sémantique, l'indexation et l'interrogation des documents Coraniques. Mémoire de magister. Alger. 2009. p.27-29.

**[Hercek, 2010]**

Serge Hercek de STS Benelux. Archivage électronique. 2010. P.5.

**[jora, 1988a]**

Journal officiel .27 janvier 1988.p.100.la loi n°88-09 .art.2

**[jora, 1988b]**

journal officiel .27 janvier 1988.p.100.la loi n°88-09 .art.3

**[Livre blanc, 2011] :**

Livre blanc Version 1.44 janvier 2011 . La maîtrise du flux documentaire .p.11.

**[micaelli, 2010]**

Eric micaelli. «Système d'archivage électronique (SAE)» des exigences et des spécifications qui tiennent compte d'un environnement [GED-SAE(PAE)-PGA] . 1er octobre 2010 .p6-9.

**[Midi-Pyrénées, 2008]**

Midi-Pyrénées. «L'archivage électronique des documents à valeur probante dans les collectivités territoriales». Septembre 2008.p.4.

**[Mouette, 2007a]**

Solène Mouette. La place et le rôle du Records management dans les besoins d'archivage d'un établissement public 23 octobre 2007.p18/19.

**[Mouette, 2007b]**

Solène Mouette. La place et le rôle du Records management dans les besoins d'archivage d'un établissement public : le cas de l'Institut National de la Consommation. le 23 octobre 2007.p. 13.

**[mouhoubi, 2002a]**

zahri mouhoubi. Interface généralisé EAD-SEDA.2010.p.13.

**[mouhoubi, 2002b]**

zahri mouhoubi. Interface généralisé EAD-SEDA.2010.p.18.

**[Müller, 2011]**

Bertrand Müller. «Historiographie & archivistique». Université de Genève. 12 janvier 2011 .En ligne consulter le 28/4/2011 sur <[http://tristan.ubourgogne.fr/umr5605/publications/historiographie/B\\_Muller.html](http://tristan.ubourgogne.fr/umr5605/publications/historiographie/B_Muller.html) >.

**[Ou Rabah. 2008a]**

soualah Mohammed Ou Rabah. Numérisation des manuscrits arabes : Catalogage et accès multilingue. Mémoire de magister. Alger. 2007/2008.p.64.

**[Ou Rabah. 2008b]**

soualah Mohammed Ou Rabah. Numérisation des manuscrits arabes : Catalogage et accès multilingue. Mémoire de Magister.alger. 2007/2008P59

**[ouarzeddini, 2008a]:**

El Hadi ouarzeddini. «Ergonomie et interface des bibliothèques numérique». Mémoire de magister. janvier 2008. p.33.

**[ouarzeddini, 2008b]:**

El Hadi ouarzeddini. «Ergonomie et interface des bibliothèques numérique». Mémoire magisterJanvier .2008.P39.

**[ponsot, 1995]**

Dominique ponsot. «Valeur juridique des documents conservés sur support photographique ou numérique». Septembre 1995.p.13.

**[records, 2008] :**

Records management. FICHE 3 | NORME ISO/IEC 15489.2008.

**[rietsch et al, 2006a]**

Jean-Marc rietsch. Marie-Anne chabin. Eric caprioli. «L'archivage a l'usage du dirigeant».2006. P.16-17.

**[rietsch et al, 2006b]**

Jean-Marc rietsch. Marie-Anne chabin. Eric caprioli. «L'archivage a l'usage du dirigeant».2006. p.8.

**[SAID L'HADJ, 2009]:**

Said l'hadj Lynda. Recherche Conceptuelle d'Information Modèle d'Indexation Mixte : concepts-mots. Mémoire de magister .alger INI .2008/2009.p.11-12.

**[Sédallian, 2002]**

Valérie Sédallian. L'archivage de l'acte électronique. Juriscom.net, 8 juillet 2002, <<http://www.juriscom.net>>.

**[sellal, 2007] :**

sellal Achour. Les Glissements sémantiques dans la terminologie archivistique. RIST Vol. 17 N° 1-2 Année 2007. P.37.

**[STS, 2009] :**

STS Group 2009. Livre blanc. Archivage et valeur probatoire .p.5.

**[yahiaoui, 2008a]**

Samir yahiaoui. Structuration de métadonnées en vue d'un filtrage d'information sur le Web .mémoire de magister. Alger. 2007/2008.p3.

**[yahiaoui, 2008b]**

Samir yahiaoui. Structuration de métadonnées en vue d'un filtrage d'information sur le Web. de Magister.alger. 2007/2008. P.81.

## **Web graphie**

1. **[Web1. 11]:** <http://docfetcher.sourceforge.net/fr/index.html> 15/5/2011
2. **[Web2. 11] :** <http://france.abbyy.com> 15/5/2011

Expert PDF Evaluation