

## GESTION ELECTRONIQUE DES DOCUMENTS

*Le monde est fait pour aboutir à un beau livre  
(Stéphane Mallarmé)*

### 1. Introduction



*Hervé Haut*

*est ingénieur civil et Docteur en Sciences Physiques. Après plusieurs années de recherche en physique théorique à l'UCL, il a occupé diverses fonctions informatiques dans le secteur privé. Il a rejoint la SmalS-MvM en 1998 comme consultant à la section des Recherches où il effectue principalement des missions de consultance pour des projets de gestion documentaire.*

*Contact : 02/509.58.16*

L'objectif de ce document est de donner un aperçu général de l'ensemble des outils que l'on peut être amené à utiliser lorsque l'on entreprend la mise en place d'une gestion électronique des documents. Nous verrons comment ces outils s'intègrent entre eux pour gérer au mieux les documents d'une entreprise, y compris dans l'aspect dynamique de cette gestion avec, parfois, des influences sur les procédures de travail.

#### 1.1 Un peu d'histoire

Depuis plus de deux mille ans, l'écrit a constitué le principal véhicule de transmission de la connaissance humaine. Depuis la pierre de Rosette et les manuscrits de la Mer Morte (datant d'environ deux siècles avant notre ère), l'homme a développé de multiples moyens pour immortaliser sa pensée (papyrus, vélin, tissu, papier, gravures sur pierre, sur bois puis sur métal...). L'invention de la typographie par Gutenberg a permis de développer progressivement les techniques d'impression et nous disposons aujourd'hui d'un nombre incalculable de documents témoins de notre histoire.

Tôt dans notre histoire, est apparu le souci de rassembler, d'organiser et de gérer les documents jugés importants aux différentes époques. C'est ainsi que se sont développées nos grandes bibliothèques, de la légendaire Bibliothèque d'Alexandrie aux prestigieuses bibliothèques d'aujourd'hui. C'est également ainsi que se sont développés les scriptoria, ces espaces réservés à l'écriture des manuscrits dans les monastères, qui furent le lieu essentiel de la transmission du savoir littéraire durant la majeure partie du Moyen Age.



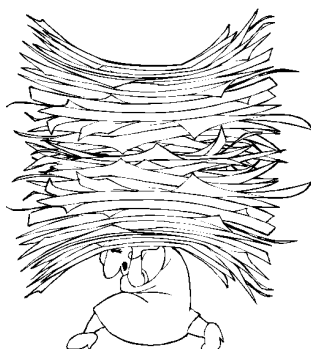
## GESTION ELECTRONIQUE DES DOCUMENTS

Dans nos entreprises, le document est encore aujourd'hui le principal container de l'information non structurée et représente la sève de nos organisations, leur capacité à communiquer tant à l'intérieur qu'à l'extérieur. Et pourtant, malgré leur importance stratégique, les documents restent une des ressources les moins bien gérées et les moins automatisées dans l'entreprise, par ailleurs de plus en plus informatisée.

C'est, paradoxalement, le développement de la technique et de l'informatique qui a conduit à cette situation.

Avant l'apparition de la machine à photocopier, on trouvait encore, dans la plupart de nos entreprises, un ou plusieurs archivistes garants de la bonne gestion des archives papier et, en quelque sorte, gardiens de la connaissance de l'organisation. Le développement de la photocopieuse a permis de multiplier aisément les documents, de se libérer de la contrainte d'un archivage unique et rigide mais a conduit par ailleurs à des archivages multiples et redondants, souvent synonymes de désorganisation de la documentation et de dispersion de la connaissance. La fonction d'archiviste a pratiquement disparu, chaque employé consacre une partie de son temps à cette fonction, avec sa propre méthode, avec une motivation plus personnelle qu'organisationnelle.

### 1.2 La GED, archiviste des temps modernes?



Le développement de l'informatique n'a fait qu'aggraver la situation. Dès l'apparition des ordinateurs, la facilité avec laquelle on pouvait produire des listings a engendré des volumes de papier de plus en plus difficiles à gérer. Une étude de 1996 aux Etats-Unis affirmait "que l'ensemble des entreprises américaines produisait environ 90 milliards de documents par an, que ces documents étaient photocopiés en moyenne 11 fois conduisant ainsi à environ 1000 milliards de pages; l'empilement de ce papier atteindrait l'altitude de 300 000 kilomètres!".

La SmalS-MvM recycle chaque année une quantité de papier correspondant à 30 000 pages par travailleur! Nous verrons que l'on pourrait facilement mettre en oeuvre des solutions plus économiques.

On s'est rapidement rendu compte qu'il était nécessaire d'organiser les données dites "structurées" et l'on a vu se développer tous les outils de bases de données pour organiser ce type d'information et la rendre utilisable.

Au niveau des documents produits dans l'entreprise (que l'on qualifie généralement de données "non structurées"), il a fallu attendre les années nonante pour voir se développer et s'implanter progressivement des outils permettant de gérer électroniquement cette masse documentaire.

Dans la suite de cet article, nous allons détailler ces outils de GED (Gestion Electronique des Documents) en passant en revue les principaux modules d'une gestion documentaire. Notre but est de décrire les outils existants, de préciser leur fonctionnalité et de montrer comment ils s'intègrent entre eux. Nous resterons à un niveau général sans entrer dans le détail d'outils particuliers et sans faire de comparaison entre outils de différents constructeurs.

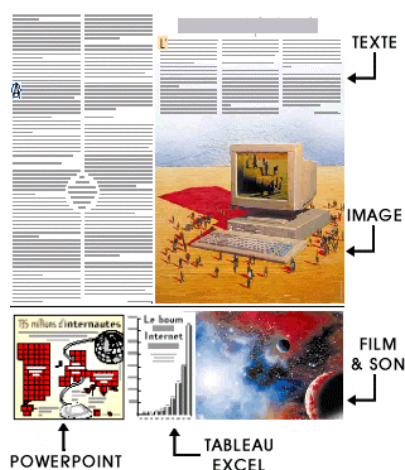


## 2. Le document dans l'entreprise

Avant de parler de la gestion des documents, il est utile de définir ce que l'on entend par document aujourd'hui, de rappeler son importance dans l'entreprise et de mentionner les principales normes actuelles en matière de documents.

### 2.1 Qu'est un document?

Le Larousse donne comme définition d'un document: "*renseignement écrit ou objet servant de preuve, d'information ou de témoignage*". C'est probablement l'image d'un écrit papier qui nous vient d'abord à l'esprit quand nous évoquons l'idée d'un document. Pourtant, le développement des techniques digitales a manifestement étendu les modes de représentation des documents en y incluant les fichiers produits par les outils bureautiques (traitements de texte, tableurs...), les images, la vidéo, le son...: on parlera alors de documents composites.



Au sein d'un même document, on trouvera de plus en plus souvent, des informations provenant de différentes sources comme, par exemple, un texte provenant d'un rapport, avec un tableau construit à partir d'une base de données et des illustrations scannées dans une brochure commerciale.

Ces documents composites ne se limitent plus à être publiés sur papier mais sont souvent conservés et diffusés également sous forme électronique. Dans cette forme électronique, des éléments de navigation (hyperliens) seront souvent ajoutés dans le document (ou dans un ensemble de documents) pour en rendre la lecture plus interactive.

### 2.2 L'importance du document

Même si sa représentation a évolué, le document, au sens large où nous le considérerons dorénavant, garde toutes ses valeurs d'information, de renseignement ou de témoignage et son importance au sein de l'entreprise n'a cessé de s'accroître.

En effet, on estime généralement que l'information dans l'entreprise se partage en 20% de données structurées (bases de données...) et 80% de données non structurées c'est-à-dire de documents. Cette masse d'information non structurée est disséminée dans l'entreprise et est généralement "gérée" par les individus sans motivation collective. De plus, ces documents sont le plus souvent "classés" (une ou plusieurs fois!) et utilisés sous forme papier bien que la grande majorité d'entre eux soient d'origine informatique.

Il peut en résulter des pertes financières importantes si l'on sait que<sup>1</sup>:

- ◇ chaque employé n'utilise que 65% des documents classés dans ses armoires;
- ◇ le classement et la recherche d'un document coûte en moyenne 20\$ et ces coûts peuvent aller jusqu'à 120\$ s'il est mal classé;
- ◇ l'employé consacre entre 10 et 40% de son temps à rechercher ou à attendre de l'information;

<sup>1</sup> Business Week 1995.



## GESTION ELECTRONIQUE DES DOCUMENTS

◇ même si les documents sont gardés sous forme électronique, ils ne sont généralement accessibles qu'à leur auteur.

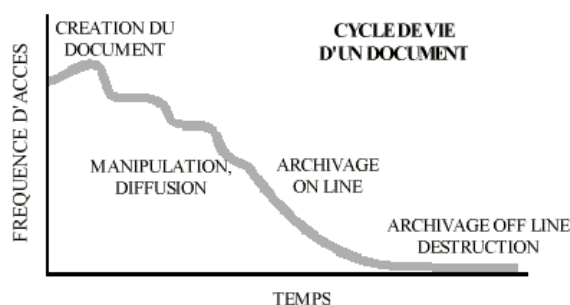
On ajoutera à ces constatations les coûts de stockage des volumes de papier engendrés, le danger de perte d'information capitale pour l'entreprise, la perte d'unicité de l'information suite aux duplications des documents, les risques liés à la circulation possible de documents non validés ou à la diffusion non contrôlée de données confidentielles.

Tous ces coûts et risques sont difficiles à estimer mais, au vu des considérations ci-dessus, il apparaît évident que la plupart des sociétés auraient beaucoup à gagner en mettant tout en oeuvre pour gérer leur information non structurée au même titre que leur information structurée. C'est l'objectif des outils de la Gestion Electronique des Documents.

Nous allons passer en revue ces outils que la technologie moderne met à notre disposition, en suivant le cycle de vie d'un document.

### 2.3 Le cycle de vie d'un document

Tous les documents ont un cycle de vie qui peut être divisé en quatre phases: la création du document, une phase de "manipulation" pendant laquelle le document est finalisé par des mises à jour successives (par l'auteur ou par un ensemble de personnes), une phase au cours de laquelle il est archivé en restant aisément accessible mais sans plus subir de modifications importantes et enfin un dernier stade où il est soit archivé "off line", soit détruit.



A ces différentes phases correspondent différents outils de gestion documentaire, ces outils étant le plus souvent intégrés entre eux autour d'un "repository", c'est-à-dire d'une base de données gérant l'ensemble de la documentation.

## 3. La création des documents

Nous ne nous étendrons pas sur les outils bureautiques dont nous disposons pour créer des documents. Nous utilisons tous un traitement de texte qui permet d'inclure des tableaux de notre tableur, des images ou plus généralement n'importe quel objet provenant d'une autre application. Il existe bien entendu des outils de composition de documents beaucoup plus complexes, utilisés par les spécialistes de l'édition mais nous ne les détaillerons pas ici.

A côté de ces logiciels de création "directe" de documents électroniques, nous détaillerons deux autres outils permettant d'alimenter une base de données documentaire à partir de documents papier ou de listings provenant de mainframes.

Mais avant, nous voulons insister sur l'importance des normes documentaires.



### 3.1 Les normes documentaires

A l'intérieur d'une entreprise, on impose généralement un standard d'outils bureautiques comme les outils Microsoft à la SmalS-MvM. Ce faisant, on résout bien entendu le problème de communication et de partage de documents à l'intérieur de l'entreprise.

Mais en ce qui concerne les garanties de pérennité de l'information, c'est loin d'être suffisant. Nous savons tous que certains documents doivent être conservés pendant parfois un temps très long, particulièrement dans le secteur où nous travaillons. Quelle garantie avons-nous qu'un document Word, stocké aujourd'hui, sera encore "lisible" dans une vingtaine d'années? Est-il raisonnable de penser que Microsoft offrira encore, à cette époque, des outils compatibles avec ses outils d'aujourd'hui? La réponse est bien entendu négative et c'est la raison pour laquelle il conviendrait de choisir un format durable et portable en s'alignant sur des normes "universellement reconnues" pour archiver les documents dont le temps de vie dépasse quelques années.

Il n'est pas toujours aisé de se retrouver dans le monde des standards: il existe de nombreuses organisations (tant nationales qu'internationales) de normalisation, certaines imposant des standards, d'autres se contentant de les recommander. A côté de ces organisations, il y a les sociétés informatiques qui développent également des formats propriétaires que le marché choisit parfois comme standard de facto. Dans la suite de ce paragraphe, nous mentionnerons quelques-uns de ces standards sans vouloir être exhaustif.

En matière de contenu du document, trois standards émergent du marché:

- ◇ **SGML** (Standard Generalized Markup Language) est un standard ISO (International Standard Organisation, une organisation des Nations Unies qui publie des standards). C'est aujourd'hui le standard le plus flexible et le plus sophistiqué pour construire des documents complexes qui soient indépendants de tout vendeur et portables sur la plupart des operating systems. La principale caractéristique de SGML réside dans le fait que les documents ainsi produits se décrivent eux-mêmes: ils contiennent une description complète du corps du document tant au point de vue de la forme que de la structure (par exemple les niveaux de titres, la table des références...).
- ◇ **XML** (eXtensible Mark-up Language) est un sous-ensemble très simplifié du langage SGML. Ce langage est proposé par le consortium W3C (World Wide Web Consortium). Un de ses objectifs est d'enrichir le langage **HTML** (HyperText Mark-up Language) utilisé dans les pages Internet en tenant compte de la structure des documents mais avec moins de lourdeur et de difficulté que SGML.
- ◇ **PDF** (Portable Document Format) est un format propriétaire de Adobe utilisé pour la publication électronique de documents. Il présente l'intérêt de représenter un document dans un format indépendant du logiciel qui l'a produit, du matériel et du système d'exploitation utilisé. Sa facilité d'intégration dans une page Internet, la mise à disposition gratuite de l'outil de lecture (le reader Acrobat) de ce type de document et sa richesse en terme de possibilités de recherche et de navigation en font un standard de fait aujourd'hui<sup>2</sup>. Bien que propriétaire, son format est parfaitement documenté et nous pensons que c'est un format d'avenir pour l'archivage et la publication des documents (lorsque les documents sont finalisés dans leur forme et dans leur contenu).

---

<sup>2</sup> Par exemple, le Gardner Group distribue à ses membres un CD-ROM reprenant, en format PDF, le texte des conférences de son symposium annuel.





## GESTION ELECTRONIQUE DES DOCUMENTS

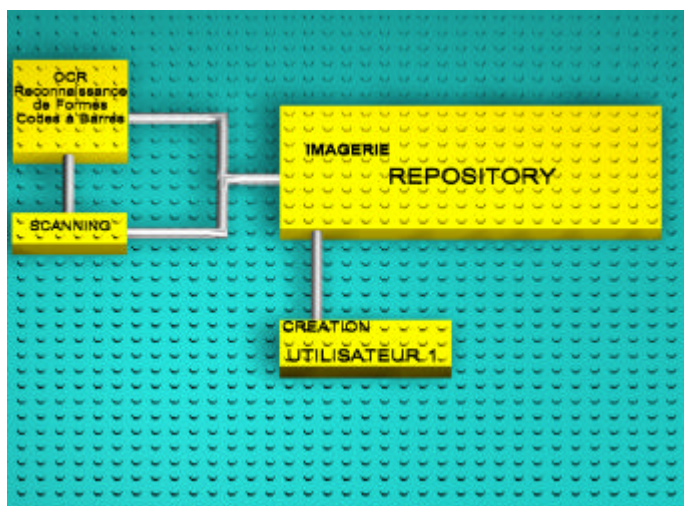
En ce qui concerne les images et l'audiovisuel en général, on notera les standards ISO suivants:

- ◇ **TIFF** (Tag Image File Format) qui est un format utilisé industriellement pour le transfert ou le stockage d'images. Ce format peut se présenter sous forme non compressée ou sous forme compressée sans perte de données (suivant les normes de compression CCITT groupe III et IV).
- ◇ **JPEG** (Joint Photographic Experts Group) est un autre standard graphique qui est compressé avec plus ou moins de perte de données suivant le taux de compression choisi.
- ◇ Enfin, pour ce qui concerne la musique et la vidéo numérique, les divers formats **MPEG** (Moving Picture Experts Group) sont parmi les plus utilisés.

A côté de ces standards liés au format du document, il est utile de savoir qu'il existe d'autres standards relatifs à l'interopérabilité des modules de gestion documentaire, mais nous ne détaillerons pas ici ces standards.

### 3.2 L'imagerie

Au coeur d'un système de gestion documentaire se trouve un "repository": c'est le module qui contient la base de données permettant de gérer l'ensemble des documents; c'est lui qui joue le rôle d'intégrateur des différents autres modules. Nous allons y greffer un premier module d'imagerie (ou de capture de documents) utilisé pour alimenter le système documentaire à partir de documents papier. C'est la première pièce de notre assemblage.



La capture des documents s'avère utile lorsque l'on veut conserver des documents internes qui n'existent que sur papier ou des documents externes reçus sous forme papier. Après scanning, ces documents seront stockés dans le système sous forme d'image. Cette forme de stockage présente le désavantage de consommer beaucoup plus de place disque qu'un document électronique (on compte

généralement 50 kbytes par face papier). Ce format sera cependant préféré pour les documents dont l'existence plus que le contenu présente un intérêt pour l'entreprise (un contrat, une facture, un acte notarié...), ce contenu étant souvent repris dans d'autres bases de données.



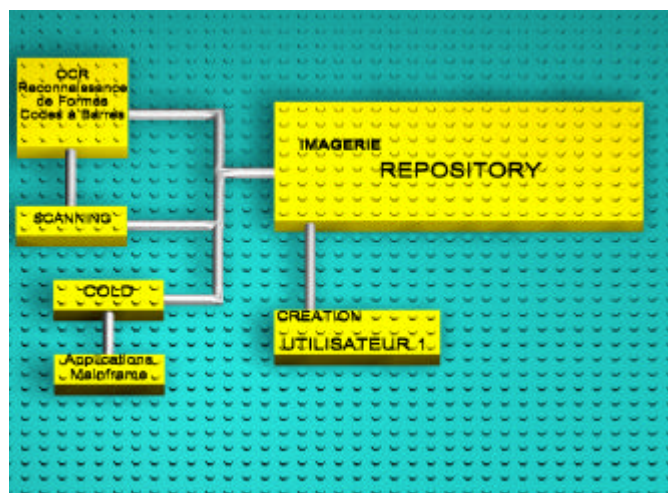
## GESTION ELECTRONIQUE DES DOCUMENTS

Lorsque c'est davantage le contenu du document qui est important, on emploiera diverses techniques pour identifier le contenu du document:

- ◇ L'OCR (Optical Character Recognition) ou la reconnaissance de caractères permet d'extraire le contenu textuel d'un document en produisant un document électronique: cela présente les deux avantages de réduire considérablement la taille du document à stocker et de le rendre exploitable par des outils standards comme les traitements de texte. Les outils disponibles aujourd'hui sur le marché peuvent être considérés comme donnant de bons résultats pour autant que l'on parte de documents de bonne qualité.
- ◇ La reconnaissance de forme sera utilisée lorsque l'on veut exploiter le contenu d'un formulaire. Dans ce cas, on connaît les emplacements (la position physique) du formulaire où se trouvent les informations pertinentes. Les logiciels de reconnaissance de forme permettent d'extraire cette information, généralement pour la stocker directement dans des bases de données permettant ainsi leur exploitation. Ce type d'outil existe pour reconnaître des formulaires imprimés, des formulaires où l'on doit cocher des cases définies et également des formulaires où les zones à reconnaître sont remplies de façon manuscrite; dans ce dernier cas, les taux d'erreurs sont encore assez élevés.
- ◇ Une autre manière de coder l'information d'un document est d'y inclure un code à barres lors de son impression. Les logiciels de reconnaissance de codes à barres donnent d'excellents résultats et permettent très souvent d'extraire du document les "meta-informations" en vue de son indexation dont nous parlerons plus bas.

### 4.3 Le COLD

A côté des documents directement produits sous forme électronique par les outils bureautiques et les documents introduits dans le système après scanning, les outils COLD (Computer Output to Laser Disk) permettent de stocker dans un système documentaire les listings (ou rapports formatés) produits par les mainframes (ou autres ordinateurs).



Lors de l'impression, les grosses imprimantes traditionnelles (comme l'imprimante Xerox du centre de calcul de la SmalS-MvM) effectuent une fusion entre un fond de page et des données issues du mainframe. Les outils COLD fonctionnent d'une manière très similaire.

Lorsque les rapports générés par le mainframe doivent être archivés dans le système documentaire, les données sont redirigées vers le système COLD en même temps que les informations relatives à la structure du document. Plutôt que d'imprimer les rapports sur papier, le système COLD stocke ces rapports sur des disques optiques (d'où son appellation initiale) ou sur n'importe quel media digital.



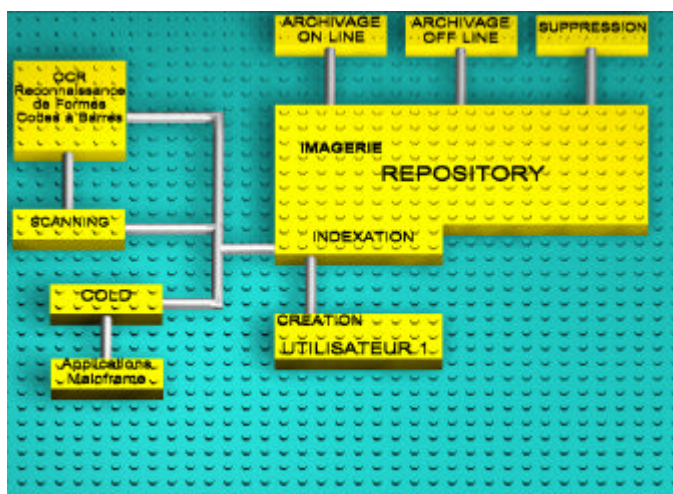
## GESTION ELECTRONIQUE DES DOCUMENTS

Ce type d'outil présente d'énormes avantages et il est étonnant qu'il ne soit pas plus utilisé dans les sociétés qui produisent chaque année des tonnes de listings:

- ◇ Pour chaque lot d'impression, le fichier créé contient le fond de page (qui n'est ainsi stocké qu'une seule fois) et l'ensemble des données du rapport; ce type de fichier est par conséquent peu volumineux.
- ◇ Ce mode opératoire permet de garantir qu'à tout instant, on pourra reproduire le listing tel qu'à son origine, même si, entre-temps, le fond de page a été modifié.
- ◇ Les données étant stockées sous forme digitale, elles peuvent être exploitées par des outils d'analyse de données allant d'un simple tableur à des outils statistiques évolués. Par ailleurs, ce mode de stockage permet également l'indexation automatique des pages du rapport.
- ◇ Il n'est plus nécessaire de distribuer l'entièreté d'un rapport à toutes les personnes qui ne sont généralement intéressées que par une partie sur laquelle elles doivent travailler: grâce aux outils périphériques au système COLD, chacun peut extraire l'information qui lui est utile sans s'embarrasser de données dont il n'a que faire.
- ◇ Il va sans dire qu'un système COLD permet d'éviter un grand nombre d'impressions évitant ainsi de nombreux gaspillages de papier.
- ◇ Enfin, toutes les études montrent qu'un système COLD est rapidement rentabilisé avec un retour sur investissement de l'ordre de 8 à 10 mois!

### 4. Les solutions de stockage

Les systèmes de gestion documentaire sont conçus pour gérer de gros volumes de documents. Il est donc nécessaire de prévoir des moyens de stockage (ou d'archivage) appropriés ainsi que des moyens de suppression à la fin de vie du document. Nous énumérerons ci-dessous les principales caractéristiques des médias utilisés.



L'objectif étant bien évidemment de pouvoir travailler efficacement avec les documents archivés, il importe de les organiser correctement afin d'en faciliter la recherche ultérieure; tout document doit donc être soigneusement indexé avant d'être archivé dans le système.

Enfin, dans tout cet ensemble d'informations, il importe de gérer correctement les droits d'accès afin de garantir la confidentialité nécessaire et de n'ouvrir à chacun que l'information à laquelle il a droit.

Notre construction se complète avec les modules d'indexation et de stockage.





## 4.1 Les média de stockage

Traditionnellement on distingue trois niveaux de stockage en fonction des temps d'accès souhaités.

Les documents auxquels on doit accéder très souvent (essentiellement pendant le temps où ils sont "en gestation") seront de préférence stockés sur disques magnétiques présentant un temps d'accès rapide. Les documents qui doivent rester "on-line" mais pour lesquels on admet un temps de récupération un peu plus long seront archivés sur des cassettes magnétiques ou des disques optiques regroupés dans des juke-boxes (on parle de stockage "near-line"). Enfin les documents que l'on doit conserver pendant un certain temps mais auxquels on n'accède qu'exceptionnellement pourront être stockés sur bandes magnétiques (archivage "off-line").

La vitesse d'accès d'un média étant généralement inversement proportionnelle à sa capacité et à son coût, il s'agira toujours de trouver un juste équilibre entre les performances attendues et les coûts acceptables. Il faut cependant noter qu'aujourd'hui, cette hiérarchie tend à se modifier et il devient rentable de travailler avec des solutions basées exclusivement sur disques magnétiques (et bandes pour l'archivage off-line) même si les volumes de documents sont importants.

D'autres contraintes que performance et coût interviennent dans le choix du média de stockage. Lorsque l'on veut implanter une force probante sur l'information archivée, c'est-à-dire garantir que l'information stockée reste fidèle à l'original et ne peut s'altérer au cours du temps, on aura intérêt à utiliser des media de stockage offrant cette garantie de par leur nature physique: les CD-WORM (Write Once Read Many) sur lesquels on ne peut écrire qu'une seule fois.

Dans un avenir proche, il est probable que la technologie DVD (digital versatile disk) remplacera les disques optiques actuels en offrant des capacités plus importantes. Lorsque la norme des disques DVD-R (write once) sera industriellement reconnue, ceux-ci remplaceront les CD-WORM.

Lorsque l'on parle de conservation des données sur de longues périodes, il ne faut pas cependant pas se leurrer en se focalisant uniquement sur les médias utilisés: même si un constructeur nous garantit qu'un disque optique sera intact dans cinquante ans, aura-t-on encore le lecteur capable de le lire et le format utilisé pour les fichiers sera-t-il encore supporté par les logiciels de l'époque?

## 4.2 L'indexation

Une bonne indexation du document lors de son introduction dans un système de gestion documentaire est une condition sine qua non pour en garantir la facilité d'accès ultérieure.

Nous ne détaillerons pas ici les outils d'indexation "full text" où un index est construit à partir de tous les mots d'un document qui n'appartiennent pas à un dictionnaire de mots à exclure (les "le", "la", "ou"... qui ne présentent aucun intérêt pour la recherche). Ces outils d'indexation et les outils de recherche correspondants pourraient faire l'objet d'une publication séparée.

L'indexation dont nous parlons ici concerne les "meta-données" d'un document qui peuvent se structurer en trois parties, les deux premières fournissant des informations sur le document lui-même, la troisième décrivant le contenu du document:



## GESTION ELECTRONIQUE DES DOCUMENTS

- ◇ Un premier ensemble d'index concerne les informations relatives au document en tant que tel: l'auteur, le propriétaire, la date de création, l'application qu'il faut utiliser pour le visualiser...
- ◇ Une deuxième série d'index décrit le cycle de vie du document: les dates de mises à jour successives, les auteurs de ces mises à jour, les numéros de version, la date d'archivage on-line, la date prévue pour la suppression...
- ◇ Le troisième ensemble d'index a pour objet la description du contenu du document. Pour décrire correctement ce contenu, il convient non seulement de confier cette tâche à une personne qui connaît parfaitement ce contenu mais aussi de l'indexer avec des termes standardisés, des mots-clé préalablement définis. On définira ces mots-clé soit dans des listes simples, soit dans des structures plus complexes de thesaurus, de classification que l'on peut décrire comme des dictionnaires de termes normalisés, organisés de manière conceptuelle, et reliés entre eux par des relations sémantiques.

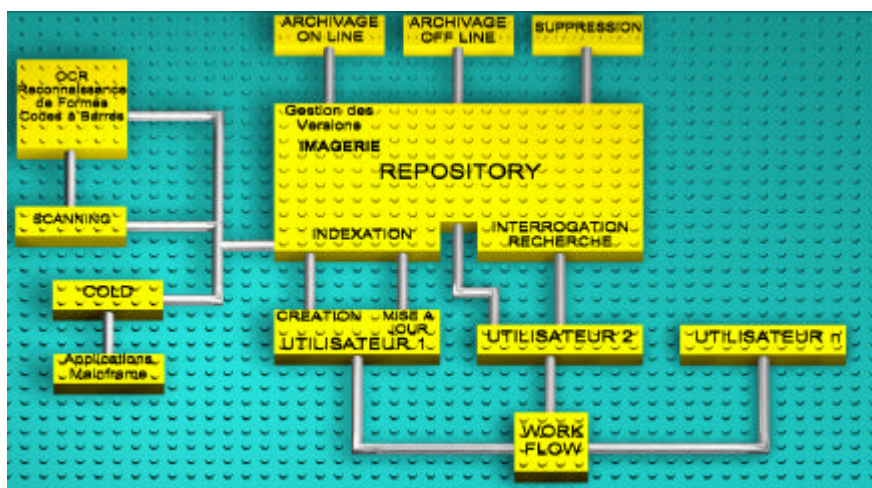
### 4.3 Les contrôles d'accès

Il est important de contrôler les accès à l'information dans un système de gestion documentaire. A chaque document seront donc associés des niveaux de sécurité permettant de définir les utilisateurs habilités à visualiser le document, à le modifier, à l'imprimer, à en couper/coller des parties, à le détruire. Dans des systèmes plus évolués, ces restrictions d'accès pourront même être définies sur des parties du document.

A côté de ces sécurités d'accès, un système de gestion documentaire offrira généralement des outils d'audit permettant de conserver une trace de toutes les actions effectuées dans le système: qui a fait quoi, quand et comment.

## 5. La manipulation des documents

Afin de pouvoir accéder rapidement à l'information souhaitée, il est nécessaire de disposer d'outils de recherche performants et suffisamment élaborés pour cerner au mieux cette information. Dans le début de son cycle de vie, un document est souvent modifié soit par son auteur soit par un ensemble de personnes travaillant en collaboration: des outils de partage et de circulation des documents permettront ces opérations. Les modules de recherche, de gestion de versions et de workflow répondent à ces besoins.



## 5.1 La recherche

L'outil de recherche mis à disposition de l'utilisateur doit lui permettre de retrouver aisément les documents dont il a besoin.

Une recherche "full text" simple (par exemple: retrouver tous les documents qui contiennent le mot "social") n'est pas souvent très utile car la réponse pourra contenir tellement de documents que ce résultat ne sera pas exploitable (c'est le cas des nombreux outils de recherche sur Internet qui vous livrent des milliers de documents en réponse à ce type de question). C'est pourquoi des outils de recherche plus puissants ont été développés: nous n'entrerons pas dans le détail de ces logiciels.

Plus classiquement, les recherches dans une base de données documentaire se feront sur les "meta-données" énoncées plus haut. Les possibilités offertes par ces outils de recherche vont des plus simples au plus complexes:

- ◇ Recherche sur un mot-clé appartenant à la liste des mots-clé autorisés;
- ◇ Opérations booléennes(et, ou, et non...) sur ces mots-clé;
- ◇ Lorsque l'on utilise un thésaurus hiérarchisé, on pourra, par exemple, retrouver tous les documents relatifs à un niveau de la hiérarchie en recherchant sur un mot d'un niveau inférieur: une recherche sur "ouvrier" donnera tous les documents relatifs aux "travailleurs";
- ◇ Des facilités multilingues pourront être offertes par certains outils de recherche: une recherche sur "travail" retrouvera également les documents indexés avec "arbeid".

A ces outils de recherche sont généralement associés des outils qui permettent de manipuler le document (rotation de l'image, zoom...), de les imprimer, d'en sélectionner des parties ainsi que de les annoter par des sortes de "post-it".

Insistons enfin sur le fait que, dans un environnement de gestion documentaire, il nous paraît fondamental qu'un seul et même outil de recherche puisse effectuer les recherches sur tous les documents introduits dans le système quelle que soit leur origine physique (bureautique, COLD ou scanning).

## 5.2 La mise à jour et le partage des documents

Comme nous l'avons déjà dit plus haut, les documents d'origine électronique (traitements de texte, tableurs...) sont généralement insérés dans le système de gestion documentaire avant qu'ils ne soient complètement finalisés.

Les utilisateurs qui ont les permissions appropriées doivent donc disposer d'outils permettant les mises à jour de ces documents en toute sécurité. Il faut en effet éviter que plusieurs utilisateurs modifient simultanément le même document sous peine de perdre, lors de la sauvegarde par le dernier utilisateur, les modifications des autres utilisateurs. Un mécanisme de "check-in/check-out" permet de résoudre ce type de problème: lorsqu'un utilisateur veut modifier un document, il le "check-out" (le verrouille), ce qui le laisse accessible en lecture aux autres utilisateurs mais le bloque en modification; lorsqu'il a terminé ses modifications, il le "check-in" (libère le verrou) de façon à permettre des modifications par autrui.



Lors de modifications, il est souvent utile, surtout si le document est élaboré collectivement, de créer de nouvelles versions à chaque modification afin d'en tracer son évolution. Les outils de gestion de versions permettent d'adresser cette problématique. Dans la plupart des cas, les modifications entre deux versions successives sont clairement identifiées soit pour en notifier les autres utilisateurs soit en attendant l'approbation de ces modifications par la personne en charge de l'acceptation du document.

Certains outils de gestion de versions offrent également la possibilité de gérer des "branches de versions" dans le but de particulariser ensuite le document en fonction de son destinataire: par exemple, le mode d'emploi d'un modèle de voiture fera l'objet d'un seul document avec cependant des parties de ce document qui se présenteront sous différentes versions en fonction des accessoires ajoutés au modèle de base.

### 5.3 Le workflow

La conception d'un document par un groupe d'utilisateurs, telle que décrite ci-dessus correspond à une approche que l'on pourrait qualifier de "pull" (tirer): lorsqu'un utilisateur veut travailler un document, il doit lui-même effectuer la démarche de le rechercher dans la base documentaire et de se l'approprier momentanément pour y effectuer ses modifications. Cette approche présente l'avantage de laisser la liberté d'initiative aux différents auteurs et par là même de favoriser leur créativité.

Il y a cependant des situations où des règles imposent une procédure bien définie pour élaborer un document: on peut imaginer que certaines parties doivent être rédigées par une personne désignée, que la rédaction des différentes parties doive se faire dans un ordre donné, que des approbations de la hiérarchie soient nécessaires à certains stades d'élaboration du document. Dans ce type de situation, on adoptera une approche de type "push" (pousser): lorsqu'une action devra être effectuée par une personne donnée, le document lui sera envoyé avec les instructions adéquates.

Les outils de workflow permettent d'automatiser les procédures de ce type. Pour un document donné (ou plutôt pour un type de document donné), on définit les tâches à accomplir, les intervenants avec leurs responsabilités, les séquences à respecter et, dès que le processus est initialisé, la réalisation du document suit son cours automatiquement.

Suivant leur degré de complexité, les outils de workflow permettront de définir des routages plus ou moins élaborés:

- ◇ des routages linéaires où à chaque étape, le document est envoyé vers une seule personne bien précise;
- ◇ des routages avec branchement où, après une étape donnée, le document est envoyé simultanément à plusieurs personnes;
- ◇ des routages où la personne à qui le document est envoyé est choisie en fonction de sa charge de travail;
- ◇ des routages conditionnels où l'intervenant suivant est choisi en fonction d'un critère lié à l'état du document;
- ◇ des routages avec alternative pour tenir compte, par exemple, de l'absence d'un collaborateur (afin de ne pas bloquer le document à une étape de son circuit);
- ◇ ...



## GESTION ELECTRONIQUE DES DOCUMENTS

La mise en place d'un workflow dans l'entreprise est certainement la partie la plus délicate dans le déploiement d'une solution de gestion documentaire. On évitera donc de mettre en place un workflow là où les procédures ne sont pas bien établies.

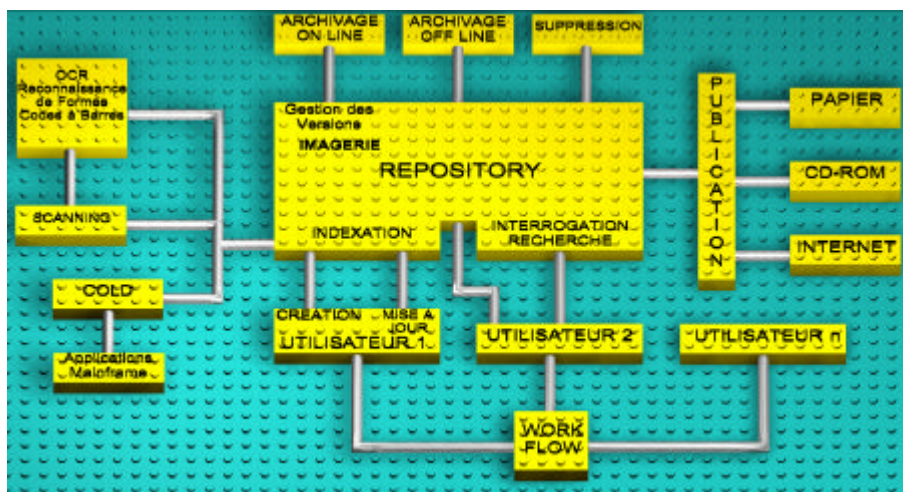
Par contre, les avantages que l'on peut en attendre sont nombreux:

- ◇ rationalisation des procédures de l'entreprise;
- ◇ diminution de leur coût grâce à leur automatisation;
- ◇ gain de temps par la transmission électronique des pièces nécessaires à leur exécution;
- ◇ amélioration du contrôle et du suivi de ces procédures.

L'analyse préalable à la mise en place d'un workflow révèle souvent les défauts d'une organisation existante et la tentation est grande de profiter de cette occasion pour procéder à un re-engineering de cette organisation. Nous pensons qu'il faut évaluer sérieusement le poids d'une telle décision et ne partir dans cette direction que si une volonté existe sur le terrain et si les dirigeants concernés s'impliquent réellement dans le projet. Psychologiquement, il est parfois plus rentable et réaliste de passer par une première phase où le workflow calque les procédures existantes même si elles sont en partie déficientes.

### 6. La diffusion des documents

Nous arrivons à la fin de la construction de notre modèle de gestion documentaire en y ajoutant la dernière pièce: la diffusion ou la publication des documents.



Comme l'indique le schéma ci-dessus, il y a trois voies principales de diffusion des documents: sur papier, sur CD-ROM et sur Internet ou Intranet.

Dans de nombreuses situations, on sera amené à publier les mêmes documents sous plusieurs formes. Il est dès lors souhaitable de disposer de ces documents dans un format électronique tel qu'il soit aisé de réaliser ces différentes diffusions avec la même qualité de présentation.

Dans la pratique, il existe aujourd'hui de nombreux outils de conversion permettant de "traduire" un texte conçu dans un traitement de texte standard en page HTML afin de le diffuser sur Internet. Ces traductions ne sont cependant pas toujours parfaites au niveau de la





## GESTION ELECTRONIQUE DES DOCUMENTS

forme, de la qualité du document produit et l'on peut espérer que les développements futurs de XML permettront de réaliser ces traductions avec une qualité plus grande.

De plus en plus, le format PDF est également utilisé pour publier sur Internet: on constate cette évolution dans les sites d'entreprises où les brochures commerciales, les rapports d'activité... sont de plus en plus souvent publiés sous format PDF; de même beaucoup d'éditeurs commencent à offrir des abonnements via Internet en utilisant ce format pour leurs publications. Outre les intérêts cités plus haut, ce format présente également l'avantage de reproduire fidèlement le document imprimé, de présenter un taux de compression élevé (et donc des tailles de fichier raisonnables pour la diffusion sur Internet) ainsi que d'offrir, depuis la dernière version, la capacité de "streaming" qui signifie qu'un utilisateur ne doit pas attendre d'avoir téléchargé l'entièreté d'un fichier avant de pouvoir en lire les premières pages. Enfin, l'outil permettant de lire ces fichiers soit directement ("reader") soit à l'intérieur d'un navigateur Internet ("plug-in") est entièrement gratuit, ce qui constitue sans aucun doute une raison importante de son succès.

La publication sur CD-ROM peut se réaliser très simplement soit en recopiant sur le CD-ROM les pages HTML déjà réalisées pour Internet, soit en y gravant des fichiers PDF (avec éventuellement le reader associé).

Cependant, il est rare que l'on se contente de ces solutions élémentaires lorsque l'on réalise un CD-ROM: celui-ci offre en effet des possibilités bien plus riches telles que l'insertion de clip sonores et vidéo et sa vitesse d'accès n'est pas un frein à l'utilisation de ces techniques. Pour réaliser de tels CD-ROM, on utilise des outils d'édition ("authoring") développés à cette fin.

## 7. Conclusion

Nous avons passé en revue les différents outils qui, assemblés entre eux, permettent de gérer l'ensemble des documents d'une entreprise.

Beaucoup d'entreprises commencent à se rendre compte de l'importance de leur information non structurée existant sous forme de documents au sens large où nous l'avons défini plus haut. La croissance du marché des outils de gestion électronique des documents montre très clairement cette tendance: en 1998, le marché mondial était estimé à 14 milliards de dollars et les études prévoient une croissance proche de 20% par an dans les quatre années qui suivent, pour atteindre 33 milliards de dollars en 2002.

Cette évolution se retrouve dans tous les secteurs de l'industrie: les banques et compagnies d'assurance où la GED est surtout utilisée pour archivage et pour améliorer le service à la clientèle; l'industrie manufacturière tant pour la gestion des procédures de fabrication que pour la production des manuels et modes d'emploi; l'industrie pharmaceutique où la documentation à soumettre aux autorités en vue d'obtenir l'autorisation de mise sur le marché d'un médicament devient quasi impossible à gérer sous forme papier; le secteur public confronté non seulement à des masses importantes de formulaires à gérer mais aussi à des contraintes légales de conservation des documents.



**GESTION ELECTRONIQUE DES DOCUMENTS**

Même si les outils actuels deviennent de plus en plus conviviaux, il ne faut pas en conclure trop rapidement qu'une gestion électronique de documents est aisée à mettre en oeuvre: il est important de déployer progressivement ces solutions afin de permettre aux utilisateurs de s'habituer progressivement à ces nouvelles méthodes de travail et de les accompagner dans ce changement. On évitera notamment de modifier trop les procédures existantes lors de la mise en place d'un workflow en se contentant de calquer ces procédures dans un premier temps et en les optimisant progressivement par la suite.

La SmalS-MvM a déjà mis en place, pour l'ONSS, un projet de recherche documentaire (Juridoc) pour améliorer l'exploitation de la documentation juridique et va entreprendre prochainement le déploiement d'une solution de gestion documentaire pour gérer, sous forme électronique, l'ensemble de la documentation de l'ONSS relative aux employeurs.

Nous pensons qu'il est intéressant d'étudier les opportunités de mise en place de solutions de type COLD afin de diminuer la production de papier par les mainframes tout en offrant aux utilisateurs de ces listings des outils modernes et conviviaux d'exploitation de l'information.

L'émergence de ces nouveaux outils de gestion de documents a encore un potentiel énorme de développement et l'on peut penser que dans un futur plus ou moins proche, on pourra réellement parler de "minimal paper office", de bureau sans papier superflu.

Ces nouvelles technologies permettront de retrouver au sein des entreprises une partie importante des connaissances qui s'étaient peu à peu dissipées au fil des ans. Mais, pour notre part, nous espérons que le livre conservera sa raison d'être et que les générations qui nous suivent pourront toujours apprécier le plaisir de la lecture d'un livre à peine sorti de presse, avec les sensations que procurent son odeur et le toucher de ses pages fraîchement imprimées.

