
Module 6 Technologies et gestion automatisée des archives

Objectifs

Objectif général

Le présent module vise à rendre le participant capable de comprendre, de façon générale, les particularités des différentes technologies ainsi que leurs applications à la gestion des archives.

Objectifs pédagogiques

Le participant sera en mesure de :

- de distinguer les principales caractéristiques des différentes technologies généralement utilisées dans les organisations contemporaines;
- de reconnaître l'impact des différentes technologies sur les archives;
- d'identifier les différentes possibilités de gestion des archives par des logiciels dédiés aux différentes fonctions archivistiques;
- de définir quelques problèmes simples reliés aux archives électroniques.

Introduction

L'ordinateur est de plus en plus présent au cœur de notre vie quotidienne. Il est à la maison, il est au travail, il est même dans notre véhicule automobile et dans nos appareils électroménagers, bref il est partout et parfois même à des endroits où l'on ne soupçonne pas sa présence. Cette omniprésence de l'ordinateur a considérablement modifié nos habitudes et certains aspects liés à notre environnement. Un de ces aspects est celui de changer la représentation que l'homme a des informations qu'il reçoit de cet environnement. En effet, nombre des informations que nous recevons sont de type analogique. Elles sont perçues de façon continue : intensité lumineuse, position d'un objet, hauteur d'un son et elles sont transmises sous forme d'une modulation, d'un courant électrique ou d'une réaction chimique. Ces mêmes informations, si elles doivent être traitées ou analysées par un ordinateur, doivent être transformées sous forme numérique c'est-à-dire codées en une suite de chiffres binaires 0 et 1. Par ce processus de numérisation, les données passent par un traitement informatique dont le principal outil est l'ordinateur duquel dépendent d'autres outils comme l'imprimante, le photocopieur, le numériseur et des fonctions comme l'accès aux réseaux, la gestion des tâches, la récupération des données, etc. L'ordinateur est désormais au centre de toute activité. Mais l'informatique avec l'outil ordinateur ne permet plus de représenter à elle seule tout l'environnement technologique. Il faut lui associer les télécommunications, et tous les nouveaux outils liés à leur synergie et leur convergence. L'ensemble de ces outils forment ce qui est communément appelé les NTIC soit les « nouvelles technologies de l'information et de la communication ».

Cette révolution technologique a évidemment eu un impact important sur le monde en général mais aussi sur celui des archives. Ses effets se sont particulièrement faits sentir sur la création de l'information, sur ses modes de transmission, de gestion et de conservation. De plus, elle a généré de nouvelles technologies applicables à la gestion des archives.

1. Présentation des technologies de l'information et de communication

Les technologies de l'information et de la communication recouvrent tous les nouveaux moyens et outils qui permettent de traiter (machines et logiciels de traitement), transmettre (réseaux et moyens de transmission et d'échange), de conserver (les supports de stockage) l'information électronique. Ces technologies existent depuis plus d'un quart de siècle et elles ne cessent d'évoluer pour devenir de plus en plus rapides et performantes. Par leurs capacités, leur omniprésence et les défis qu'elles posent dans la gestion de l'information, elles sont maintenant situées au cœur des préoccupations des personnes responsables de la gestion de l'information et, en particulier, les archivistes. Voyons en quoi consistent ces technologies et des principales étapes qui ont marqué leur développement.

1.1 Le matériel

Il ne fait aucun doute que l'avènement des ordinateurs a provoqué un bouleversement complet dans le monde de l'information. Ce bouleversement est proportionnel à l'augmentation de la puissance de ces ordinateurs et ceux-ci ne cessent d'accroître en performance tout en se miniaturisant constamment. □ titre de comparaison, les ordinateurs de la première génération (ordinateur central ou mainframe) pouvaient peser près de 25 tonnes, disposaient de quelques milliers d'octets de mémoire et pouvaient exécuter une centaine d'instructions par seconde. Les ordinateurs d'aujourd'hui sont dotés de microprocesseurs qui ne pèsent que quelques grammes, disposent de 256 à 512 meg (millions d'octets¹) de mémoire et peuvent exécuter plus de 1 milliard d'instructions par seconde. Toute cette puissance prend place dans un petit boîtier auquel on a relié un écran, une souris et un clavier et cet ensemble forme ce qu'on appelle un PC (personal computer ou ordinateur personnel).

□ l'augmentation de la puissance des ordinateurs, est également associée une diminution des coûts de production ce qui entraîne une présence de plus en plus accrue des ordinateurs personnels dans les foyers mais aussi et surtout dans les organisations de tous genres. On estime à plus de 100 millions, le nombre d'ordinateurs dans le monde et ce nombre croît de jour en jour.

¹ Un octet est un groupe de huit bits (chiffre binaire 0 ou 1) adjacents traités comme une unité.

Cette situation démontre sans aucun doute l'importance actuelle de la science de l'ordinateur soit l'informatique et de la place qu'elle occupe dans notre quotidien.

L'ordinateur est constitué de plusieurs composantes. Voici un bref rappel des principales ainsi qu'une brève description de chacune d'elles.

- Le microprocesseur : aussi appelé UTC (unité centrale de traitement) ou CPU (central processing unit) est le cerveau de l'ordinateur. C'est un circuit intégré, de la taille d'un timbre poste, chargé d'exécuter toutes les opérations de calcul de la machine. Il coordonne et vérifie le bon fonctionnement et le déroulement des opérations de l'ordinateur.
- La carte mère (motherboard) : c'est l'ossature électronique permettant à tous les composantes de communiquer entre eux. C'est un large circuit imprimé rassemblant l'essentiel des composants électroniques de l'ordinateur : le processeur, les barrettes de mémoire vive, le boîtier de la mémoire morte, le contrôleur. La carte mère, par l'intermédiaire de ses circuits électroniques, assure la bonne communication entre le microprocesseur et les périphériques tels que le disque dur, l'écran, le clavier, la souris, la carte de son, etc.
- La mémoire RAM (Random Access Memory) : mémoire interne dont la fonction est de gérer les données et programmes en cours d'utilisation. Il est possible d'y écrire et d'y lire mais les données se perdent au moment de couper l'alimentation électrique de l'ordinateur. Elle prend la forme de barrettes qui sont insérées sur la carte mère.
- La mémoire ROM (Read-Only Memory) : est celle qui renferme le programme de base permettant de démarrer l'ordinateur. Contrairement à la mémoire vive, l'utilisateur ne peut la modifier. Elle prend la forme d'un petit boîtier sur la carte mère.
- Le contrôleur (driver); est responsable de la gestion des données de lecture et d'écriture des lecteurs de disquettes et de CD-ROM et il se charge d'établir les connexions vers les différents périphériques de l'ordinateur (imprimantes, la souris, modem, etc.)
- Le disque dur : c'est un disque magnétique utilisé pour emmagasiner les données, les logiciels et le système d'exploitation. C'est le support de stockage interne de l'ordinateur.

- Les lecteurs; dispositif de stockage et de récupération de données inscrites sur un support quelconque. Ex. lecteur de disquettes, lecteur de CD-ROM (Compact Disk Read Only Memory) et plus récemment le lecteur de DVD (Digital Video Disk)
- La carte graphique ou vidéo : est l'élément qui se charge de la conversion analogique/digital de l'information visuelle entre l'unité centrale et l'écran.
- Carte de son : permet d'écouter et d'enregistrer de la musique et tout autre son.

Les composantes énumérées ci-dessus sont toutes insérées dans un boîtier ou une tour. □ cette dernière, il est possible de relier divers périphériques c'est-à-dire des dispositifs externes qui effectuent les tâches utiles pour l'ordinateur. Les principaux périphériques sont l'écran, le clavier, la souris, l'imprimante, le numériseur (scanner), le modem, le microphone, le lecteur CD-ROM, le lecteur DVD-ROM et le graveur de CD et de DVD.

1.2 Les logiciels

Les logiciels (software) sont des produits immatériels permettant de traiter des instructions humaines. Un logiciel est le résultat d'une série de commandes, d'instructions, de phrases écrites dans un langage formalisé soit un langage de programmation compris et reconnu par la machine. Un logiciel désigne ainsi tout programme dont l'objectif est de réaliser une tâche bien particulière (traitement de texte, gestion de bases de données, communication à distance, etc.). Le logiciel transforme l'ordinateur d'un outil qui peut théoriquement résoudre un problème en un outil qui le résout en pratique.

Les progrès dans le logiciel sont beaucoup plus lents que ceux du matériel. Ces progrès se situent surtout dans le domaine des langages qui offrent maintenant de meilleures structures de contrôle de programmes, de meilleurs environnements de programmation et des outils de programmation plus puissants.

Le progrès le plus marqué dans le domaine des logiciels concerne l'interface utilisateur. Ce dernier est la partie du logiciel qui sert d'intermédiaire entre l'utilisateur et le programme. Les écrans en mode caractères sont désormais remplacés par des écrans graphiques du type WYSIWYG (« What You See Is What You Get » pouvant être traduit par « ce que vous voyez à

l'écran est ce que vous obtiendrez à l'impression » ou encore « tel écran - tel écrit »). Le confort d'utilisation s'améliore par l'introduction de fenêtres, de menus, d'icônes et d'un outil de pointage. Les fenêtres rendent possible la représentation simultanée sur l'écran de plusieurs activités. Les menus permettent de choisir les prochaines actions. Les icônes représentent des objets informatiques sous forme concrète. L'outil de pointage, généralement la souris, permet de sélectionner les fenêtres, menus et icônes. L'exemple le plus concret et connu de cette transformation est le passage de l'interface DOS à celle du logiciel Windows de Microsoft. Cette évolution a donné naissance à une nouvelle génération de logiciels interactifs qui facilitent considérablement la tâche de l'utilisateur. Ce changement majeur est attribuable au fait que, jadis, l'interface utilisateur était la dernière partie à être conçue par le programmeur. Maintenant, elle est devenue la première. Beaucoup d'énergie et d'attention sont ainsi déployées à la conception de l'interface graphique d'un logiciel.

Il existe des milliers de logiciels sur le marché. Ceux-ci se répartissent parmi les principales catégories suivantes :

- Les logiciels « système ». Ce sont les logiciels qui assurent le fonctionnement de base de l'ordinateur et celui du matériel. Parmi ces logiciels on retrouve le BIOS (Basic Input-Output System) qui consiste en quelques programmes de soutien stockés de façon permanente dans les circuits électroniques de l'ordinateur (une puce de mémoire non-effaçable). Le BIOS prend charge de l'ordinateur au moment du démarrage. Il effectue alors des procédures de diagnostic qui vérifient le bon fonctionnement des principaux éléments du matériel. Ensuite, il entreprend ses procédures d'amorce qui visent à mettre en mémoire le logiciel ou système d'exploitation et à lui remettre le contrôle des opérations. Le système d'exploitation est constitué d'un ensemble de programmes qui aident l'ordinateur à gérer ses ressources. Il s'occupe de plusieurs fonctions internes et permet aux utilisateurs d'interagir, de manière transparente, avec les programmes lors de leur exécution. Le système d'exploitation le plus connu est Windows de Microsoft. Les logiciels Mac OS et Unix et Linux font aussi partie de cette catégorie.
- Les logiciels utilitaires, qui recouvrent tous les domaines d'application, s'avèrent indispensables à tout utilisateur conscient des problèmes issus de l'usage quotidien de

l'ordinateur; récupération de fichiers détruits par erreurs, organisation du disque dur, optimisation de la mémoire, compression de données, logiciels antivirus, sauvegarde, etc.

- Les logiciels de bureautique, généralement intégrés dans une « suite bureautique » telle Microsoft Office, offrent à l'utilisateur les traitements de texte (Word), les tableurs (Excel), gestionnaire de base de données (Access), les logiciels de présentation (Power Point) et les agendas et gestion du courrier électronique (Outlook). Tous ces logiciels visent à faciliter et accélérer les travaux de bureau et de secrétariat. Les logiciels de traitement de texte sont utilisés pour créer, éditer, sauvegarder et imprimer des documents. Ils sont particulièrement utiles pour effacer, déplacer, insérer et remplacer du texte. Un tableur permet de créer un chiffrier électronique. Constitué de rangs et de colonnes, le chiffrier sert à présenter et à analyser des données. La feuille de calcul est délimitée par des identificateurs de colonnes (haut) et de lignes (gauche). L'intersection d'une colonne et d'un rang s'appelle cellule ; sa position, une adresse. Le pointeur de cellule (le curseur du tableur) indique où se fera la prochaine entrée. Un gestionnaire de bases de données ou système de gestion de bases de données (SGBD), sert à organiser une grande quantité de données de façon à pouvoir en retrouver facilement les éléments désirés. Les données sont structurées en champs et en enregistrements desquels il est possible d'extraire, afficher, trier calculer et formater des données. Les logiciels de présentation permettent de créer des diaporamas en assemblant des textes, des images, des sons et des animations. Ils servent essentiellement à exposer des idées clairement ou à expliquer des projets afin de les mettre en valeur. Un gestionnaire d'informations personnelles (GIP) aide à organiser et à demeurer organisé. Les GIP sont des logiciels qui remplacent les agendas, les carnets d'adresses et les blocs-notes de même que la liste des tâches à accomplir. Il contient aussi un gestionnaire de courrier électronique.
- Les logiciels de traitement et de création d'applications multimédias qui regroupent diverses catégories de programmes spécialisés selon le média choisi (image, son, vidéo) et des logiciels permettant de réunir ces différents médias pour réaliser des séquences multimédias.
- Les logiciels de communications et de navigation qui sont destinés soit à la gestion des réseaux locaux, soit, et surtout, à la communication à distance. Ils se réorganisent actuellement autour des standards d'Internet et sont de plus en plus intégrés à l'ensemble des

autres programmes. Ils permettent également le partage des imprimantes, numériseurs et modem. C'est avec eux que la notion de communication devient le centre de l'informatique. Parmi ces logiciels, nous pouvons inclure les logiciels de réseaux Netware de Novell, LAN Manager de Microsoft et les navigateurs Internet Explorer et Netscape. Les logiciels de groupware (Lotus Notes, Novell Groupwise et Microsoft Exchange) et workflow qui soutiennent les processus de travail et le partage de tâches en réseau peuvent également s'inscrire dans cette catégorie.

- Les logiciels professionnels qui ont été créés pour satisfaire l'ensemble des besoins spécifiques à une profession; archivistes, bibliothécaires, dentistes, médecins, avocats, comptables, etc. □ cette catégorie, on peut également inclure les applications spécialisées à un type d'entreprise (aéronautique, banque, imprimerie) et qui ne se retrouvent pas sur le marché.
- Les logiciels de programmation qui permettent eux-mêmes de créer de nouveaux logiciels ou de nouvelles applications. Ces logiciels s'appuient sur les langages de programmation suivants : JAVA, Pascal, Langage « C » et « C+ », Basic et autres.
- Les logiciels éducatifs (dictionnaires, encyclopédies, atlas, apprentissage des langues, correcteurs, traducteurs, etc.) et de jeux dont le nombre et la qualité ne cessent de croître.

Plus récemment, nous observons l'accroissement du développement et de l'utilisation de logiciels de gestion d'entreprises soit les ERP (Enterprise Resource Planning) qui effectuent, de façon plus ou moins intégrée, les principales tâches de gestion d'une entreprise : comptabilité, paye, facturation, gestion des stocks, etc.

1.3 Les réseaux, Internet, Intranet et Extranet

1.3.1 Les réseaux : local et longue distance

Un réseau est un système qui permet l'intercommunication entre différents ordinateurs. C'est un ensemble de matériels et logiciels informatiques reliés entre eux de manière à permettre à plusieurs ordinateurs de fonctionner sur les mêmes logiciels, de partager des ressources et

d'échanger des données. Tout réseau est ainsi formé d'ordinateurs, de logiciels et de système de transmission. Un réseau est généralement bâti sur une architecture client/serveur. Le « serveur » est un ordinateur dédié à l'administration du réseau. Il gère l'accès aux ressources et aux périphériques et les connexions des différents utilisateurs. Il est équipé d'un logiciel de gestion de réseau : un serveur de fichiers prépare la place mémoire pour des fichiers, un serveur d'impression gère et exécute les sorties sur imprimantes du réseau, enfin un serveur d'applications rend disponible sur son disque dur les programmes pouvant être appelés à travers le réseau. Le « client » quant à lui est l'ordinateur qui demande ou reçoit les services ou informations disponibles sur le serveur distant. Cet ordinateur dispose ainsi de la partie « cliente » des logiciels pour dialoguer avec les serveurs.

Il existe deux sortes de réseaux; les réseaux locaux et ceux à longue distance. Un réseau local ou LAN (Local Area Network) constitue une connexion physique et logique d'ordinateurs dans un environnement de bureau. Tous les ordinateurs sont habituellement reliés au réseau par câblage. Ils doivent tous utiliser le même protocole de communication et la même infrastructure dont la plus connue est Ethernet. En ce qui concerne le réseau longue distance ou WAN (Wide Area Network), celui-ci est composé de systèmes situés en différents espaces physiques et géographiques pouvant même être dispersés dans le monde entier.

1.3.2 Internet et le World Wide Web (WWW)

Techniquement, Internet est un réseau mondial qui relie des réseaux d'ordinateurs entre eux. C'est en quelque sorte le réseau de tous les réseaux. Il est né, vers la fin des années 60, des besoins de l'armée américaine et des chercheurs de communiquer entre eux par la voie d'ordinateurs indépendants reliés ensemble par un réseau appelé ARPANET². Ce réseau, parce qu'il était formé de plusieurs réseaux, était en quelque sorte moins vulnérable aux attaques de l'ennemi : en cas de bris ou de destruction de l'une des parties du réseau, les communications pouvaient prendre un autre chemin pour arriver à destination permettant ainsi aux opérations de poursuivre leur procédure normale. Parallèlement à l'expansion de ce réseau, plusieurs universités ont mis sur pied un nombre considérable de réseaux dans le but de faciliter les

² Arpanet est la juxtaposition des mots « ARPA » (Advanced Research Project Agency) et NET qui signifie « toile ». Il a été créé au début de la guerre froide entre les États-Unis et la Russie au moment où planait une menace d'attaque nucléaire.

échanges internationaux. Ce n'est que par l'intégration des réseaux militaire, éducationnel et commercial qu'il est possible de parler d'un réseau international, soit l'Internet.

Une des nombreuses fonctions de l'Internet est de permettre le partage de différentes ressources entre plusieurs réseaux et des utilisateurs aux quatre coins du monde. Son contenu informationnel n'a de limites que celles de l'imagination humaine. Outre l'information, l'internet offre plusieurs services dont les plus connus sont le courrier électronique³, l'émulation de terminal (Telnet qui permet de se connecter à distance à tout type d'ordinateur), le transfert de fichier (FTP), les newsgroups (Usenet), l'internet relay chat (IRC) et bien d'autres.

Jusqu'en 1993, l'Internet était presque uniquement utilisé par les militaires et les chercheurs. Le mode de communication utilisé alors était basé sur le courrier électronique tel que nous le connaissons aujourd'hui. □ partir de cette date, l'Internet devient de plus en plus accessible à l'ensemble de la population mondiale, mais uniquement sous un format texte et sans image. Cependant, en 1995, l'arrivée du logiciel graphique ou navigateur «Netscape » est à la base d'un développement fulgurant d'Internet dans sa forme graphique telle qu'on la connaît aujourd'hui par le World Wide Web (WWW). Ce dernier est le fruit du travail de chercheurs suisses du Centre européen de recherche nucléaire (CERN) qui ont imaginé un système destiné à simplifier l'accès aux informations de cet immense ensemble mondial d'ordinateurs. Ils ont construit un système hypertexte permettant d'associer des données textes et des données multimédia utilisant l'ensemble des ordinateurs interconnectés comme une seule base de données universelle ; le World Wide Web. Les trois éléments clés du WWW sont : URL, HTTP et HTML.

- L'URL (Uniform Resource Locator) permet de créer une adresse électronique universelle à un document. Ainsi, chaque document a une adresse⁴ mondiale unique qui peut être référencée dans n'importe quel document. Ce sont les liens entre tous les documents qui tissent une toile d'araignée mondiale qui ont donc donné son nom au World Wide Web.

³ Le courrier électronique est un service de correspondance sous forme d'échange de messages électroniques, à travers un réseau informatique. Il permet à deux utilisateurs qui possèdent des adresses Internet de communiquer rapidement par l'envoi de messages écrits et de fichiers par le biais de serveurs faisant office de boîtes postales. Actuellement, le courrier électronique représente une très grande partie du trafic sur l'internet.

⁴ Cette adresse peut être représentée sous deux formats soit une adresse de type domaine telle www.umontreal.ca et un numéro IP formulé uniquement avec des nombres séparés par des points, Ex. 142.169.1.16. Seules les adresses de type domaine sont utilisées par les internautes.

- Quant au HTTP (Hyper Text Transfer Protocol) celui-ci permet de repérer les documents à travers l'ensemble du réseau par l'hypertexte qui permet, par un système de renvois ou hyperliens, de passer directement d'une partie d'un document à une autre, ou d'un document à d'autres documents choisis comme pertinents par l'auteur.
- Finalement, il y a le HTML (Hyper Text Markup Language), langage maintenant universel par lequel il est possible de créer des documents hypertextes et multimédias.

Mais l'élément déterminant du succès du WWW a été l'intégration de l'hypertexte, du multimédia et de tous les services existants dans un seul logiciel de présentation graphique, le navigateur (browser) tel Internet Explorer ou Netscape. La grande simplicité d'utilisation des navigateurs permet à tout utilisateur d'avoir accès à de l'information disponible dans les quatre coins du monde par de simples « clics » de souris.

1.3.3 Intranet et Extranet

Le succès d'Internet s'explique surtout par le fait que son infrastructure est indépendante de toute application et permet de supporter des applications aussi différentes que le transfert de fichiers ou la vidéoconférence. Par l'Internet, les ordinateurs de tout type peuvent communiquer et échanger des informations parce qu'ils utilisent un protocole de communication commun appelé TCP/IP (Transmission Control Protocol/Internet Protocol). Ce protocole a été mis au point au début des années 1970. Il autorise la libre circulation d'informations quel que soit le type de machine ou de réseau et fournit une interopérabilité totale entre utilisateurs différents et élimine ainsi toute incompatibilité de technologies. □ l'heure actuelle, la plupart des communications des réseaux locaux ou à distance ont tendance à passer par le système Internet pour bénéficier des nombreux avantages du protocole TCP/IP. Utilisé dans le cadre des réseaux locaux d'entreprise, le réseau prend le nom d'Intranet et d'Extranet lorsqu'il relie des sociétés et des organismes différents.

Un intranet est un réseau local et privé (entreprise) qui utilise les technologies de l'Internet : Web, e-mail, etc., mais ne s'ouvre pas aux connexions publiques. En résumé, l'intranet est un internet privé. L'intranet répond aux besoins de communication, de partage, d'efficacité et de souplesse des organisations. Il constitue le pilier du commerce électronique qui s'établit de plus en plus entre les organisations.

L'intranet se transforme en extranet pour les partenaires, les clients privilégiés et les fournisseurs. Il est un réseau de télécommunication et de téléinformatique constitué d'un intranet étendu pour permettre la communication avec certains organismes extérieurs, par exemple des clients ou des fournisseurs. L'extranet constitue donc un intranet auquel on a autorisé des accès privilégiés à certains partenaires, de fournisseurs ou de clients. L'extranet est un privilège, une permission d'accès octroyée aux partenaires d'une organisation. L'intranet d'une organisation est normalement restreint à ses employés, mais par souci d'efficacité, l'entreprise peut en ouvrir certaines sections à des partenaires. .

1.4 Le Groupware et le Workflow

1.4.1 Le Groupware

L'explosion des réseaux et le développement des technologies ont créé des nouvelles méthodes de travail dans les organisations. Parmi ces méthodes, on remarque la mise en place du travail en équipe assisté par ordinateur (Computer Supported Cooperative Work ou CSCW) appuyée sur le concept de Groupware (parfois traduit en français par collecticiel ou synergiciel) qui désigne à la fois le processus humain et organisationnel du travail en groupe et les outils technologiques nécessaires à l'accomplissement de ce travail. Ces outils permettent d'améliorer grandement le travail coopératif entre les différents membres d'une même équipe ou entre les membres d'équipes différentes, séparés ou réunis par le temps et l'espace, qui participent à un même projet ou une même activité et qui ont des informations à partager. Les principaux outils d'un Groupware sont les suivants : l'échange d'informations au sein d'un groupe (principalement par messagerie électronique), l'aide à la réalisation de tâches communes, l'aide à la planification et à la synchronisation du travail (calendrier et agenda de groupe), les facilités de stockage et de gestion documentaire et la conférence électronique.

1.4.2 Le Workflow

Mais, parmi les nombreuses possibilités offertes par le Groupware, on retrouve celle de la coordination du travail par le « Workflow » (pouvant être traduit en français par « flux de travail »). « Le Workflow est une application Groupware spécialisée qui se charge de superviser

la circulation sous forme électronique de documents ou de dossiers à traiter, lesquels s'insèrent dans une routine de travail où interviennent plusieurs personnes» (Levasseur, 1998, p. 22). Le Workflow est la gestion intelligente des flux d'informations partagées grâce à un logiciel capable d'assurer le traitement global des objets informationnels (courrier, rapport, transaction, etc) en définissant, modélisant, documentant, déclenchant et contrôlant l'enchaînement des tâches réalisées dans un groupe de travail.

Les voies des systèmes de Workflow favorisent le passage des tâches d'une activité, d'une fonction ou d'un procédé à une autre, en conformité avec les règles d'affaires en vigueur. Les logiciels de Workflow permettent de relier l'ensemble des activités d'une organisation selon une séquence logique et font le suivi des états fonctionnels. Les tâches (une demande du service à la clientèle, par exemple), peuvent être déclenchées de différentes manières, y compris par un message de courrier électronique, un appel téléphonique ou la réception d'un document. La nature du système de Workflow dépend du type de transactions commerciales ou d'activités opérationnelles de l'organisme. Le Workflow ponctuel soutient des procédés qui sont normalement amorcés par des participants à l'intérieur de ces derniers, plutôt que par des intervenants de l'extérieur. On peut en donner comme exemples la préparation et la présentation de formules de remboursement de frais de déplacement ou le processus d'évaluation des employés. Ces processus sont habituellement associés aux fonctions d'administration ou de soutien, plutôt qu'aux activités fonctionnelles.

Dans les applications de WorkFlow, on distingue classiquement quatre catégories :

- le Workflow de production, qui correspond à la gestion des processus de base de l'entreprise. Les procédures supportent peu de changements dans le temps, et les transactions sont répétitives. On peut y trouver par exemple la production de contrats d'assurance, la gestion de litiges, la gestion de réclamations clients, etc.
- le Workflow administratif, qui correspond à tout ce qui est routage de formulaires, basé en général sur une infrastructure de messagerie.
- le Workflow ad-hoc pour la gestion des procédures non déterminées, ou mouvantes.

- le Workflow coopératif, gérant des procédures évoluant assez fréquemment, et liées à un groupe de travail restreint dans l'entreprise.

1.4.3 Internet, Intranet et Groupware

Le concept de Groupware a été lancé et largement répandu par l'arrivée du logiciel « Notes » de Lotus. Le créateur de ce logiciel s'est largement inspiré de la notion de newsgroup (forums de discussion) présente sur l'internet. Aujourd'hui l'Intranet tend de plus en plus à intégrer les outils requis au travail de groupe de manière à devenir l'une des voies du Groupware.

1.5 Le multimédia

Le multimédia est l'ensemble des technologies informatiques qui permettent de traiter, de façon intégrée, plusieurs médias; texte, image, son ou vidéo. Il est né au début des années 1980 de la numérisation de l'ensemble de l'information (texte, voix, image et vidéo) et de leur intégration sur un même support soit le vidéodisque. □ partir de 1985, l'apparition du cédérom a rapidement supplanté le vidéodisque. Aujourd'hui, on retrouve des documents multimédias sur cédéroms, cartouches de jeux vidéo, DVD, bornes interactives, et évidemment sur Internet.

Un produit est généralement qualifié de « multimédia » s'il répond aux critères suivants :

- Il est sur support numérique; ce peut être fixe, il s'agit soit d'un cédérom ou d'un DVD; ou mobile, il s'agit alors d'Internet, d'un intranet ou d'un extranet.
- Il doit avoir un contenu; le contenu d'un produit peut prendre diverses formes, soit des images, des données audio, du texte et de la vidéo. Certains analystes considèrent qu'un produit multimédia doit contenir au minimum deux des quatre dimensions précitées, d'autres vont jusqu'à demander trois des quatre éléments précités
- Il doit être interactif : Un produit interactif est un produit qui permet d'utiliser un mode conversationnel, dans lequel l'utilisateur converse avec l'application par l'intermédiaire d'un

terminal, un ordinateur généralement, et peut intervenir pendant le déroulement des interventions.

Les domaines d'application pour le multimédia sont très nombreux. Les plus courants sur le marché et dans la vie de tous les jours sont :

- Le divertissement ; c'est le domaine d'application le plus connu du multimédia. On retrouve notamment les produits suivants dans ce domaine d'application : films à la carte, films sur vidéodisques numériques (DVD), télévision interactive, jeux vidéo, centres de jeux virtuels.
- L'éducation (enseignement) et la formation ; ces deux domaines bénéficient grandement de l'arrivée des technologies du multimédia. Les nouveaux médias ajoutent à la formation assistée par ordinateur traditionnelle un environnement plus dynamique et plus vivant grâce à l'intégration de plusieurs médias pour présenter un sujet. Les dictionnaires et encyclopédies sur cédéroms sont des outils grandement pratiques pour assurer la diffusion du savoir et des connaissances.
- La santé ; le multimédia permet, dans le domaine de la santé, une foule d'applications qui aident grandement les professionnels de la santé dans leur profession. Voici quelques applications multimédias utilisées dans le domaine névralgique de la santé publique : télémédecine, informatisation des dossiers des patients, imagerie médicale, information médicale par le biais de terminaux interactifs.
- L'animation visuelle; Ce domaine fut très bouleversé. L'arrivée des technologies associées au multimédia a littéralement bouleversé le domaine de l'animation visuelle. Que l'on parle des animations (dessins animés) ou des effets spéciaux pour le grand écran, les applications multimédias permettent d'enrichir grandement la qualité visuelle des films présentés au public. Le domaine de la publicité corporative bénéficie aussi grandement des applications multimédias. Les logiciels d'animation 2D, 3D, et les logiciels d'application pour les effets spéciaux sont les principaux produits que l'on retrouve dans ce domaine

- Les transactions financières ; Les applications multimédias permettent aux institutions financières de rendre plus conviviaux leurs services automatisés à la clientèle, tout en facilitant la gestion des transactions courantes et commerciales.
- Les arts et la culture ; L'industrie du multimédia ouvre la porte à d'innombrables possibilités pour les créateurs d'œuvres artistiques. Les simulations virtuelles, par exemple, permettent d'offrir au spectateur-acteur une expérience qui ne pourrait se réaliser sans les technologies du multimédia. La mise en ligne des oeuvres artistiques visuelles ouvre la porte à une démocratisation sans précédent de l'art.

Les services corporatifs. Les entreprises ont recours à l'expertise des producteurs multimédias pour mieux diffuser l'information concernant leurs produits et services.

1.6 Les bases de données

Une base de données (électronique) est un ensemble intégré de données réparties sur un ensemble de fichiers et d'enregistrements logiquement reliés les uns aux autres. Elle est comparable à une boîte de fiches cartonnées, dont le contenu peut être présenté et classé de plusieurs manières différentes. La plupart des bases de données sont constituées d'une suite de fiches appelées « enregistrements », toutes conçues sur le même modèle, qui comportent un certain nombre de rubriques nommées « champs » destinées à recevoir une information bien précise. Le contenu d'une base de données peut être visualisé de différentes façons et on doit pouvoir effectuer les opérations suivantes :

- L'interrogation ou la consultation,
- La sélection,
- Le tri,
- La mise à jour,
- L'insertion,
- L'élimination.

L'origine des bases de données remonte au début des années 60 et leur apparition se voyait un moyen de remédier à la situation anarchique créée par la multiplication de fichiers et de programmes disparates, redondants et incompatibles dans les grands organismes. Il illustre le souci de regrouper toutes les données utiles à un même organisme à l'intérieur d'une seule entité informatique pour permettre la réalisation par l'ordinateur de diverses opérations de gestion tout en donnant à chaque utilisateur un accès aux données le concernant dans des fichiers spécifiques. La technologie essentiellement logicielle des bases de données que constituent les fichiers a fait des progrès considérables en termes de convivialité et de facilité d'utilisation à la fin des années 1980. Il existe plusieurs types de bases de données lesquelles se distinguent par le type d'organisation logique des données. On retrouve notamment des bases de données hiérarchiques, réseaux, relationnelles et objet.

- Hiérarchique : les données sont organisées de façon arborescente. Il n'y a pas de liaison entre les branches de même niveau; c'est donc un modèle simple qui n'autorise que peu d'interrogation.
- Réseau : C'est un modèle hiérarchique qui permet d'établir des connexions entre les différents éléments ce qui permet plus d'interrogations que le modèle hiérarchique.
- Relationnel : Les données sont stockées sous la forme de relation dans des tables et chaque table est constituée de plusieurs colonnes. Ce type de structure permet d'établir des connexions au moment de l'exécution. C'est la méthode d'organisation de données la plus souple. Dans cette structure, il n'existe aucun chemin de type hiérarchique pour accéder aux données, celles-ci étant conservées dans diverses tables faites de rangs et de colonnes. On appelle relation le lien qui relie deux tables.
- Objet : Les bases de données relationnelles sont actuellement les plus répandues. Le qualificatif « relationnel » signifie que la structure est fondée sur des relations entre des informations et non, comme pour les autres systèmes de fichiers, sur leur rangement spatial. Les bases de données relationnelles sont bien adaptées à la gestion d'informations textuelles ou numériques mais elles sont beaucoup moins adaptées pour organiser et ranger des informations plus complexes de type « images », « sons » ou « séquences vidéo », qui

nécessitent de faire appel à des logiciels de type objet. Le logiciel de bases de données orienté objet est structuré autour des données plutôt qu'autour des actions comme le font les autres types de logiciel de bases de données.

Les bases de données s'appuient sur les trois principes suivants :

- éliminer les données redondantes, c'est-à-dire que celles qui, bien qu'identiques, existent dans plusieurs fichiers d'un même organisme ne sont saisies et conservées qu'une seule fois.
- assurer l'intégrité des données, en contrôlant ou en interdisant toute modification sur les données déjà saisies. Toute intervention de cette nature ne peut se faire que par une personne dûment habilitée soit le gestionnaire de la base de données.
- distinguer les données et les programmes. Par cette distinction, les mêmes données peuvent être utilisées par des programmes différents. Mais chaque application n'accède qu'aux données nécessaires et seulement à celles-là.

Les bases de données peuvent être petites ou volumineuses ; on peut les classer en quatre catégories : individuelle, centralisée, répartie et commerciale.

- Base de données individuelle : Cette catégorie se retrouve surtout sur les micro-ordinateurs. Il s'agit d'une collection de fichiers intégrés utilisée principalement par une seule et même personne.
- Base de données centralisée : Cette base est généralement constituée de fichiers sur les activités courantes d'un organisme que partagent les membres d'une organisation. Les données sont habituellement intégrées à un ordinateur central administré par un informaticien reconnu comme l'« administrateur de bases de données⁵ ». Les utilisateurs de la base de données y ont accès à partir de leur micro-ordinateur via un réseau local.

⁵ L'administrateur de base de données est la personne responsable de la gestion de la base de données. Il aide à déterminer la structure des données la plus appropriée pour répondre aux besoins des utilisateurs. Il s'efforce de maximiser le rendement de l'ensemble du système et décide le type d'accès (consultation et modification) de chacun des utilisateurs aux informations. Enfin il voit à assurer la sécurité physique de la base et à la protéger contre le feu, le vandalisme ou tout autre désastre. De plus, l'administrateur doit se préoccuper de la sécurité et de la protection de la vie privée.

- Base de données répartie : Il peut arriver que les données d'une organisation soient stockées à plusieurs endroits plutôt qu'à un seul. Dans ces circonstances, les utilisateurs ont accès aux données à l'aide d'un réseau de communications et d'un serveur.
- Base de données commerciale : Ce sont les banques d'informations rassemblant des informations sur une vaste gamme de sujets, mises à la disposition du public moyennant un abonnement ou des frais de consultation.

Le fonctionnement global de la base de données est assuré par un logiciel appelé « système de gestion de bases de données (SGBD) ou DBMS : Data Base Management System » dont la fonction est d'organiser les données et de contrôler leur usage. Les SGBD les plus connus pour les micro-ordinateurs sont les logiciels Access, File Maker, Paradox et dBase. Un SGBD est constitué d'un dictionnaire de données et d'un langage de requêtes. Le dictionnaire de données contient la description de la structure des données utilisées dans la base de données. Le dictionnaire définit le nom utilisé pour désigner un champ en particulier et le type de données que contient ce champ (alphabétique, numérique ou autre). Il peut aussi préciser la taille en nombre de caractères (octets).

L'accès à la majorité des bases de données est assuré grâce à un langage normalisé dont le plus répandu est le SQL (Structured Query Language). C'est un langage simple, qui permet toutes les opérations possibles sur la base de données. Il comporte des commandes comme AFFICHER, MODIFIER, COMPARER, GÉRER, LISTER et METTRE □ JOUR

Une distinction est à faire entre la base de données et la banque de données. Cette dernière correspond à un ensemble de fichiers apparentés rassemblant des données dans un domaine défini de connaissances et organisé pour être offert aux consultations d'utilisateurs. Le concept de banque de données est davantage lié à l'exploitation et au stockage des données quand celui de bases de données est lié à l'organisation des données.

1.7 Les télécommunications

Notre survol sur les technologies se doit d'être complété par quelques notions sur les communications et télécommunications afin de couvrir l'ensemble des notions couvertes par le

sigle NTIC (nouvelles technologies de l'information et de communication). Voyons sommairement ce qu'elles sont et comment elles se sont développées.

Les télécommunications, ou communications à distance, sont nées au XIX siècle avec la télégraphie et la télégraphie sans fil. Vint ensuite la téléphonie qui permit la mise sur pied du premier véritable réseau mondial. Puis on vit apparaître les télécopieurs, les téléphones cellulaires, les satellites de communication et la câblo-distribution. Toutes ces technologies permettent aux individus de communiquer entre eux mais elles représentent également les moyens par lesquels les ordinateurs peuvent communiquer entre eux via les différents réseaux.

Lorsque deux ordinateurs se retrouvent côte à côte, on peut facilement imaginer un lien direct par un câble pour les unir. Cependant, quand les distances deviennent importantes, l'établissement du lien physique pose le problème de la faisabilité et, surtout, du coût. C'est pourquoi le téléphone est apparu comme la solution idéale pour la transmission des données informatiques. Le réseau téléphonique était déjà bien établi et ses ramifications recouvraient la majorité des zones habitées. L'utilisation de ce réseau posait par contre un petit problème. Le téléphone a été conçu à l'origine pour transmettre la voix humaine, donc du son, il nécessite un signal analogique, c'est-à-dire un signal ayant la forme d'une onde, d'une oscillation continue. Quant aux ordinateurs, ils émettent et reçoivent des signaux numériques, composés de suites d'impulsions électriques : les signaux 1 ou 0. Alors, pour convertir les signaux numériques de l'ordinateur en signaux analogiques et vice versa, on doit utiliser un modem (MO pour modulateur et DEM pour démodulateur). Grâce à ce dernier, les ordinateurs transmettent des données numériques au moyen de lignes analogiques téléphoniques. Cependant, les dizaines de millions de micro-ordinateurs dans le monde qui utilisent les centaines de milliers de serveurs connectés à l'Internet ont provoqué une explosion des besoins en télécommunications. Les échanges de données entre ordinateurs nécessitent des capacités et des vitesses de transmission beaucoup plus élevées que pour les câbles téléphoniques. Il fallait donc augmenter considérablement les capacités des réseaux de télécommunications pour répondre à la croissante utilisation des réseaux. L'utilisation du câble coaxial⁶ a contribué à l'amélioration des communications de type réseau mais c'est par la mise en place et l'utilisation de plus en plus répandue des câbles de fibre optique comme technologie de communication que les réseaux ont

⁶ Câble de transmission à haute fréquence qui remplace les multiples fils téléphoniques par un gros conducteur recouvert d'une gaine isolante. C'est le type de câble qu'utilisent la plupart des cablo-distributeurs.

considérablement amélioré la vitesse et l'efficacité de la transmission des données. Avec ces câbles, les données sont transmises sous forme d'impulsions lumineuses à travers des filaments de verre. Les capacités de ces câbles sont de 26 000 fois celles des câbles téléphoniques.

Parallèlement aux développements des réseaux terrestres, il y eu celui des télécommunications par satellites. Ce genre de communications fait appel aux micro-ondes qui utilisent des ondes radio de haute fréquence qui circulent en ligne droite dans les airs. Puisque les ondes ne suivent pas la courbure de la Terre, leur portée est limitée – elles ne couvrent que de courtes distances. De fait les micro-ondes représentent une voie de transmission fiable pour envoyer des données entre les édifices d'une ville. Pour parcourir de plus longues distances, elles doivent être relayées par des stations de relais ; celles-ci, munies d'antennes paraboliques, sont installées, entre autres, sur les gratte-ciel et les montagnes. Pour couvrir les grandes distances qui séparent les continents on utilise comme stations de relais micro-ondes, des satellites qui tournent à environ 35 000 kilomètres de la terre.

Tous ces médias constituent la toile de fond de la télématique, c'est-à-dire, l'ensemble des services de nature ou d'origine informatique pouvant être fournis à travers un réseau de télécommunications.

Les technologies de l'information et de communications font appel à d'innombrables outils et notions dont la nature varie considérablement en terme de diversité et de complexité. Le but de cette partie visait à présenter sommairement les technologies les plus courantes et les plus accessibles par de simples utilisateurs. Toutes ces technologies sont en constante évolution et il est impératif de la part des archivistes de suivre leur développement afin d'adapter leurs méthodes de travail et de répondre aux nouveaux besoins créés par l'intégration de plus en plus massive de ces technologies dans les organisations.

2. Les technologies appliquées aux archives

Nous avons, jusqu'à maintenant, passé en revue et examiné les différentes technologies de l'information et de communication qui caractérisent le quotidien des interventions archivistiques. Nous allons maintenant voir comment ces technologies peuvent être appliquées

aux archives et surtout, nous allons identifier quelques problèmes reliés aux archives électroniques et les principaux défis à relever par leur gestion et leur conservation.

Parmi les nombreuses technologies présentes et disponibles qui ont eu un impact dans le domaine des archives, nous pouvons identifier celles qui ont marqué davantage le domaine des archives sont celles des logiciels de gestion des documents et d'archives et l'imagerie électronique dans laquelle on peut inclure la GED et la GEIDE.

2.1 Les logiciels de gestion de documents et d'archives

Une des principales applications de l'informatique est celle d'automatiser les tâches effectuées normalement par l'homme. De nombreux domaines ont profité de l'automatisation et le domaine des archives ne fait pas exception à ce processus. C'est par l'intégration de logiciels de gestion des documents et des archives que s'est effectué ce processus d'automatisation dans le domaine des archives. Il existe de nombreux logiciels de gestion de documents sur le marché. Parmi les nombreux logiciels disponibles, certains sont spécialisés et s'adressent à une forme déterminée d'organisation ou de documents. Par exemple, certains logiciels de gestion des documents s'adressent quasi-exclusivement à la gestion des documents scolaires. D'autres logiciels sont destinés à la gestion d'un type de documents en particulier par exemple, les comptes rendus de réunion ou la correspondance. Mais la plupart des logiciels sont plus généraux et s'adressent à des clientèles diversifiées. Généralement, ils permettent, en tout ou en partie :

- d'automatiser les fonctions archivistiques de création, de classification (plan de classification), d'évaluation (calendrier de conservation), de description, de diffusion et d'accès (instrument de recherche et index de repérage)
- d'automatiser la gestion des dossiers ou documents actifs, semi-actifs et inactifs.

De plus, certains logiciels offrent certaines options telles la gestion des dépôts des semi-actifs ou la gestion des prêts de dossiers. Notons que ces types de logiciels sont principalement destinés à gérer des documents et ne donnent nullement accès à leur contenu. Ce sont les systèmes de GED, que nous verrons plus loin, qui sont conçus pour offrir ce type d'accès.

On retrouve sur le marché plusieurs logiciels de gestion dite « intégrée » d'archives. Le terme « intégré » signifie que le logiciel prend en charge les fonctions archivistiques essentielles soit la classification, le calendrier de conservation, la description et le repérage ainsi que des dossiers et des documents (actifs, semi-actifs et inactifs), quel que soit le support d'information (papier, électronique, microfilm).

Un logiciel de gestion de documents et d'archives est habituellement développé sur une plate-forme d'un logiciel de bases de données relationnelles. Par conséquent, pour bien comprendre la structure et le fonctionnement de ce type de logiciel, il faut se référer aux principales caractéristiques reliées aux systèmes de gestion de bases de données. Entre autres, un logiciel de gestion de documents et d'archives permettra l'insertion, l'interrogation ou la consultation, la sélection, le tri, la mise à jour et l'élimination ou la destruction d'information. Plus particulièrement, ce type de logiciel vise à :

- éliminer certaines opérations mécaniques,
- faciliter la production d'index (en vocabulaire libre ou contrôlé) et assurer une uniformité de présentation des index,
- faciliter la mise à jour et le contrôle du plan de classification et du calendrier de conservation en établissant une restriction d'accès pour toutes modifications,
- produire des listes de déclassement pour les documents ayant terminé leur période active ou leur période semi-active selon le calendrier de conservation,
- rendre le repérage rapide et facile,
- produire les listes de documents à détruire selon le calendrier de conservation,
- contrôler le prêt de documents ou dossiers,
- produire des instruments de recherche et des rapports de statistiques diverses (entrées, dossiers communiqués ou prêtés, nombre de dossiers traités, etc).

L'utilisation de logiciels de gestion documentaire comporte de nombreux avantages dont les principaux sont :

- La rapidité de repérage des dossiers,
- La diminution importante des erreurs humaines,
- La rapidité de la production d'index,
- L'application informatique du plan de classification et du calendrier de conservation,

Par contre, les coûts d'acquisition de ces types de logiciels sont parfois très élevés et leur acquisition implique obligatoirement une formation du personnel. De plus, le choix de ce type de logiciel doit être le résultat d'une démarche de préparation au cours de laquelle on a soigneusement évalué les besoins et identifié des critères de sélection tels que :

- Le coût du logiciel : un choix de logiciel doit inévitablement être fait en fonction des contraintes budgétaires mises à la disposition de l'organisation.
- La réputation de la compagnie. On devra s'assurer de l'expérience de la compagnie en consultation d'autres organisations ayant acquis le logiciel pour connaître leur niveau de satisfaction.
- Le service après-vente : en l'absence d'un service informatique ou d'une personne spécialisée en informatique dans l'organisation, on aura intérêt à choisir une compagnie qui offre un bon service après-vente.
- Le temps de la formation avant l'utilisation. Certains logiciels sont simples et d'autres sont compliqués. La période d'apprentissage doit être raisonnable.
- Le recours à une personne spécialisée en gestion de documents. Si l'organisation emploie déjà une personnes spécialisée en gestion des documents ou prévoit son engagement, le logiciel choisi peut être plus sophistiqué. Si au contraire, l'exploitation du logiciel sera faite par des personnes non spécialisées, le logiciel choisi devra être simple.

- La gestion du plan de classification. Le logiciel choisi doit pouvoir assurer les opérations suivantes : l'intégration du plan de classification, la possibilité d'ajouts, de retraits ou de modifications, l'inscription automatique du descripteur officiel à l'inscription de la cote lors de l'enregistrement de dossiers, la possibilité d'impression totale ou partielle du plan de classification.
- La gestion du calendrier de conservation ; Le logiciel choisi doit pouvoir assurer les opérations suivantes : l'intégration du calendrier de conservation, la possibilité d'ajouts, de retraits ou de modifications, l'inscription de remarques, la possibilité d'impression totale ou partielle du calendrier de conservation.
- La production de liste de déclassement. Le logiciel choisi doit permettre l'application informatique du calendrier de conservation et produire des listes de déclassement claires et complètes.
- La production d'index. La production des index varie beaucoup selon les logiciels. Certains logiciels sont très flexibles et c'est l'utilisateur qui détermine les items et l'ordre de présentation. Si la présentation des index possibles est prédéterminée, il faut vérifier si ces index standards offrent une présentation satisfaisante pour l'organisation.
- Le repérage rapide et facile : La flexibilité et la vitesse de réponse sont les deux éléments de base pour l'étude du repérage.
- L'intégration possible de normes pour la description. Les normes de description ISAD(G) et RDDA étant de plus en plus utilisées pour la description des documents d'archives, il faut donc tenir compte de cet aspect lors du choix de logiciel.
- Gestion des documents sur différents supports. : Le logiciel doit permettre la gestion de documents sur des supports autres que le papier ; des photographies, des plans, des vidéos voire même des documents électroniques.
- La gestion du prêt de dossiers. Si une des caractéristiques de l'organisation est une importante circulation interne et externe des dossiers, la gestion du prêt des dossiers devient une fonction essentielle.

- Le réenregistrement automatique des dossiers. Certains dossiers sont répétitifs d'années en années. Le réenregistrement automatique permet d'éviter de recommencer chaque année l'enregistrement des mêmes items. Le logiciel reproduit les mêmes informations et ne change que l'année.
- La gestion des emplacements des dépôts de semi-actifs et la salle des archives. C'est le volume de la masse documentaire d'une organisation qui détermine la nécessité de ce critère. Si l'organisation gère plusieurs dépôts de semi-actifs et une salle d'archives à plusieurs emplacements, il est recommandé de les gérer via le logiciel de gestion de documents et d'archives
- La production d'étiquettes. Lors de l'enregistrement d'un dossier, certains logiciels offrent la possibilité d'imprimer, sur étiquettes, tous les renseignements d'identification du dossier.

□ ces critères, on doit également tenir compte des critères généraux, techniques, fonctionnels et spécifiques à l'archivistique dont l'étudiant pourra prendre connaissance à la lecture de l'article d'Estelle Bouthillier identifié en lectures complémentaires.

L'avancement technologique actuel dans les logiciels de gestion documentaire offre maintenant des possibilités quasi illimitées. L'acquisition d'un de ces logiciels constitue une démarche importante. Elle justifie largement tous les efforts qui doivent être investis aux études préalables à l'achat de ce type de logiciel dans le but d'acquérir un produit conforme aux exigences et au budget de l'organisation.

2.2 La GED et la GEIDE

Parmi les possibilités offertes à la gestion des documents et des archives par les nouvelles technologies, la GED est certes celle qui prend une place de plus en plus prédominante. Le terme GED correspond à la Gestion électronique de Documents, concept qui a pris naissance dans le milieu des années 1980 pour répondre aux besoins « d'archivage électronique ». □ cette époque, on voyait dans la GED une façon de remplacer les salles d'archives ou les microfilms en substituant l'image numérisée du document à l'original. On associait alors la GED au processus

d'imagerie électronique ou numérique par lequel on photographie, à l'aide d'un numériseur ou scanner, des documents sur papier pour les convertir en image et finalement les conserver sur un support optique. Dès lors, on pouvait avoir accès aux documents numérisés et les consulter en ayant recours à des logiciels de recherche. Aujourd'hui, la GED est devenue une composante des systèmes d'information des organisations et elle prend en compte tout type de fichiers numériques créés ou exploitables par des ordinateurs ou des systèmes numériques. C'est pourquoi on parle davantage de GEIDE - Gestion Électronique d'Informations et de Documents Existants- et dans cet ensemble, il est possible d'associer, au sein d'un même dossier, des fichiers issus de logiciels de bureautique (traitements de textes, tableurs, etc) ou de systèmes centraux, des télécopies reçues par fax-modem, des plans ou des dessins industriels, des photographies numériques, des courriers électroniques, des documents récupérés sur Internet ou encore du son et de la vidéo. L'avantage de la GEIDE est de réunir les moyens et méthodes qui permettent de classer et de réutiliser intelligemment et rapidement toutes ces informations. Concrètement, la GEIDE correspond à l'ensemble des techniques qui permettent d'accéder rapidement et le plus économiquement possible aux masses d'informations et de documents générés ou reçus par une organisation.

Par conséquent, la GEIDE c'est la :

- **G**estion, c'est la définition des règles et des méthodes d'utilisation et de gestion des flux d'informations.
- **É**lectronique, ceci implique que cette information doit être dématérialisée et accessible au traitement par ordinateur.
- **I**nformation, c'est considérer que l'information est une notion d'élément de connaissance, de renseignement, quel que soit la forme ou le support.
- **D**ocuments, il désigne le support qui contient les informations de référence.
- **E**xistants cela sous-entend que la GEIDE n'a pas pour vocation de créer ou concevoir l'information, mais d'en assurer la gestion.

Ainsi, la GEIDE couvre l'ensemble des systèmes logiciels et matériels destinés à organiser, gérer et distribuer, sous forme électronique, les informations existantes dans le cadre de solutions intégrées et réparties dans les organisations. La GEIDE est ainsi un véritable concept qui fait partie intégrante d'un projet global d'entreprise où l'on associe de plus en plus le pouvoir à la connaissance (concept de gestion de la connaissance).

2.2.1 Les applications GEIDE

La GEIDE est un terme général qui recouvre de multiples compétences parfois très différentes. Il est donc préférable de parler d'applications de GEIDE que l'on peut regrouper en cinq grandes catégories et solutions GEIDE:

- La GEIDE Administrative. Ce type d'application concerne généralement ce qu'il est convenu d'appeler la Gestion de Dossiers électroniques et correspond au classement de documents administratifs divers et variés, souvent sous une forme d'image numérisée. On trouve parmi ces documents des dossiers étudiants (bulletin, acte de naissance, baptistère, correspondance, etc). Généralement la GEIDE Administrative fait partie d'une application globale de gestion et permet à l'utilisateur d'accéder rapidement aux documents dont il a besoin, sans avoir à se déplacer ou à encombrer son bureau de dossiers physiques. Bien souvent ce type de GEIDE intègre des outils de communication (fax) et de gestion des flux (workflow) à l'ensemble de l'application; il permet d'échanger des documents via le réseau, de demander un complément d'information à un collègue, de soumettre à sa réflexion les documents extraits d'un dossier, de valider un document avec une signature électronique, etc.
- La GEIDE Bureautique. Les outils de GEIDE Bureautique, dont le plus représentatif est Lotus Notes, sont en concordance avec les plates-formes bureautiques classiques et qui se veulent de plus en plus communicantes dans un concept de travail de groupe « groupware ». Ils permettent de manipuler des documents dans leur format bureautique d'origine (Word, Excel...), de centraliser leur classement sur un ou plusieurs serveurs, d'échanger ces documents par messagerie électronique. Bien souvent, ils intègrent aussi des fonctions de distribution et de télécopie, d'agenda électronique.

- La GEIDE COLD. Le terme de **COLD**, qui est l'acronyme de Computer Output on Laser Disk, est utilisé en référence aux applications de COM (Computer Output on Microfilm) car c'est une technique qui se substitue à celle de la micrographie. Ce type d'application est la toute première application de la GEIDE. Elle permet de stocker et d'indexer automatiquement l'ensemble des documents générés par des ordinateurs centraux destinés à l'impression. Il peut s'agir, par exemple, de factures, de bons de livraison, de relevés de paie, etc.
- La GEIDE Documentaire. La GEIDE Documentaire est directement issue de la recherche documentaire particulièrement utilisée dans le cadre d'applications du type bibliothèque, documentation scientifique, et autres. La GED a apporté à cette catégorie d'application l'accès à l'image du document physique ou directement à son contenu. Elle implique l'organisation et la gestion des informations de référence et la mise en oeuvre de modules de recherche documentaire. Ce type de GEIDE se caractérise principalement par ses méthodes d'indexation et de recherche qui font le plus souvent appel à des thesaurus, au plein texte (indexation et recherche en texte intégral) avec opérateurs de proximité, à une recherche par concept et parfois à un mode d'interrogation en langage naturel.
- La GEIDE Technique. La GEIDE Technique, aussi appelée dans certains cas GEIDE Métier, concerne toutes les applications dans lesquelles le concept de Gestion Electronique de Documents est directement lié à une profession ou à un métier. Les applications de GEIDE Technique en sont l'exemple de plus courant et, parmi elles, les plus nombreuses sont celles pour bureaux d'études, pour la gestion de plans ou la gestion de la documentation technique sur un projet industriel particulier.

Une solution GED ou GEIDE est un processus au cours duquel se succèdent les fonctions :

- de saisie/création⁷ en format numérique
- de formatage
- de traitement
- d'indexation,
- de stockage,
- de recherche,

⁷ On retrouve aussi l'expression « acquisition » ou « capture » pour identifier l'opération de saisie de l'information.

- de consultation
- de diffusion.

Même si les fonctions de la GED sont nombreuses et pas forcément les mêmes d'une application à l'autre, l'on retrouve toujours un noyau de fonctions de base que sont (1) la saisie/création, (2) le stockage et la conservation et (3) la recherche/consultation et la diffusion.

2.2.2 La saisie/création

Les solutions de GED ET GEIDE impliquent nécessairement que l'information soit sous une forme électronique c'est-à-dire une suite de bits à 0 ou à 1. Pour ce faire, il faut donc acquérir ou convertir numériquement l'information et ce processus constitue le premier maillon de la chaîne GED. Le symbole de l'application de GEIDE est le **scanner** ou **numériseur** qui transforme le document papier en une image électronique manipulable et communicable sur ordinateur. A chaque forme de document correspond un type d'appareil ou un dispositif logiciel : **scanner**, carte fax, logiciel de transfert de fichier. Les conditions de l'acquisition sont étroitement liées aux types de documents, aux volumes, aux contraintes de l'organisation : service de numérisation associé à un service courrier, répartition des tâches sur les postes de travail des utilisateurs, importation de documents, recours à un centre de numérisation, etc.

Il existe plusieurs types de scanners ou numériseurs et chaque type est adapté à une utilisation particulière. Les principaux types de scanner/numériseur sont les suivants :

- **scanner à main** : en déplaçant le scanner sur le document, l'utilisateur découpe ce document en lignes. Certains de ces scanners sont motorisés afin de réguler le déplacement;
- **scanner à défilement** : le document est entraîné par un système de rouleaux qui le fait défiler devant la ligne de lecture;
- **scanner à plat** : le document reste fixe mais le système de lecture se déplace devant lui;
- **scanner volumétrique** : le document reste fixe et les systèmes de lecture (multiples) sont pivotants afin de corriger les erreurs de parallélisme (utilisé notamment dans le cas d'anciens registres reliés).

Outre la numérisation, la reconnaissance optique de caractères constitue une autre technologie d'acquisition de l'information. L'ordinateur souffre d'un handicap majeur : il ne sait pas lire et le clavier représente un frein important à l'entrée massive de données et une cause fréquente d'erreurs. C'est ici que peut intervenir la technologie OCR (Optical Character Recognition) ou ROC (Reconnaissance Optique de Caractères) pour suppléer à ces deux faiblesses. L'OCR permet de transformer les dessins de lettres en caractères manipulables par l'ordinateur. Les logiciels de reconnaissance optique de caractères, jumelés à l'utilisation d'un scanner, visent ainsi la conversion de documents imprimés sur papier en fichiers en ASCII⁸ traitables par l'ordinateur. L'OCR offre de nombreuses possibilités :

- de retraiter l'information dans un traitement de texte (copier/coller, reprise de textes) ou autre outil bureautique, bases de données, tri automatique, édition électronique, messagerie, etc. ;
- d'indexer automatiquement l'information contenue dans le texte (indexation de type intégral) ;
- de réduire la taille du fichier formé par un document ;
- d'améliorer la rapidité et la qualité d'impression et de visualisation à l'écran ;

En contrepartie, même si l'OCR demeure une technologie performante, elle est très dépendante de la qualité des documents qui lui sont soumis :

- les caractères tordus, cassés, les caractères qui se touchent seront mal reconnus. Exemple : IVI sera reconnu M. Le taux de reconnaissance est supérieur sur des polices proportionnelles ;
- les marques de stylo ou crayon : si ces marques ne croisent pas le texte, certains systèmes sauront les identifier et les séparer (signature, marque dans la marge) ;
- le mauvais contraste ;
- la présence d'un fonds coloré.

⁸ Acronyme formé par les mots anglais : American *Standard Code for Information Interchange*. C'est une norme de codage de 128 caractères alphanumériques sur 7 bits. Les versions étendues sur 8 bits, 256 caractères sont adaptées suivant les pays. (Définition tirée du dictionnaire de l'informatique disponible sur le site <http://www.dicofr.com>)

Même si le numériseur et l'OCR sont des moyens très répandus pour l'acquisition de l'information, il faut rappeler que la GEIDE a pour vocation d'intégrer et de gérer tous les types de documents, y compris ceux qui se trouvent déjà sous une forme électronique. En plus des documents papier, les types de documents les plus fréquemment rencontrés sont les suivants : télécopies, documents bureautiques issus de logiciels de traitement de texte, de tableurs, graphiques et plans, fichiers informatiques et en particulier les fichiers spool, sons, images animées, documents composites. Les techniques d'acquisition doivent être adaptées aux types de documents concernés, aux conditions de l'acquisition et aux volumes en cause.

Nous pouvons ainsi conclure qu'il existe trois possibilités de faire de la saisie numérique de l'information :

- La conversion en numérique de documents analogiques : il s'agit d'effectuer une conversion en faisant appel à différentes technologies telles que le numériseur ou l'OCR
- La saisie directe de l'information en mode numérique : cela va du cas le plus simple comme la saisie directe de texte par un logiciel de traitement de texte jusqu'à des cas plus sophistiqués tels que l'acquisition de photo ou vidéo avec des appareils numériques.
- La collecte et l'assemblage de documents déjà numériques : dans ce cas l'information de base est déjà numérique mais souvent éparpillée sur différents postes et serveurs, et disparate dans son format et dans sa forme. Ceci implique généralement un travail axé sur la conversion de format.

Les formats d'encodage et d'enregistrement.

La saisie et la création de l'information numérique impliquent nécessairement que cette information soit représentée sous forme d'un code binaire composé de 0 et de 1 (les bits). De nombreux formats d'encodage permettent ensuite de présenter cette information sous forme textuelle, image, vidéo, sons, etc. Un format est une donnée technique qui informe sur la méthode d'enregistrement d'un document. Un format de document électronique est généralement identifié par l'extension figurant après le nom du fichier. Cette extension indique aux logiciels à quel type de format le document appartient afin qu'il puisse le décoder et par conséquent l'afficher correctement. Par exemple, un fichier enregistré par le logiciel traitement de texte

Word aura une extension « .doc ». Il existe de nombreux formats d'enregistrement. Ces formats peuvent être regroupés selon :

- la nature des informations enregistrées : format textuel (TEXT, UNICODE, ASCII, RTF, etc.) format image/graphique (TIFF, BMP, PICT, GIF, JPEG, etc.), format vidéo (JPEG, M-JPEG, MOV, MPEG, AVI, etc.), format son (WAVE, DAT, MP3, RM, etc.).

- le niveau d'encodage :
 - ouverts (si les spécifications du logiciel sont publiques).
 - fermés (si les spécifications du logiciel sont tenues secrètes par le développeur du logiciel).
 - propriétaires, c'est-à-dire spécifiques à une application d'un fournisseur donné. Les formats propriétaires sont sous le contrôle exclusif du producteur du logiciel. Les producteurs sont ainsi libres de publier ou non les spécificités du format utilisé par leur logiciel et de faire évoluer leur format comme bon leur semble d'une version de logiciel à une autre.
 - standards ou normes, c'est-à-dire produits et reconnus par un organisme de normalisation (ISO, AFNOR, W3C) ; ils sont ouverts et non propriétaires.

- le type d'utilisation :
 - format natif (ou format d'acquisition) :c'est-à-dire le format dans lequel le document a été produit. Exemple : tous les formats propriétaires dont l'exemple le plus courant est le document créé et enregistré par le logiciel Word (propriétaire et fermé). L'information contenue dans ce type de fichier ne devient lisible que grâce à l'interaction de son encodage avec le logiciel qui l'a créé et du matériel informatique permettant de l'interpréter.
 - format d'échange : c'est le format dans lequel le format natif a été transcrit pour en permettre l'interprétation et la modification par d'autres logiciels. Exemple : RTF, PDF.

- format de diffusion ou d’affichage : le format dans lequel le format natif a été transcrit, pour en permettre l’affichage, et éventuellement l’impression, par d’autres logiciels. Exemple : HTML, PostScript.
- format de conservation : format dans lequel le fichier natif (éventuellement à partir d’un format d’échange) a été transcrit pour en permettre la conservation à long terme.

En ce qui concerne les formats destinés à la conservation à long terme, nous devons souligner que, suite à plusieurs études effectuées dans le domaine des archives et à l’observation des tendances dans le domaine informatique, il a été constaté qu’aucun format n’est en mesure d’offrir une garantie d’accès des documents à très long terme. Cependant, on s’accorde sur certains critères dont on doit tenir compte pour faire un choix judicieux et ce, en fonction des besoins et des capacités de l’organisation :

- on choisira donc, de préférence, des langages et formats standards ou faisant l’objet de recommandations internationales.
- en cas d’impossibilité, on retiendra des formats ouverts (PDF pour les documents textuels, TIFF, ou PNG pour les images) et on évitera les formats fermés qui n’offrent aucune garantie de pérennité et qui impliquent les services d’archives à devenir de véritables musées de technologies informatiques pour être en mesure de récupérer les données enregistrées dans ce type de format.

Le lecteur se référera à l’annexe 1 pour connaître les différents formats qui sont présentés comme standards, leurs caractéristiques de même que leur statut comme format pour la conservation.

2.2.3 Le stockage et la conservation des données

□ la puissance des technologies, on doit associer la production d’une quantité d’information de plus en plus importante. Ces masses d’informations doivent être conservées, temporairement ou définitivement. Il faut donc en organiser le stockage ou l’archivage sur des supports appropriés. □ l’instar des autres composantes informatiques, les supports dédiés au

stockage et à la conservation des données ont grandement évolué. Fondamentalement, on utilise aujourd'hui trois types de support de données ; magnétiques, optiques et magnéto-optiques.

Les supports magnétiques

Les supports magnétiques ont recours à des matériaux magnétiques enduits sur des bandes de plastiques ou des disques rigides. Le principe de l'enregistrement des données fait appel à un courant électrique pour magnétiser une petite zone de matériau magnétique. Les données sont représentées par la présence ou l'absence d'une impulsion magnétique. Les supports magnétiques fournissent une grosse capacité de stockage pour un accès rapide. En revanche, ils présentent une durée de vie limitée impliquant nécessairement un recopiage périodique ou une migration en fonction de l'évolution de la technologie. Parmi les supports magnétiques les plus courants, on retrouve : les bandes et les rubans magnétiques, les cassettes, les disques magnétiques rigides et souples (disque dur et disquette, disque amovible style « zip ») et les cartouches.

Les supports optiques

Les supports optiques utilisent le balayage par rayon laser qui effleure des microcuvettes (1-2 micron) dont le changement de réflectivité permet le codage de données. Cette technologie a été utilisée pour le compact disc audio introduit en 1983. Ce compact disque a été rapidement utilisé dans le monde de l'informatique sous le nom de CD-ROM par analogie avec les ROM (Read Only Memory) car il n'était pas réinscriptible. Toutefois, les sons, les images et surtout la vidéo numérique nécessitant des capacités beaucoup plus importantes, les principaux fabricants mondiaux se sont mis d'accord fin 1995 sur le standard DVD qui permet de stocker 4.7 milliards de caractères par face soit une capacité dix fois supérieure à celle du CD-ROM pour la même surface.

Les supports optiques les plus utilisés sont le CD-ROM (Compact Disk Read Only Memory) et les disques optiques numériques DON WORM (Write Once Read Many) ou CD-WORM aussi appelé CD-R. Par rapport aux supports magnétiques, les CD-ROM ont une capacité de mémoire relativement limitée et un temps d'accès plus lent. Si les informations ne

peuvent être gravées qu'une seule fois, elles peuvent en revanche être lues plusieurs fois. En comparaison des supports magnétiques, le CD-ROM a une durée de vie significativement plus élevée. Les disques WORM, contrairement aux CD-ROM, permettent plusieurs enregistrements successifs. Cependant, les secteurs du disque une fois écrits ne peuvent être réutilisés. Par leur grande capacité de stockage, les supports WORM sont utilisés pour sauvegarder des bases de données volumineuses.

Les supports magnéto-optiques

La technologie magnéto-optique combine des propriétés optiques, thermiques et magnétiques sur des supports permettant l'effacement et la modification de données. L'exemple le plus courant de ce type de support est le CD-RW (CD- Rewritable) sur lequel il est possible de réécrire des données jusqu'à 1 000 fois de manière répétitive. Il fournit aux utilisateurs un média haute densité, sûr, durable et à un faible coût.

Les microformes

Les microformes comprennent principalement les microfilms, les microfiches et les sorties d'ordinateurs sur microfilm (COM). Cette technique permet à l'information contenue dans l'ordinateur d'être imprimée directement sur microfilm à l'aide d'une unité d'impression. Cet appareil se compose habituellement d'un écran cathodique étroit qui affiche la sortie d'ordinateur une ligne à la fois. Le film est en contact avec l'écran et se déplace en synchronisation avec l'affichage. Cette technologie d'impression sur microfilm offre un moyen rapide et efficace de stocker des documents. De plus, les microformes ont la particularité d'avoir une très bonne durée de vie, soit environ une centaine d'années dans de bonnes conditions climatiques et matérielles. Elles ont aussi l'avantage d'être lisibles sans intervention d'ordinateurs. En contrepartie, la microforme prive des facilités de traitement et de recherche que présente le numérique. Il peut devenir fastidieux de rechercher une image en particulier.

Le choix d'un support.

Le choix d'un support doit tenir compte de plusieurs considérations dont les principales sont :

- Le niveau d'intégrité requis pour les données. On doit privilégier un support qui donne la meilleure garantie. Pour les supports optiques, c'est la technologie du disque optique WORM, qui empêche toute modification des données enregistrées.
- Les moyens financiers disponibles. Le niveau de ces moyens conditionnera directement le rythme de renouvellement des supports ; le minimum indispensable (c'est-à-dire le niveau en dessous duquel il est illusoire de prétendre faire de l'archivage électronique) est ce qui permet d'effectuer l'opération deux ans au moins avant la date de péremption des supports. Il faut donc tenir compte des coûts impliqués par les migrations de données d'un support à l'autre.
- Le volume des données à conserver à long terme et leur fréquence d'utilisation. On doit faire un choix qui concilie à la fois la capacité de stockage et la facilité de communication. La taille des supports doit en principe être adaptée à la taille des fichiers : un seul fichier conservé sur un support qui lui est bien adapté quant au volume, est plus rapidement repéré par un système de gestion automatisée, que de nombreux petits fichiers stockés sur un support de forte densité.
- La longévité. Ce critère est certainement à considérer mais il ne doit pas primer sur les autres car même si un support offre une bonne longévité, rien ne garantit que les matériels de lecture supporteront encore dans quelques décennies leur format. Car un support de bonne qualité dans des conditions optimales de conservation peut être conservé sans altération, alors que sa technique et ses outils de lecture seront devenus obsolètes et difficiles à trouver sur le marché.

En raison de leur capacité de recherche, les supports optiques sont de plus en plus privilégiés pour la conservation à long terme. Plus particulièrement, le support optique à technologie laser de type non-réinscriptible (ou de type WORM-Write Once Read Many) est à privilégier car la déformation de la surface sensible du support est définitive et irréversible ce qui assure l'intégrité des données. Le support à utiliser actuellement est *le Compact Disk Enregistrable ou CD-R* qui présente les avantages suivants :

- compatibilité du support avec la grande majorité des lecteurs de CD ;
- fiabilité du support (une conservation d'une dizaine d'années est observée dans la mesure où la conservation est effectuée dans des conditions normales d'humidité relative et de température,
- faible coût,
- normalisation,
- une capacité située entre 550 et 600 Mo (millions d'octets) de données utiles.

Plus récemment, est apparu le disque numérique *DVD-R* (Digital Video Disk) qui offre les mêmes avantages que le CD-R mais avec des capacités de stockage beaucoup plus élevées soient de 5 à 20 Go (milliards d'octets).

Vu les rapides développements dans le domaine des supports de données et des normes d'enregistrement, il sera nécessaire de prévoir des conversions à intervalles réguliers des documents électroniques pour une conservation à long terme.

2.2.4 Le repérage/consultation et diffusion

La recherche, la consultation et la diffusion permettent d'exploiter l'information ou le document électronique. Lorsque le document a été numérisé, indexé et stocké sur mémoire électronique, il est destiné à être recherché pour consultation, traité à nouveau et éventuellement transmis ou diffusé à d'autres destinataires. Ce sont les buts ultimes de la GEIDE.

La consultation et la diffusion des documents recouvrent plusieurs technologies pouvant être combinées selon le besoin. Elles comprennent l'affichage à l'écran, l'impression (par imprimantes) et les différentes formes de communication sur réseau local ou distant, comme la messagerie ou le Workflow. La consultation et la diffusion de l'information fait généralement suite à un repérage ou une recherche dont la qualité dépend de l'indexation conçue au moment de la saisie de l'information. Cette indexation est le cœur de la GED: elle doit être la représentation la plus fidèle et la plus exhaustive du contenu et du sens du document afin de faciliter les

recherches et d'en garantir la pertinence. L'indexation d'un document est l'opération qui consiste à décrire le document afin de permettre une exploitation ultérieure sans nécessairement recourir à la consultation du document lui-même. L'enjeu de l'indexation est donc, compte tenu de l'exploitation qui sera faite ultérieurement des documents, d'en exprimer le contenu sous la forme de descripteurs⁹. Cette opération fait appel à différentes techniques et outils¹⁰ (langages d'indexation libre ou contrôlés) pour effectuer l'indexation. Cette dernière peut être manuelle lorsque l'utilisateur saisit les descripteurs suite à l'analyse du contenu ou automatique lorsqu'on a recours à un procédé qui effectue automatiquement le traitement du contenu du document (par exemple la technique d'indexation « plein-texte »).

La finalité des techniques d'indexation ou de représentation de contenu est de constituer une base de données des différents éléments de description des documents permettant d'établir les liens entre descripteurs et documents. C'est à partir de cette base de données que se feront les recherches et les sélections de documents.

Comme c'est le cas pour l'acquisition d'un logiciel de gestion des documents et des archives, la mise en place d'une solution ou d'une application GED ou GEIDE implique un travail de préparation considérable. Elle doit être l'aboutissement d'un long processus d'analyse qui doit tenir compte d'une multitude de facteurs et surtout des impacts potentiels sur les méthodes de travail à l'intérieur de l'organisation.

Comme bon nombre d'applications informatiques, la GED ne reste pas à l'écart des bouleversements provoqués par la montée en puissance des réseaux internet, intranet et extranet. La majorité des éditeurs de programmes de GEIDE ou de Workflow ont conçu des passerelles entre leurs logiciels et la consultation via ces réseaux de manière à ce la GEIDE devienne une composante du système d'information et de communication de l'organisation.

⁹ Un descripteur est un « terme (mot ou expression) retenu, parmi un ensemble de synonymes, de quasi-synonymes et de termes apparentés, pour représenter, de manière univoque, un concept susceptible d'intervenir dans des documents et des questions traités dans un système documentaire donné et dès lors, inclus dans le thésaurus de descripteurs de ce système » définition de l'AFNOR citée dans *La gestion électronique documentaire*, Jean-Yves Prax, Interéditions, 1998.

¹⁰ Le module 4 intitulé « Traitement et description des documents » fournira à l'étudiant les informations concernant les techniques d'indexation.

3. Les problèmes et défis reliés aux documents et aux archives électroniques

Le papier comme support aux documents d'archives existe depuis des siècles. Sa présence domine encore largement le paysage des organisations et ce, en dépit du fait que les ordinateurs personnels envahissent de plus en plus les environnements de travail. Cette présence soutenue du papier peut cependant camoufler une réalité : en parallèle au papier, on produit de plus en plus de documents sur support électronique et ceux-ci ne sont pas systématiquement pris en compte par les systèmes de gestion documentaire existants lesquels sont majoritairement conçus et orientés vers la gestion des documents sur support papier. Ainsi, un nombre croissant de courriers électroniques, de transactions électroniques produits dans le cadre du commerce électronique, de pages WEB, de bases de données et de documents d'archives électroniques créés par les multiples applications informatiques et utilisés dans la conduite des affaires courantes de l'organisation ne sont pas intégrés dans le système de gestion documentaire de l'organisme. La pratique la plus courante et la plus observée en ce qui concerne les documents électroniques est de les imprimer pour ensuite les gérer comme tous les autres documents papier. L'intégration du document électronique dans le système de gestion documentaire dépend donc ici de la décision personnelle du créateur d'imprimer ou non un document. Dans ces circonstances, on peut supposer qu'un bon nombre de documents électroniques ne sont pas intégrés au système de gestion documentaire et que l'organisation se voit perdre à jamais des informations dont elle pourrait avoir besoin comme valeur de preuve et de témoignage.

Les divers projets¹¹, efforts et études théoriques et pratiques menés dans le domaine des documents électroniques jusqu'à aujourd'hui nous démontrent qu'il existe d'ores et déjà des moyens efficaces pour gérer et conserver efficacement les documents produits sur support électronique tout en préservant leur valeur de preuve et de témoignage. Cependant, la clé du succès de ces projets réside dans une bonne connaissance et compréhension de la nature

¹¹ Projet InterPARES

VERS - Cadre des normes, des conseils et des projets de mise en place dont le but est de garantir un archivage fiable et authentique des enregistrements électroniques créés ou gérés par le gouvernement de l'état de Victoria en Australie

DIRKS – *Design and implementation of recordkeeping systems* des Archives nationales d'Australie

Travaux du Public record Office (PRO) en Grande Bretagne

Travaux du Comité sur les documents électroniques du Conseil International des Archives (CIA)

MOREQ – Model Requirements for the Management of Electronic Records de la communauté européenne

document électronique et des systèmes informatiques à l'origine de leur création et d'un changement d'attitude des archivistes face à ces documents en acceptant que :

- le papier ne soit plus considéré comme le support principal de communication et de conservation de l'information.
- les documents électroniques soient considérés comme les originaux à consulter et de les reconnaître comme des sources d'informations très importantes (toutes aussi importantes que le papier) puisqu'ils supportent aussi les activités et la mission de l'organisation.

Dans cet esprit, il est également nécessaire d'adapter nos méthodes de travail afin de mettre en place des programmes de gestion documentaire qui tiennent compte des particularités des documents électroniques.

3.1 Qu'est qu'un document électronique ?

Suite à leurs travaux de 1997, le Comité sur les documents électroniques, sous la responsabilité du Conseil International des Archives (CIA), a développé une définition du document qui tient compte des nouveaux défis posés par l'arrivée et la prolifération des documents électroniques :

« Un document¹² est une information enregistrée qui est produite ou reçue dans le cadre de la mise en œuvre, de la réalisation ou de l'achèvement d'une activité institutionnelle ou personnelle et qui englobe le contenu, le contexte et la structure permettant de prouver l'existence de l'activité »
(Conseil International des Archives, p. 23).

Par cette définition on peut constater qu'un document d'archives se distingue de l'ensemble des ressources informationnelles de l'organisation par le fait qu'il :

- 1- est le reflet d'un processus d'affaires, d'une transaction ou d'une activité individuelle. Il est ainsi la conséquence ou le produit d'un événement.

¹² Le mot document de cette définition est traduit du mot anglais « record ». Le terme anglais « record » désigne les documents considérés dans leur dimension de preuve, par opposition à « document » qui ne prend en compte que leur contenu informatif et à « archives » qui vise la portée historique.

2- fournit la preuve de ce processus, transaction ou activité.

Un document électronique répond donc à cette même définition mais il se distingue des autres documents par le fait qu'il est consigné sur un support électronique sous forme de symboles soient des chiffres binaires 0 ou 1. □ première vue, cette différence semble assez banale mais elle implique une toute autre façon de faire en ce qui concerne la gestion de ces documents puisque dans un contexte électronique, on voit l'émergence de nouveaux types de documents tels les pages Web, les bases de données, etc. De plus, la notion de document change. Un document électronique peut être constitué d'un ensemble de fichiers déposés dans des répertoires différents et/ou produits dans des formats variés, textes images, sons. Souvent créés dynamiquement, le document est assemblé à partir de différentes sources mais il ne prend existence qu'une fois toutes ces sources réunies, c'est les cas notamment d'une page web. De plus, on peut également constater que l'unité de gestion qui, sur support papier, est le dossier, peut passer au niveau plus détaillé du document.

3.2 Les caractéristiques et particularités du document électronique

Pour bien comprendre les documents électroniques, il est nécessaire d'identifier les caractéristiques et particularités qui les distinguent des autres documents. Voici les principales :

3.2.1 Leur matérialité ou immatérialité.

Contrairement au document papier qui est directement accessible et lisible par l'homme, le document électronique est un ensemble de codes, des chiffres binaires 0 ou 1, dont l'accès ou la lecture fait obligatoirement appel à des équipements informatiques et des logiciels pour interpréter ses codes. Un document électronique est donc « invisible » et « virtuel » jusqu'au moment où il est enregistré sur un support, magnétique ou optique et qu'il soit décodé par le matériel requis pour ensuite être capté par un écran d'ordinateur ou imprimé sur du papier.

3.2.2 Leur dépendance au matériel et au logiciel

Les documents électroniques sont obligatoirement créés par des équipements informatiques et par une grande variété de logiciels tels les traitements de textes, les tableurs, les bases de données. Or, ces équipements sont en perpétuels changements. Ils deviennent rapidement obsolètes et les documents risquent éventuellement de ne plus être accessibles et lisibles si les équipements qui ont permis leur création ne sont plus disponibles. Les documents électroniques sont donc très dépendants de solutions technologiques particulières. □ titre d'exemple, un document créé à partir du logiciel Microsoft Word et enregistré en format Word ne peut être lu que par une version de ce même programme ou d'un autre programme qui permet de reconnaître ses « codes ». Si le fichier est envoyé dans un environnement où le logiciel Word n'est pas disponible alors les informations contenues dans ce fichier deviennent inaccessibles. Ce document est donc entièrement dépendant du logiciel qui l'a créé en raison du format propriétaire de ce logiciel, c'est-à-dire un format dont les spécifications ne sont connues que par le développeur du logiciel.

3.2.3 Leur indépendance envers les supports.

Le contenu et le support d'un document papier sont inséparables. Les documents électroniques, par contre, ne sont pas liés de manière fixe ni durable à un support. Ils peuvent régulièrement être enregistrés sur différents supports ou être transférés d'un support à un autre, (ex. disquette à disque dur, disque dur à disque compact, etc.). Or, les supports (magnétiques ou optiques) sur lesquels sont enregistrés et conservés les documents électroniques, ont une durée de vie beaucoup plus limitée que celle du papier ou du microfilm. Ainsi, afin de préserver l'accès aux informations enregistrées sur un support contenant des documents électroniques, il faut donc prévoir recopier les données sur des nouveaux supports qui vont offrir les meilleures garanties de conservation. Cette solution entraîne nécessairement des coûts parfois très élevés et peut porter atteinte à l'intégrité du document.

3.2.4 Leur intégrité, fiabilité et authenticité

Les documents électroniques peuvent facilement être modifiés, altérés ou effacés sans que personne ne puisse s'en apercevoir. Ils sont également très vulnérables aux dommages accidentels voire même intentionnels. Ils peuvent parfois même être effacés et modifiés d'une seule touche de clavier. Il est donc important de s'assurer que le document électronique conserve son intégrité, sa fiabilité et son authenticité tout au long de sa durée de vie afin de maintenir sa valeur de preuve et de témoignage.

- On dit qu'un document est intègre s'il n'a pas été altéré.
- On dit qu'un document est authentique s'il peut être prouvé qu'il est bien ce qu'il prétend être (auteur et destinataire, date).
- On dit qu'un document est fiable si le contenu peut être considéré comme la représentation complète et exacte des opérations, des activités ou des faits.

Un document électronique ne peut donc être accepté comme preuve d'une activité que s'il démontre, avec un certain niveau de confiance, qu'il est bien ce qu'il prétend être – qu'il est authentique – et qu'il n'a pas été altéré – qu'il est intègre et qu'il est fiable.

Pour s'assurer de la fiabilité, de l'intégrité et de l'authenticité d'un document électronique, il faut tenir compte des trois aspects suivants :

Le contenu du document

La notion de contenu réfère aux données, texte, nombres, images, liens, sons, etc. produits par le créateur du document. Par le fait que le document d'archives constitue une preuve d'une activité, les informations contenues dans le document doivent être une image fidèle de cette activité et l'utilisateur de ce document doit avoir suffisamment confiance en cette réalité.

Le contexte du document

Une des particularités du document électronique est qu'il n'a pas à être entreposé dans un endroit physique (classeur, tiroir, dossier, chemise). Ceci implique que les informations

contextuelles intrinsèques à la plupart des documents papier sont très souvent absentes pour le document électronique. Le contexte à l'origine de la création des documents électroniques et des liens qui les relient avec les autres documents peuvent facilement être perdus. Cette connaissance du contexte est primordiale pour bien comprendre un document. Ainsi une lettre n'a de sens que si l'on connaît le contexte dans lequel elle a été créée. Le créateur et les destinataires de cette lettre comprennent bien le rôle de ce document et sa place qu'il occupe dans les activités courantes et au moment de sa création, mais quel sens ce document aura-t-il dans 100 ans quand celui-ci sera consulté par les chercheurs.

La valeur d'un document, peu importe son support, dérive non seulement des informations qu'il contient mais aussi et surtout de son contexte et des circonstances à l'origine de sa création. Le document acquiert une valeur de preuve que si son contexte est décrit. L'information sur le contexte est l'un des éléments qui fournit la preuve de l'existence de l'activité et détermine la provenance du document. Par le contexte, on connaît les informations relatives aux liens qui existent :

- entre le document (ou l'activité qu'il décrit) et les individus ou l'organisation,
- entre le document et l'activité à l'origine de sa création,
- entre le document et les autres documents de l'activité à laquelle il est lié.

C'est par le contexte qu'il est possible d'appliquer le premier degré du principe de respect des fonds puisqu'il permet d'identifier le créateur et l'activité à l'origine d'un document.

Le contexte permet également d'établir un certain niveau de fiabilité. La notion d'original utilisée avec le document papier devient plus complexe avec un document électronique. Ce dernier est facile à manipuler, à dupliquer pouvant se retrouver ainsi en plusieurs exemplaires, en divers formats d'encodage ou en différentes versions. Comment retrouver et identifier le document qui fait preuve d'une activité ? Dans ces circonstances, leur valeur de preuve d'un document est sérieusement compromise voire même absente. Pour assurer la valeur de preuve des documents électroniques, il est essentiel de documenter le contexte d'utilisation, c'est à dire les processus au cours desquels un document a été accédé, modifié ou communiqué. Le degré de confiance à partir duquel un utilisateur considère d'un document est fiable varie considérablement sur la confiance qu'il éprouve face aux procédures de contrôle (techniques et autres) qui ont conditionné la création d'un document et ses utilisations subséquentes.

La structure du document

Un document sur support papier se caractérise par le fait que sa forme physique contient presque tous les éléments utiles à son identification. Étant donné la dépendance des documents électroniques aux équipements matériels, cette notion de structure n'apparaît que dans la mesure où le document est affiché sur un écran d'ordinateur. La structure d'un document, peu importe son support, se situe à deux niveaux : physique et logique. La structure physique d'un document papier décrit sa mise en page, ses différentes parties du document, leur agencement les uns par rapport aux autres, ainsi que l'ensemble de leurs caractéristiques : police, couleur, gras, italique, etc. La structure physique des documents électroniques reste en grande partie cachée à l'utilisateur. Elle comprend les formats des enregistrements, les règles de compression et de décodage, les appareils d'enregistrement, les procédures de base de données, la définition des masques de saisie et d'affichage, etc. La structure physique, principalement destinée aux ordinateurs, ne permet pas à l'utilisateur de déterminer si un document est authentique.

La structure logique correspond au marquage des parties internes du contenu d'un document (ex. champs d'un tableau, marges, paragraphes, titre, sous-titres, paragraphes, etc) ainsi que l'ensemble des liens hiérarchiques et/ou logiques qui les lient les uns aux autres. Elle permet l'identification de chaque document et la représentation de chacun de ses éléments constitutifs. Habituellement, la structure logique correspond à celle créée à l'écran par l'auteur du document. Afin qu'il puisse être considéré comme complet et authentique, le document doit conserver cette structure.

Comme nous venons de le voir, les archivistes ont plusieurs préoccupations concernant les documents électroniques dont les principales sont :

- la perte ou la destruction de documents qui auraient dû être conservés pour des fins légales, de preuve ou de témoignage ;
- la confusion possible entre différentes versions d'un document dont il peut exister des multiples copies sans savoir laquelle a autorité sur les autres ;
- l'authenticité, l'intégrité et la fiabilité des documents mises en doute en raison de la facilité de la manipulation du contenu des documents ;
- l'insuffisance ou l'absence de contexte qui permet de bien comprendre le document ;

- les changements technologiques (équipements et logiciels) qui peuvent compromettre, à long terme, l'accessibilité et l'exploitabilité des documents électroniques.

Maintenir, au fil du temps, une continuité d'accès aux documents d'archives en format électronique dans le contexte d'un environnement technologique en perpétuel changement et garantir l'authenticité, la fiabilité, l'intégrité et l'exploitabilité de ces documents tout au long de leur cycle de vie, voilà les principaux défis auxquels les archivistes sont confrontés dans cette ère de l'information électronique.

3.3 Les stratégies à adopter ou les choses à changer.

Les caractéristiques et les particularités des documents électroniques influencent les stratégies et les méthodes pour leur gestion et leur conservation à long terme. Ceci implique pour les archivistes à repositionner quelques principes et à revoir certaines de leurs méthodes de travail et en adopter des nouvelles.

«Lorsqu'ils sont confrontés aux défis et aux opportunités que présentent les documents électroniques, les archivistes doivent comprendre qu'il ne s'agit plus d'évoluer mais de changer radicalement, changement visant à acquérir une vision d'ensemble plus large, changement en matière de formation et d'expertise, et changement quant aux tâches et aux rôles».
(CIA, p. 31)

Voici les principaux changements :

3.3.1 L'évaluation et la classification

Au niveau de l'évaluation, la gestion des documents électroniques amène un changement d'approche. Habituellement l'évaluation est appuyée sur un inventaire de documents qui permet d'identifier les types de documents pour ensuite fixer leur délai de conservation. Par cette approche, le calendrier de conservation est généralement organisé en fonction des dossiers ou des séries de documents utilisés par chaque unité.

Dans ce monde moderne où un très grand volume de documents électroniques est créé et où ces documents ne sont que des entités logiques et non physiques, l'approche traditionnelle

axée sur le contenu et sur les documents eux-mêmes s'avère difficile d'application. L'approche privilégiée dans un contexte de l'électronique est celle qui permet de « préserver la preuve ». On s'oriente donc davantage vers une approche d'évaluation dite « fonctionnelle » par laquelle on détermine les opérations, les transactions et les activités dont nous voulons conserver la preuve. Or, il s'avère que la documentation la plus appropriée et la plus complète pour établir cette preuve soit celle provenant de l'analyse de la fonction, l'activité, ou la transaction qui a généré le document plutôt qu'en examinant le document lui-même. Bref, c'est le contexte et non le contenu du document qui doit être le point de départ de notre quête de preuve et par conséquent la valeur des documents.

En ce qui concerne la classification, l'approche fonctionnelle est celle qui devrait aussi être privilégiée pour la gestion des documents électroniques. Tout comme pour l'évaluation, cette approche a l'avantage de bien situer le contexte de création et les cadres de classification issus de cette approche survivent bien aux multiples changements organisationnels qui peuvent survenir au cours de l'existence d'un organisme.

3.3.2 La modélisation des processus d'affaires.

Les nouvelles approches dans les méthodes d'évaluation basées sur les fonctions et les activités de l'organisation s'appuient sur une bonne connaissance des processus d'affaires¹³ d'une organisation ou les activités d'un individu. Pour bien comprendre ces processus d'affaires, les sciences de la gestion ont développé une discipline appelée « analyse de systèmes » dont l'objectif est d'étudier les problèmes et les besoins d'une organisation en vue de déterminer comment les systèmes et les technologies de l'information peuvent mieux résoudre des problèmes et apporter des améliorations pour les affaires. Les résultats de cette procédure sont soit l'amélioration des processus d'affaires, l'amélioration des systèmes d'information ou l'implantation de nouvelles applications informatiques. Le produit d'une analyse de système est un modèle de processus d'affaires qui représente les fonctions, les activités et les transactions d'affaires de même que les intrants et les extrants requis pour effectuer ces fonctions, activités ou transactions. Cet exercice donne généralement lieu à la conception d'un diagramme qui

¹³ Un processus d'affaires est constitué d'activités qui s'enchaînent les unes aux autres dans un ordre logique. Il est caractérisé à la fois par la nature des activités et des enchaînements possibles entre elles.

décompose les fonctions, les activités et les transactions qui ont lieu dans l'organisation et qui identifie comment ces éléments interagissent entre eux au niveau des flux d'information. Un modèle de processus d'affaires permet donc aux archivistes de représenter précisément, quand, où et comment sont créés des documents. Il fournit un modèle conceptuel basé sur une représentation réelle du contexte des activités à l'origine de la création des documents et conséquemment fournit l'information nécessaire pour identifier et définir les documents qui doivent être conservés comme preuve d'une activité.

3.3.3 Les métadonnées

Un des défis majeurs posé par la gestion des documents électroniques est relié au fait que les documents électroniques, contrairement aux documents papier, sont des entités logiques et virtuelles. Conséquemment, les documents électroniques ne peuvent être vus de la même manière que les documents papier dont le contenu, le contexte et la structure font partie intégrante du document. Pour assurer la valeur de preuve des documents électroniques et en garantir l'authenticité, l'intégrité et la fiabilité des documents électroniques, il est donc nécessaire de documenter les aspects concernant leur contenu, leur contexte et leur structure.

C'est essentiellement par les métadonnées qu'il est possible de documenter tous ces aspects. Au sens large du terme, les métadonnées sont des données sur les données. Dans un contexte archivistique on dira plutôt que les métadonnées sont des données (ou informations) sur un document. Une métadonnée devrait être associée à chaque document et elle est normalement constituée d'une liste de caractéristiques qui distinguent un document des autres documents ou des autres types d'information. La métadonnée d'un document peut inclure plusieurs éléments dont les plus courants sont le titre, le nom du créateur, une liste des mots-clés, la date de création, les dates de modification, la localisation, les droits d'accès, les délais de conservations et les modalités de disposition. De plus les métadonnées recueillent les informations qui décrivent la structure du document de manière à ouvrir et lire un document tel qu'il a été originalement créé. Plusieurs de ces éléments peuvent être automatiquement enregistrés par le système informatique, mais la plupart des métadonnées doivent être complétées manuellement. Une métadonnée peut être soit incluse dans le document, l'accompagner ou y être associée par un lien.

Divers modèles de métadonnées ont été élaborés parmi lesquelles on retrouve des modèles axés sur le repérage de l'information comme celui du Dublin Core Metadata Initiative (DCMI) et le GILS (Government Information Resource Locator), des modèles développés expressément pour garantir l'authenticité et l'intégrité des documents d'archives comme celui du Commonwealth Recordkeeping Metadata Standard (initiative australienne), du Pittsburgh Model Metadata, et du modèle canadien « Métadonnées requises pour le tenue des dossiers du gouvernement du Canada » et finalement des modèles destinés à la préservation comme celui du PADI (Preserving Access to Digital Information) de la bibliothèque Nationale d'Australie et celui du OAIS (Open Archival Information System).

Les possibilités et le choix des éléments d'une métadonnée demeurent très nombreux et il peut parfois s'avérer très difficile de déterminer le nombre d'éléments à inclure dans une métadonnée pour permettre d'atteindre un niveau de profondeur et de détails adéquat et suffisant pour à la fois garantir la valeur de preuve des documents et amasser les informations qui pourront être éventuellement utilisées pour la préparation des instruments de recherche. Cependant, les différentes parties d'une métadonnée étant appelées à être complétées par les créateurs de documents (parce qu'ils connaissent les documents qu'ils créent, les créateurs de documents sont les personnes les plus adéquates pour fournir les informations concernant le contexte, le contenu et la structure du document qu'ils créent), les archivistes se voient confronter au dilemme de décider de concevoir une métadonnée simple et facile d'application telle celle proposée par le modèle Dublin Core et risquer de ne pas capter toutes les informations nécessaires pour établir la valeur de preuve des documents à celle de produire une métadonnée qui, au contraire, contient tous les éléments nécessaires pour établir cette preuve mais qui, en revanche, devient une procédure longue et ardue pour le créateur. Les conclusions des expériences de conception et d'application de métadonnées dans les différents projets gouvernementaux et dans le domaine privé qui ont présentement cours dans le monde pourront nous donner un éclairage sur les voies à privilégier dans ce domaine.

3.3.4 Le système de gestion des documents électroniques

Le rôle et les objectifs d'un système de gestion des documents dans un contexte de l'électronique demeurent les mêmes que ceux d'un contexte papier. Tous deux doivent garantir

des documents fiables, intègres et authentiques afin de préserver aux documents toute leur valeur de preuve d'une activité et de témoignage pour les besoins immédiats et futurs de l'organisation de même que pour les besoins de la recherche.

Qu'est-ce qu'un système de gestion des documents électroniques ? Le terme « système » est utilisé pour représenter la mission, les processus d'affaires, les politiques, les pratiques, les mécanismes humains et automatisés qui sont nécessaires pour atteindre les objectifs de produire des documents fiables, intègres et authentiques.

Même si les objectifs d'un système demeurent les mêmes pour l'électronique que pour le papier, les moyens pour y arriver sont différents et surtout, ils ne sont pas encore très bien compris comme c'est le cas pour le papier. Depuis près d'une décennie maintenant, les archivistes se sont attardés à identifier et définir les exigences minimales que doit posséder un système de gestion des documents électroniques pour produire des documents fiables, intègres et authentiques. Parmi toutes les nombreuses exigences identifiées, il semble exister un consensus autour de certains points critiques :

- Le système doit être entièrement conforme aux exigences légales, réglementaires et administratives, aux normes nationales et internationales et aux pratiques et principes reconnues en gestion documentaire.
- Les politiques et procédures concernant le système doivent être suffisamment documentées, que les équipements informatiques et les logiciels utilisés pour la gestion des documents doivent être régulièrement testés et vérifiés afin d'assurer que le système traite et gère les documents conformément aux règles et procédures établies
- Le système doit permettre la saisie et l'enregistrement de tous les documents de même que les métadonnées associées à chaque document.
- Le système doit être en mesure de maintenir et gérer adéquatement les documents, en les protégeant des altérations ou des destructions accidentelles ou intentionnelles.
- Le système doit assurer que toutes les parties composant un document, métadonnées, notes, attachements, etc., sont accessibles, récupérables et gérables pour former une unité complète du document dans sa totalité et il doit permettre l'application du calendrier de conservation.

- Finalement, le système doit assurer l'utilisation future des documents ce qui implique qu'il doit être en mesure de recréer le contenu du document et de ses métadonnées dans un nouveau système sans qu'il y ait perte d'information vitale.

Le nombre et la nature des exigences requises pour la conception d'un système sont en étroite relation avec le niveau de fiabilité, d'intégrité et d'authenticité que l'on désire des documents pour qu'ils puissent remplir leur rôle de preuve. Néanmoins, de plus en plus de pays adoptent des lois qui reconnaissent la validité des documents électroniques comme recevables à titre de preuve dans une cours de justice. Cependant, ces mêmes lois exigent que les systèmes conçus pour la conservation et la gestion des documents démontrent qu'ils sont en mesure de garantir un certain niveau de fiabilité, d'intégrité et d'authenticité des documents qui y sont enregistrés. C'est dans cette optique que fut élaborée la norme ISO 15489 «Records Management», dont la Partie 1 définit les limites dans lesquelles un système de gestion des documents doit être établi, comment il doit être structuré et comment il peut être contrôlé et vérifié. La Partie 2 de la norme contient les lignes directrices et la méthodologie pour l'implantation du système pour être en parfaite compatibilité avec la Partie 1. Cette norme¹⁴ jumelée aux technologies de la cryptographie¹⁵ et celles d'identification, notamment par les certificats¹⁶, contribue à la mise en place d'un système offrant les meilleures dispositions possibles pour assurer l'intégrité et l'authenticité des documents.

3.3.5 La préservation

Quelle est la stratégie pour préserver les documents électroniques ? Cette question est très difficile à répondre. Jusqu'à maintenant, les archivistes, les bibliothécaires et les informaticiens ont bien réalisé les défis qu'engendre le seul fait d'essayer de répondre à cette question sans toutefois réussir à y parvenir vraiment. Cependant, les professionnels oeuvrant dans le domaine de la préservation à long terme ont identifié et s'entendent sur les problèmes qui

¹⁴ La norme ISO 15489 vise à la fois les documents sur support électronique et papier et exclut les documents inactifs (archives définitives).

¹⁵ L'art de la cryptographie consiste à transformer le message, un texte lisible en clair, en un message chiffré et vice et versa

¹⁶ Acte écrit par lequel une personne, en sa qualité d'officier public ou à titre personnel, garantit un fait dont elle a connaissance. En matière de communication en réseau, le certificat est un « document électronique qui referme les justificatifs d'identité d'une entité et qui est signé par l'autorité de certification qui a vérifié ces justifications ».

sont à l'origine de ces défis : le caractère obsolète et éphémère des équipements et du matériel informatiques, la dépendance aux logiciels et la détérioration rapide des supports de conservation. Voici les principales stratégies qui sont envisagées pour relever ces défis.

- la constitution de musée de l'informatique dans lequel tous les équipements matériels et les logiciels ayant été utilisés pour créer les documents sont conservés.
- l'émulation qui consiste à simuler le fonctionnement d'un équipement à l'aide d'un autre plus récent et non conçu primitivement pour cette fonction.
- la migration ou la conversion des données qui consiste à transférer périodiquement les données dans de nouveaux formats avant que les technologies actuelles ne deviennent obsolètes.

Chacune de ces stratégies ou techniques de préservation comporte de nombreux avantages et inconvénients mais, chez les professionnels du domaine, un consensus semble se diriger vers la technique de la migration même si cette procédure nécessite beaucoup de ressources humaines et matérielles et que les risques de perte et d'altération d'information sont élevés durant le processus. Cette technique dont on voit habituellement l'application pour la conversion d'un document codé en un format propriétaire c'est-à-dire un format dont les codes sont sous le contrôle exclusif du producteur du logiciel (ex. Wordperfect) vers un autre format propriétaire (Microsoft Word) a dorénavant recours à l'utilisation de format standard c'est-à-dire produit et reconnu par un organisme de normalisation (ISO, AFNOR, W3C) tel que le HTML (Hypertext Markup Language) et le XML (eXtensible Markup Language). Ces formats étant ouverts et non propriétaires, ils constituent des solutions intéressantes et plus sûres pour la survie des documents. De plus, ces formats répondent aux besoins des organisations dans leur volonté d'échange d'information entre les systèmes puisque les documents codés par l'utilisation de ces formats sont complètement indépendants des technologies matérielles et logicielles. Les professionnels de l'information voient donc dans le format HTML, mais encore plus dans le format XML, une solution efficace et pérenne pour la préservation et l'échange des documents électroniques. C'est ce que l'avenir nous dira car il existe toujours un doute qu'un nouveau format puisse venir prendre la relève du XML.

3.3.6 Le cadre d'intervention pour la gestion des documents électroniques

Rappelons ici qu'il existe deux cadres théoriques d'intervention pour la gestion des documents : celui basé sur le cycle de vie des documents et celui qualifié de « continuum ». Le premier est principalement un modèle américain proposé par Theodore Schellenberg qui décrit la vie d'un document à travers une succession de périodes ou de stades ; la création, la vie active, semi-active et inactive. Ce modèle décrit non seulement ce qui arrive aux documents au cours de ces périodes mais il définit aussi qui (gestionnaire de documents ou archiviste¹⁷) aura la responsabilité de gérer le document pour chacune des étapes. Au moment de la création et de la période active, ce sont principalement les créateurs de documents eux-mêmes qui ont la responsabilité de gérer leurs documents. L'intervention à différents degrés d'un gestionnaire de documents demeure cependant possible. Au stade semi-actif, ce sont essentiellement les gestionnaires de documents qui prennent charge et assument les principales responsabilités de la gestion des documents. Finalement, au stade inactif, ce sont les archivistes qui sont responsables des questions de préservation, description et d'accès aux documents devenus alors « archives ». En résumé, le modèle basé sur le cycle de vie des documents a largement contribué aux Etats-Unis à une nette démarcation des responsabilités entre la profession d'archiviste et celle de gestionnaire de documents (records manager). Dans le contexte de l'électronique ou plusieurs décisions importantes concernant les métadonnées, le classement, l'accès, l'archivage, et le sort final du document doivent être prises au moment de la conception des systèmes dans le but d'assurer la création de documents adéquats, fiables et authentiques, ce modèle d'intervention devient inadéquat et tend à disparaître pour laisser place à un nouveau cadre d'intervention basé sur un « continuum » au cours duquel les archivistes ou les gestionnaires de documents sont appelés à intervenir à tout moment mais prioritairement au moment de la conception et de la création. Alors que le modèle du cycle de vie est plutôt relié aux documents eux-mêmes, le modèle de continuum concerne davantage le système qui permet de gérer les documents. Le continuum réfère ainsi à un mode consistant et cohérent de processus qui s'activent dès le moment de la conception du système de gestion des documents jusqu'à l'étape d'archivage ou de destruction des documents. Ce modèle ou cadre permet donc d'éliminer toute distinction entre gestionnaires de documents et les archivistes au niveau des responsabilités et confirme le rôle et

¹⁷ Le Module 1 contient tous les renseignements qui permettront à l'étudiant de bien comprendre la différence entre ces deux rôles.

la place des archivistes à l'étape de création et de conception des systèmes et non plus uniquement à la fin de leur cycle de vie des documents.

Les changements que nous venons de décrire sont loin d'être exhaustifs mais ce sont ceux qui bouleversent davantage la profession et les tâches de l'archiviste. □ l'intérieur du champ d'action déjà reconnu aux archivistes, ceux-ci doivent nécessairement établir des stratégies, des politiques et des procédures dédiées spécifiquement à la gestion des documents électroniques afin de créer une culture propice à une saine gestion de ces documents. Ils doivent également travailler en étroite collaboration avec les autres professionnels de l'information, les analystes de système, les informaticiens, les administrateurs et les utilisateurs pour arriver à concevoir des systèmes qui sauront répondre aux besoins de l'organisme tout en étant facilement opérables. Les archivistes doivent inévitablement avoir l'appui de la haute direction dans cette démarche et réussir à sensibiliser l'ensemble des utilisateurs à l'importance d'intégrer les documents électroniques dans le système de gestion documentaire pour assurer la poursuite des activités de l'organisation et la constitution de la mémoire organisationnelle. Les archivistes sont également appelés à faire beaucoup de formation pour sensibiliser les membres d'une organisation à bien comprendre la nature des documents électroniques et à promouvoir « les bonnes pratiques » qui inciteront les utilisateurs à intégrer et enregistrer leurs documents électroniques dans le système pour ensuite les gérer adéquatement.

Il faut également souligner que les efforts investis pour arriver à une saine gestion des documents électroniques sont certainement profitables à long terme pour les organisations. En plus d'être en mesure de produire des documents qui supportent la mission et les activités de l'organisme, un bon système de gestion électronique des documents est un pré-requis essentiel et constitue la clé de voûte pour toute organisation qui désire entrer dans le monde du commerce électronique ou celui de la prestation de services électroniques. Les différents projets gouvernementaux d'interopérabilité qualifiés de E-Gouvernement ou E-Gov tels ceux du gouvernement britannique (E-ENVOY), australien (E-PERMANENCE), américain (FIRST-GOV) et français (ATICA) ont bien identifié cette problématique et investissent tous les efforts et les ressources nécessaires pour arriver à mettre en place des systèmes performants pour la gestion des documents électroniques. □ l'issue de cette constatation, nous pouvons donc conclure que les défis concernant la gestion des documents électroniques sont bien présents et identifiés mais la situation est loin d'être sans espoir puisqu'il existe aussi des solutions.

Conclusion

Comme nous venons de le constater, l'avènement et l'évolution des ordinateurs et des nouvelles technologies de communication ont considérablement modifié notre environnement et plus particulièrement notre milieu de travail. Les archivistes doivent inévitablement revoir leurs méthodes de travail ou adapter les plus traditionnelles afin de répondre aux nouvelles exigences créées par la présence de plus en plus répandue de l'ordinateur. Ceci implique plusieurs changements pour les archivistes mais aussi de nouveaux défis puisque, dans cette ère de l'électronique, on voit incontestablement le rôle des archivistes associé à celui des informaticiens, des analystes de système et des autres professionnels de l'information pour la conception de systèmes dédiés à de gestion des documents électroniques. Or, la tendance pour l'avenir est de mettre en place des systèmes qui gèrent à la fois les documents d'archives sur tout support de même que toutes les autres sources d'informations de l'organisation. Ce sont ces types de systèmes d'information qui sous-tendent le récent concept en sciences de l'information de « gestion de la connaissance » (connu en anglais sous l'expression « Knowledge Management »). Au sein de ces systèmes, les documents d'archives sont insérés et traités parmi l'ensemble des autres sources d'information mais ils reçoivent un traitement particulier parce que ces documents sont reconnus pour leur rôle distinctif et exclusif : celui d'être la preuve des activités de l'organisation.

Ainsi, les principes et les procédures qui sont la base d'une saine gestion des archives peuvent aussi s'appliquer à l'ensemble des ressources informationnelles de l'organisation. Ceci confère aux archivistes toutes les compétences et l'expertise pour les rendre aptes à prendre en charge l'ensemble de la gestion des ressources informationnelles de l'organisation. Voici donc l'ouverture d'une nouvelle avenue dans laquelle peuvent s'engager les archivistes afin de faire reconnaître la place qui leur revient au sein des sciences de l'information.

ANNEXE 1 - Tableau des standards pour l'encodage

Composant	Standard	Etat du standard	Commentaires
Jeux de caractères et alphabets	UNICODE	Recommandé	
Types de formats d'affichage des documents	TXT	- Recommandé, - pérenne, - ouvert, - très utilisé	Les fichiers texte sont les plus sûrs dans le temps car les plus simples.
	XML	- Recommandé, - pérenne, - ouvert, - faiblement utilisé	XML (Extensible Mark-Up Language) est un dérivé de SGML. Il a été conçu et promu par l'association W3C (World Wide Web Consortium) qui est le référent dans le domaine Internet.
	SGML	- Possible, - pérenne, - ouvert, - faiblement utilisé	SGML (Structured General Mark-Up Language) n'est pas à proprement parler un format, mais plutôt un langage.
	HTML	- Possible, - pérenne, - ouvert, - très utilisé	HTML (Hyper text Mark-Up Language) est une implémentation simplifiée du SGML et est de fait très utilisé sur les sites Internet.
	RTF	- Possible, - peu pérenne, - pas ouvert, - très utilisé	Le format RTF (ou Rich Text Format) est un format commun au monde bureautique (Microsoft© et Apple©).
	PDF	- Possible, - pérenne, - pas ouvert, - très utilisé	Le format PDF, ou Portable Document Format, est un standard propriétaire dont la diffusion et la promotion sont assurées par la société Adobe au travers de la diffusion gratuite du logiciel de lecture Acrobat Reader. Il s'agit d'un format dérivé du format Postscript.
Echange d'information graphique/images fixes	UIT T4	- Recommandé, - pérenne, - ouvert, - très utilisé	Ce format est un dérivé du format TIFF. Il est le plus répandu grâce à son exploitation par les télécopieurs.

Composant	Standard	Etat du standard	Commentaires
	PNG	<ul style="list-style-type: none"> - Recommandé, - pérenne, - ouvert, - faiblement utilisé 	PNG (Portable Network Graphics). Il vise à remplacer le format GIF. Ce format est promu par l'association W3C.
	TIFF	<ul style="list-style-type: none"> - Possible, - peu pérenne, - ouvert, - faiblement utilisé 	TIFF (Tagged-Image File Format). Il utilise des types de compression permettant de garantir la qualité et l'absence de perte.
	GIF	<ul style="list-style-type: none"> - Possible, - peu pérenne, - pas ouvert, - faiblement utilisé 	GIF (Graphic Interchange Format) est un format très répandu. Ce format étant propriétaire, il y a lieu de veiller à migrer dès que possible les fichiers GIF en fichiers PNG.
	JPEG	<ul style="list-style-type: none"> - Possible, - pérenne, - ouvert, - très utilisé 	JPEG (Joint Photographic Experts Group). Cette norme est très utilisée dans le monde Internet. Elle permet un haut niveau de compression (de l'ordre de 1/40). La contrepartie est une perte d'information.
Echange d'images animées / information audiovisuelle	MPEG2	<ul style="list-style-type: none"> - Recommandé, - pérenne, - ouvert, - très utilisé 	MPEG 2 (Moving Picture Expert Group). Cette norme de compression pour les images vidéo et son fonctionne sur toutes les plates-formes.
	MPEG 3	<ul style="list-style-type: none"> - Recommandé, - pérenne, - ouvert, - très utilisé 	MPEG 3 Cette norme est une évolution de MPEG 2 et utilise uniquement la partie relative au son. Le son est stéréophonique.
	WAV	<ul style="list-style-type: none"> - Possible, - peu pérenne, - pas ouvert, - faiblement utilisé 	WAV (RIFF WAVE). l'origine, format de fichier sonore de Microsoft Windows, il est maintenant élargi à d'autres plates-formes. Son stéréophonique.
Base de données	Texte plat	<ul style="list-style-type: none"> - Recommandé, - pérenne, - ouvert, - très utilisé 	Le contenu des bases de données est à conserver sous la forme d'un texte à plat doté de séparateurs (virgule de séparation).

Composant	Standard	Etat du standard	Commentaires
Plans	CGM	- Recommandé, - pérenne, - ouvert, - faiblement utilisé	CGM (Computer Graphics Metafile). Norme pour la mémorisation et l'échange de données graphiques à deux dimensions.
	STEP	- Recommandé, - pérenne, - ouvert, - faiblement utilisé	STEP (Standard for Exchange of Product Data). Normes pour l'échange d'informations sur les produits industriels. Ces normes sont répandues.
	DXF	- Possible, - pérenne, - pas ouvert, - très utilisé	DXF, format propriétaire utilisé par le logiciel Autocad qui est très répandu.

Tableau tiré du Guide de conservation des informations et des documents numériques produit dans le cadre des travaux français de l'ATICA disponible à l'adresse suivante :

http://www.atica.pm.gouv.fr/interop/CCI/standards/standards_formats_supports.shtml

Listes des lectures complémentaires

Bouthillier, Estelle (1997-1998). « Instrument d'évaluation de progiciels en gestion documentaire ». *Archives (Revue de l'Association des archivistes du Québec)* vol. 29 , No 1, p. 99-121.

Duranti, Luciana (2000). « Bâtir un avenir pour les documents électroniques : le projet InterPARES ». Association des archivistes du Québec, 29^e congrès. *Pour que survive la mémoire vive*. Québec, AAQ.

Levasseur, Denis (1998). « Les solutions réseau pour optimiser la gestion des documents et leur diffusion dans les organisations (Groupware, Worflow, Intranet) ». *Argus*, vol. 27, n°2, p. 18-26.

Lévesque, Michel (1998-1999). « Les documents électroniques et le calendrier de conservation ou les considérations d'un pauvre archiviste qui essaie de suivre l'évolution technologique ». *Archives (Revue de l'Association des archivistes du Québec)*, vol. 30, n°3-4, p. 39-54.

Méta-données requises pour la tenue des dossiers du Gouvernement du Canada.

Saulnier, Carole (1997-1998). « Prolégomènes à la gestion des documents administratifs électroniques » *Archives (Revue de l'Association des archivistes du Québec)* vol. 29 , No 1, p. 57- 68

Bibliographie générale

Auzou, Philippe, dir. (2000). *Informatique et technologies d'aujourd'hui*. Paris : Éditions Philippe Auzou.

Bantin, Philip (2002). «Electronic records Management – A review of the Work of a Decade and a Reflection on the Future Directions». *Encyclopedia of Library and Information Science*, Volume 71, Supplement 34, p 47-81. Aussi disponible à <http://www.indiana.edu/~libarch/ER/>

Conseil international des archives. Comité sur les documents électroniques (1997). *Guide pour la gestion archivistique des documents électroniques*. Paris : Conseil international des archives.

Dhérent, Catherine (2002). *Les archives électroniques. Manuel pratique*. Direction des archives de France.
<http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/DAFmanuel%20version%207.html>

Fondin, Hubert (1998). *Le traitement numérique des documents*. Paris : Éditions HERMES.

Gaudin, F. (2000) «Quelle normalisation pour les documents numérisés en vue d'une conservation et d'une consultation à long terme ? ». *Document numérique*, vol. 4, n°3-4, p. 199-217.

Gauthier, Guy (1997). *Introduction aux matériels et logiciels informatiques*. Ottawa, Canada : Éditions Vermette.

Guide de l'information numérique. Comment traiter les données lisibles par machine et les documents numérique (1997). Édition mise à jour et amplifiée. Luxembourg : Office des publications officielles des Communautés Européennes.
<http://europa.eu.int/ISPO/dlm/documents/gdlines-fr.pdf>

Hofman, Johannes (2002). «An expanding Universe. Metadata and Accessibility of Digital Information». Paper presented at the 3rd DLM Forum titled: *@ccess and preservation of electronic information: best practices and solutions*. Barcelona, May 7-8.

Hollier, Anita (2001). «The Archivist in the Electronic Age». *Help Libraries Webzine*. Issue 3, March . <http://library.cern.ch/HELPW/3/papers/5/>

Internet, Intranet, Extranet : Comment en tirer profit (1998). Montréal (Qué.) : Les Éditions transcontinental.

O’Leary, Timothy J. (2001). *Éléments d’informatique*. 4^e édition. Montréal (Québec) : Chenelière/McGraw-Hill.

Prax, Jean-Yves (1998). *La gestion électronique documentaire. Manager les flux d’information dans l’entreprise*. 2^e édition. Paris : Masson/Interédition.