

# La GED : générateur de gains à tous les niveaux



**L**a GED (ou GEIDE), le workflow et le traitement électronique de dossiers ou de documents sont entrés dans les pratiques administratives des entreprises et des institutions. Ces outils informatiques s'intègrent désormais dans la majeure partie des environnements; ils se greffent aux applications existantes auxquelles ils apportent de nouvelles fonctionnalités sur la façon de traiter documents numériques et fichiers. Sans parler d'amélioration de la productivité, la GED et ses technologies connexes poussent à rationaliser les méthodes de travail, d'archivage et de consultation des informations et des documents, à réagir très vite sur les dossiers ou tout simplement à gérer sans délai les différents documents générés ou reçus par les entreprises. La GED comme le workflow et le traitement automatique de documents sont générateurs d'économies qui se mesurent par des retours sur investissement rapides. Dans les capacités qu'elles apportent, il y a à la fois sélectivité et sécurité. Elles offrent la possibilité d'accéder de façon sélective à l'ensemble des documents au travers de réseaux locaux ou via internet/intranet, ce qui assure une communication instantanée ou permet de décentraliser une partie des traitements ou l'exploitation des informations. Mais dans le même temps, elles verrouillent l'accès aux données sensibles. Dans cette introduction à la neuvième édition de cet annuaire des fournisseurs de la GED et de l'archivage, nous allons examiner les différentes techniques de GED et les supports d'archivage qu'elles mettent en œuvre.



## **«Organiser, gérer et distribuer les informations et les documents sous forme électronique»**

### **Quelle définition pour la GED ou GEIDE ?**

On rencontre plusieurs définitions et divers acronymes quand on commence à s'intéresser à la gestion électronique de documents et d'informations. Le terme GED pour Gestion Electronique de Documents est apparu dans le milieu des années 80 et s'est substitué au terme "archivage électronique" précédemment utilisé. En 1994, les professionnels français réunis au sein de l'association APROGED ont décidé d'utiliser l'acronyme GEIDE (qui provient de Gestion Electronique d'Informations et de Documents Existants) pour dépasser ce que l'appellation GED avait de restrictif dans son aspect document et élargir la gestion à tout type d'information, en particulier aux données de source informatique. Dans les deux cas, ces acronymes désignent les mêmes technologies et les mêmes solutions.

### **Pourquoi la GED ?**

Le rôle de la GED ne peut pas se résumer aux seuls trois mots de gestion, circulation et archivage. Et contrairement aux idées reçues, elle ne s'intéresse pas seulement aux documents préalablement numérisés. Elle prend en compte tout type de fichiers numériques créés ou exploitables par des ordinateurs ou des systèmes numériques. Dans un ensemble de GED ou GEIDE, il est possible d'associer au sein d'un même dossier des fichiers issus de logiciels de bureautique (traitements de texte, tableurs, etc.) ou de systèmes centraux (comme des fichiers d'impression ou "spool"), des télécopies reçues via des fax-modem, des plans ou des dessins industriels (de type vectoriel ou raster), des photographies numériques, des courriers électroniques, des documents récupérés sur internet ou encore du son, de la vidéo, etc. Cet inventaire est loin d'être exhaustif. L'avantage de la GED est de fédérer les moyens et méthodes qui permettent de classer et de réutiliser intelligemment ces informations. L'un de ces outils a vu son importance croître ces dernières années car de lui dépend la récupération rationnelle de l'information. Cet outil est le moteur d'indexation en texte intégral qui en indexant chaque mot ou composant permet de le retrouver en quelques secondes ou de créer des dossiers électroniques complets associant de multiples types d'informations issues de sources diverses. Le couplage aux bases de données ou aux programmes intégrés de gestion existants dans les entreprises donne au système de GED la maîtrise de l'ensemble des données qui forment le tissu de l'entreprise. Nous sommes loin des traditionnelles archives. Grâce à la GED, le document n'est plus isolé: c'est un élément vivant du système d'information des entreprises. La GED, c'est aussi la possibilité de contrôler qui a accès à quoi grâce à une protection électronique des informations sensibles. Ce sont des erreurs en moins: erreur de saisie, erreur de reclassement lors de manipulations fréquentes de dossiers ou perte par inadvertance ou par malveillance. La GED s'insère dans la majeure partie des systèmes informatiques modernes des entreprises et des administrations tirant avantageusement parti des environnements client-serveur. Elle s'insère sans imposer un bouleversement complet des modes de fonctionnement, sans remettre en cause l'existant mais en l'enrichissant de nouvelles fonctions.

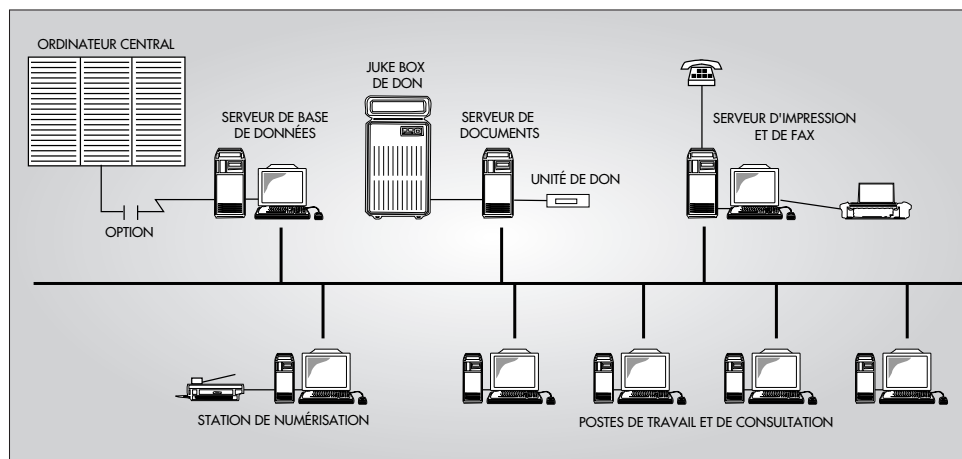
## INTRODUCTION A LA GEIDE

### La GED en complément des applications informatiques

La GED est, au fil des ans, devenue une composante des systèmes d'information des entreprises dans lesquels elle introduit des fonctions de gestion, de partage ou de circulation et d'archivage, fonctions qui viennent seconder des applicatifs et des programmes bureautiques, de gestion ou de traitement de documents. Historiquement, les solutions GED ont pris la suite des systèmes dits "d'archivage électronique" apparus dans le milieu des années 80. A cette époque, les concepteurs proposaient des solutions autonomes ou parfois connectables à des réseaux locaux, dont la vocation principale était de remplacer les salles d'archives en substituant l'image numérisée du document à l'original ou aux microformes (microfilms ou microfiches). Cette vision n'est plus d'actualité. La GED vient désormais se greffer à une structure informatique aux côtés et en complément d'autres applicatifs et de gestionnaires de bases de données. De même, des fournisseurs de systèmes d'exploitation pour micro-ordinateurs vont jusqu'à fournir "gratuitement" des modules de visualisation et de restitution d'images numérisées. C'est le cas pour Windows 95/97 et NT. Il est également possible d'utiliser les possibilités offertes par les technologies intranet/internet pour accéder sélectivement aux bases d'index ou de données puis d'afficher soit les fichiers, soit les images numérisées en utilisant un visualiseur internet classique.

Concrètement, la GED est la réponse aux problèmes de gestion, de stockage, de recherche, de consultation, de traitement et de circulation de fichiers ou de documents réunis sous forme de dossiers ou non. C'est 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 un organisme, qu'il s'agisse d'une entreprise ou d'une administration. La GED est à vocation universelle. Elle trouve son utilité partout, dans tous les contextes et à tous les niveaux. Elle le démontre là où les documents abondent comme dans la documentation technique, les dossiers d'assurances, de prêts ou du personnel, les processus industriels, les dossiers d'allocataires, etc. C'est elle qui permet d'accéder sélectivement à des documents en quelques secondes à partir d'un simple micro-ordinateur connecté à un serveur via un réseau local alors que la gestion manuelle des dossiers aurait nécessité plusieurs minutes, si ce n'est plusieurs heures, pour localiser puis regrouper le dossier. Comme nous le verrons par la suite, la GED comporte dans ses aboutissants des notions de sécurité logique et de pérennité de l'archivage si celui-ci est réalisé sur un support adéquat.

Dans l'évolution de l'informatique, la GED est une étape: celle où il devient possible d'associer des documents numériques de toute nature à des applications bureautiques, administratives, industrielles ou de production. Les solutions GED sont avant tout des ensembles fédérateurs d'outils qui viennent compléter des programmes bureautiques, ou autres, de fonctions de traitement de documents électroniques. Par document, il faut entendre non seulement la feuille que l'on numérise dans un scanner mais aussi le fichier issu d'ordinateurs centraux, le fichier de son numérique, de vidéo numérique, etc., fichiers que l'on peut relier entre eux au sein d'un dossier en créant des liens logiques.



**Schéma de principe d'un système GED/GEIDE dans une architecture client/serveur**

## INTRODUCTION A LA GEIDE

La GED, c'est aussi le partage entre groupes de travail de documents sous forme d'images électroniques afin de restreindre la circulation des originaux. Ce partage va au delà de la simple diffusion. Si on rajoute un programme de workflow à celui de GED, le partage peut être sélectif et permettre la répartition de tâches administratives entre plusieurs agents suivant un circuit préétabli. Le workflow est un outil particulièrement intéressant pour le traitement de dossiers dans le cadre de travaux collaboratifs en groupe; mais aussi pour le cheminement automatisé, sous forme électronique, de documents nécessitant l'intervention de plusieurs agents ou techniciens.

L'objectif de la GED, c'est de faciliter, au besoin de sécuriser, l'accès et la consultation des documents et des fichiers fédérés par une base de données. Quelques secondes suffisent pour accéder à l'ensemble des éléments constituant un dossier, les afficher sur l'écran d'un micro-ordinateur, les imprimer ou les diffuser. Les originaux ne sont plus mis en circulation; ils sont physiquement archivés une fois pour toutes, ce qui évite des pertes ou des mauvais reclassements. Les conditions de travail sont en général fortement améliorées comme le sont les relations avec les clients ou les administrés. Les solutions de gestion électronique de documents et d'informations constituent d'excellents outils pour réduire le coût des tâches administratives.

Les fonctions entrant sous le vocable de GED sont nombreuses et ne sont pas forcément les mêmes d'une application à l'autre, même si l'on retrouve toujours un noyau de fonctions de base. Puisque ce sont des documents ou des fichiers numérisés et non plus des originaux sur papier que véhicule la GED, on peut faire subir à ces images des traitements électroniques et les faire circuler dans l'entreprise par le ou les réseaux locaux ou vers l'extérieur. Mais en dehors de ces images numérisées, la GED sait aussi prendre en charge la gestion, l'indexation et l'archivage de tout type de fichiers numériques. Ceux, par exemple, générés par des logiciels bureautiques, mais également des fichiers "spools" issus de systèmes informatiques centraux ou encore des fichiers sonores ou vidéo.



Les solutions de GED peuvent faire l'objet de diverses classifications mais, par souci de simplification, nous n'avons retenu dans cette introduction que deux grandes catégories: la GED-COLD et la GED-Image. Les personnes désireuses d'approfondir la segmentation des applications GED peuvent consulter les serveurs web de l'APROGED (<http://www.mosarca.com/APROGED>) et du magazine MOS (<http://www.mosarca.com>).

### **La GED COLD pour la gestion de fichiers d'impression ou "spools"**

Sous l'appellation GED COLD (Computer Output LaserDisc) sont regroupés les programmes et les applicatifs conçus pour gérer des fichiers "spools" générés par des ordinateurs centraux ou des systèmes transactionnels. Il peut s'agir, par exemple, de factures, de bons de livraison, de fiches de paie, etc. La GED COLD vient en complément ou remplace les traditionnelles éditions de microfiches COM (Computer Output Microfilm) ou les listings. Une solution de GED COLD se compose de plusieurs modules logiciels qui permettent le formatage, l'extraction de données de fichiers spools en connexion directe avec un ordinateur central ou par lecture d'un support de stockage temporaire tel qu'une cartouche magnétique. Le logiciel de GED COLD réalise, par extraction, l'indexation des données du fichier spool suivant des critères propres à l'application. Ces informations enrichissent la base de données à partir de laquelle l'utilisateur du système de GED pourra effectuer des recherches multicritères sur son poste de travail connecté au serveur via un réseau local. Les applications



Ecran d'une application de GED/GEIDE COLD

**INTRODUCTION A LA GEIDE**

de GED COLD sont d'un usage beaucoup plus souple que les microfilms/microfiches COM ou les listings. La consultation des informations se fait en quelques secondes - en utilisant ou non une interrogation multicritères; le partage des fichiers entre plusieurs utilisateurs est instantané, la protection ou l'accès sélectif aux données est possible par mots de passe, etc. Les informations peuvent être présentées à l'écran sous plusieurs apparences telles que celle d'un listing ou, mieux, sous la forme d'une copie identique à un original par incrustation des données dans un fond de page numérisé. La GED COLD offre enfin la possibilité de mettre rapidement à jour des bases d'informations ou de données en ajoutant de nouveaux fichiers et de communiquer ces données rapidement et à distance au travers de réseaux de télécommunications, intranet/internet compris. Dans la majeure partie des applications de GED COLD proposées aujourd'hui sur le marché, le CD-R/CD-ROM peut servir à diffuser des informations vers des filiales ou des bureaux distants. De même, il est possible d'interroger la base de données puis de visualiser les fichiers en utilisant une connexion intranet/internet.

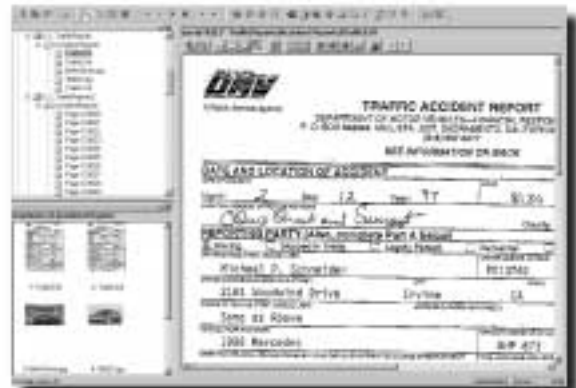
**La GED "Image"**

Sous le terme GED-Image, nous regroupons toutes les applications de gestion électronique de documents gérant des images ou des documents numérisés. Cette segmentation n'est pas restrictive car un même logiciel est en général capable de gérer à la fois des images converties électroniquement et des fichiers issus de traitements de texte, de tableurs, etc. De même, des applications de GED-COLD peuvent aussi, si on leur ajoute les programmes adéquats, gérer des documents numérisés et vice et versa. Pour capturer les images, il existe plusieurs méthodes qui dépendent de la nature des documents et des besoins de l'application.

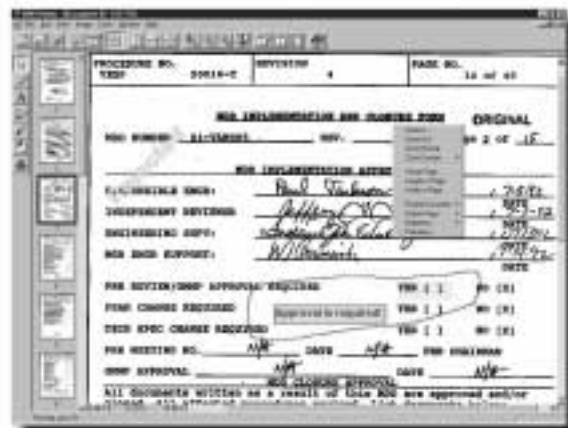
On peut numériser des documents à l'aide de scanners, importer des fichiers déjà existants ou collecter des fichiers-images à partir d'un serveur de télécopie/fax. De même, il est possible de capturer des images à partir d'une caméra vidéo couplée à une carte de numérisation adéquate placée dans le micro-ordinateur.

**La numérisation des images**

Dans le cas d'un système GED dit "image", la première étape consiste à numériser, c'est-à-dire à convertir en image électronique et numérique, des originaux qu'ils soient sur papier, sur microformes, calques, transparents, etc. Cette opération est réalisée par un numériseur ou "scanner" qui, après analyse, transmet à un micro-ordinateur ou à un serveur la copie numérisée des originaux. Ces images peuvent être, soit directement stockées sur un support d'archivage longue durée ou temporairement stockées sur le disque dur magnétique du micro-ordinateur dans l'attente d'un traitement ultérieur. Il existe de nombreux scanners ou numériseurs répondant à un très large éventail de besoins. Si le nombre de documents à saisir quotidiennement est faible, on peut se contenter d'un numériseur dit "bureautique" en gardant à l'esprit qu'il n'est pas fait pour la saisie par lots du fait de sa lenteur. Les scanners conçus pour les applications GED ou de traitement de documents sont capables de capturer et parfois de traiter



Recherche et visualisation de documents numérisés



Exemple d'annotation d'un document numérisé

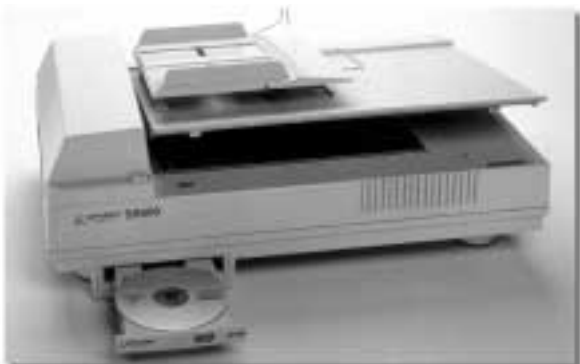
## INTRODUCTION A LA GEIDE



Scanner ou numériseur de documents Canon



Scanner ou numériseur de documents Fujitsu



Scanner ou numériseur de documents Mitsubishi



Scanner pour microfilm 16 mm Kodak

très rapidement des originaux. Certains modèles numérisent une seule face à la fois, d'autres le recto et le verso de chaque original en un temps variant de 0,8 à 2 secondes par page. Certains sont conçus pour saisir des dizaines de milliers de documents par jour et sont équipés pour cela d'une alimentation automatique de très grande capacité, de dispositifs de détection des doubles et même de dispositifs de prétraitement et de compression des images. Pour les microformes, (microfilms, carte à fenêtre, microfiches), il existe une catégorie spéciale de numériseurs qui permettent de convertir des milliers de micro-images par jour lorsqu'ils sont équipés de dispositifs d'avance et de positionnement automatique des films. Pour les documents de grande taille tels que des plans ou des dessins industriels, plusieurs industriels ont conçu des numériseurs spécifiques qui convertissent tout ou partie d'un original en image numérisée noir et blanc ou couleur. Celle-ci pourra être utilisée comme archive ou pour assurer la vectorisation des dessins et des schémas pour qu'ils puissent être retravaillés avec des logiciels de CAO/DAO. Parmi la multitude de numériseurs sur le marché, on trouve des appareils spécialisés dans la numérisation de livres ou d'ouvrages reliés, qui compensent l'épaisseur du livre par un plateau ou par un redressement électronique de la courbure de l'image.

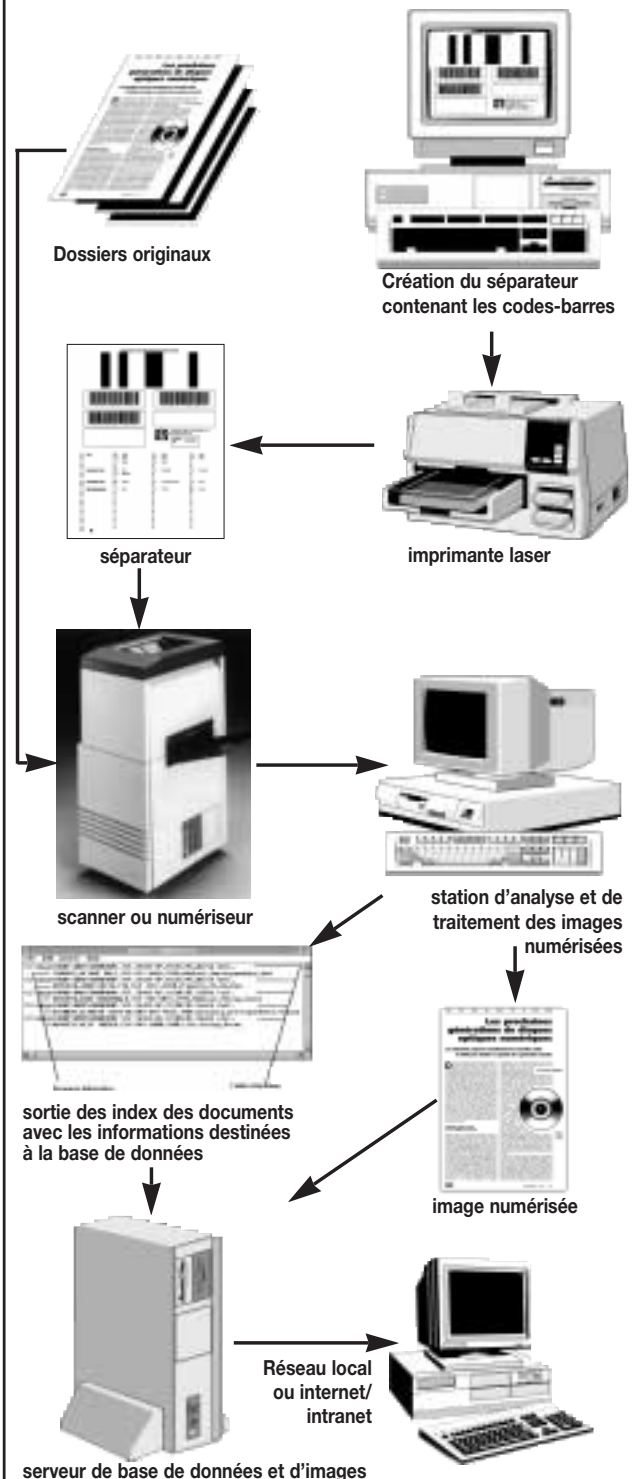
### L'indexation des documents

L'indexation consiste soit à créer une fiche descriptive d'un document ou d'un dossier, soit à analyser son contenu textuel pour indexer chaque mot. Dans le premier cas, les informations de la fiche ou des descripteurs seront les repères qui aideront à classer puis à sélectionner des documents lors de recherches ultérieures ou à les diffuser de façon sélective à des utilisateurs connectés au serveur. Dans les systèmes de GED, on fait intervenir plusieurs méthodes pour indexer des documents numérisés. Certaines partant de la description du contenu de l'original font appel à des logiciels de gestion documentaire. D'autres passent par des gestionnaires de bases de données (SGBDR) pour créer des liens entre un document ou un dossier numérisé. Dans certains cas, cette indexation sera réalisée manuellement par un opérateur qui remplira les différents champs de la fiche descriptive.

Dans d'autres, elle sera réalisée électroniquement par extraction de données déjà inscrites sur l'original. Le cas le plus fréquent est celui des codes à barres préalablement imprimés ou collés sur un original qui, après détection puis

## INTRODUCTION A LA GEIDE

### Principes de l'indexation automatique à l'aide de séparateurs contenant des codes-barres et des zones



interprétation par l'ordinateur, permettent de récupérer automatiquement des informations qui seront exportées vers la base de données. Il est également possible d'effectuer l'extraction sélective d'informations sur des zones prédéfinies d'un formulaire en combinant la segmentation d'images et la reconnaissance optique de caractères (OCR/ICR), extraction qui générera un fichier ASCII délimité exploitable par des gestionnaires de bases de données. D'autres techniques d'indexation automatique existent. Par exemple, les séparateurs de dossiers ou de documents préalablement placés entre les originaux. Il s'agit de feuilles sur lesquelles sont imprimés des codes-barres ou sur lesquelles sont cochées des cases exploitables par un programme qui assure la conversion des informations en fichiers ASCII importables dans les bases de données. Des technologies de reconnaissance qui tirent parti de la puissance de calcul des ordinateurs autorisent l'indexation automatique par interprétation du type de document, de leur contenu, etc. Ces techniques novatrices en sont à leur début opérationnel mais laissent entrevoir des perspectives intéressantes. En attendant, de nombreux fournisseurs proposent l'indexation en texte intégral pour automatiser le référencement des originaux. Dans ce cas, le contenu des documents numérisés est intégralement lu par des logiciels de reconnaissance optique de caractères (OCR/ICR) qui créent un fichier ASCII. Un programme "texte intégral" ("full text" en anglais) indexe ce fichier, complète une base d'index et crée un lien entre le contenu extrait et l'image numérisée de l'original. Les recherches ultérieures s'effectuent par interrogation de la base à l'aide d'un mot ou d'une combinaison de mots en utilisant, si cela est nécessaire, des opérateurs (et, ou, sauf, etc.). Ces mêmes méthodes d'indexation en texte intégral sont applicables aux fichiers bureautiques (traitement de texte, page HTML, etc.) ainsi qu'aux messages électroniques. On commence également à voir apparaître des logiciels capables de réaliser un référencement primaire et automatique de documents numérisés en couleur, tels que des photographies, en utilisant les dominantes de couleur, de forme et certaines autres composantes. De plus en plus, les logiciels de GED commercialisés sur le marché intègrent des méthodes automatiques d'indexation. Certains n'intègrent que la lecture de codes à barres, d'autres combinent plusieurs méthodes. Cette fonc-

## INTRODUCTION A LA GEIDE

tion permet de résoudre les problèmes que pose l'indexation des documents qui constitue souvent un goulot d'étranglement dans le système GED de l'entreprise.

### Les différents types d'image

Les systèmes GED gèrent des images numérisées dont la nature dépend des originaux et des traitements prévus par la suite. Elles peuvent être bitonales, à niveau de gris ou en couleurs. Avec les images dites bitonales, les informations d'un original sont uniquement converties en noir et blanc. Elles se rencontrent principalement là où les documents ne comportent que du texte. Les images dites à niveaux de gris restituent les nuances de gris, du gris foncé au gris pale qui donnent du relief à une photo ou un dessin. Ces nuances peuvent être plus ou moins nombreuses, de 16 à 256 ou plus, avec pour conséquence une augmentation de taille du fichier électronique. Certains numériseurs et certains programmes sont conçus pour pratiquer la numérisation bitonale et la numérisation à niveaux de gris sur un même document. Ainsi, les textes seront traités en bitonal tandis que les photographies ou les illustrations seront converties en utilisant des niveaux de gris. Quant aux images en couleurs, il est possible de les numériser en codant les couleurs avec plus ou moins de précision. En général, les numériseurs adaptés permettent de numériser sur 8 bits, 16 bits, 24, 30 ou 36 bits. Un codage des couleurs sur 8 bits permettra de restituer une image électronique en 256 couleurs différentes tandis qu'un codage sur 16 bits donnera une image de 64.000 couleurs et un codage sur 24 bits, une image de 16,7 millions de couleurs; avec toujours comme conséquence une augmentation de la taille du fichier. Un autre facteur entrant en ligne de compte dans le processus de numérisation est la résolution d'analyse. Celle-ci est exprimée en DPI (Dots Per Inch) ou points par pouce en français. La majeure partie des numériseurs GED offrent un spectre de réglages allant de 100 à 400 DPI. D'autres, notamment les appareils pour images couleurs ou pour microformes, sont capables de numériser en milliers de DPI. En général, les solutions GED utilisent des résolutions comprises entre 200 et 400 DPI (8 ou 16 points/mm), ce qui donne des résultats tout à fait acceptables.



*Exemple de traitement des images numérisées : suppression de fond et augmentation du contraste*

### La compression des images

Les images numérisées subissent plusieurs traitements électroniques avant d'être stockées. Ceux-ci sont fonction de l'utilisation qui sera faite des documents ainsi convertis. Mais il faut tout d'abord et impérativement les compresser pour réduire la taille du fichier. Cette compression réduit la place occupée sur les supports d'archivage mais permet aussi de transférer plus rapidement un fichier sur des réseaux locaux ou de type intranet/internet. Les techniques et méthodes de compression sont diverses. Elles diffèrent en fonction de la nature des originaux et de la méthode de numérisation. Pour les documents textuels numérisés sous forme d'images bitonales, c'est la compression CCITT G3/G4 qui est utilisée dans la plupart des logiciels de GED. Basée sur des techniques conçues à l'origine pour la télécopie, elle fait appel à des algorithmes normalisés et n'induit pas de pertes lors du traitement, ce qui préserve l'intégrité du document original. De création plus récente, la compression JBIG (Joint Bi-level Image Group) a fait son apparition dans les applications GED et la télécopie. Le JBIG, normalisé au niveau de l'ISO réduit la taille des documents, là aussi sans perte du contenu. Le traitement assure une compression plus élevée que le CCITT G4 et offre quelques avantages tels que la création automatique de variantes d'une même image ou d'une imagerie. Les documents couleurs ou à niveaux de gris numérisés sont généralement compressés en utilisant l'algorithme JPEG (Joint Picture Expert Group) normali-



## INTRODUCTION A LA GEIDE



*Carte de traitement d'images numérisées*

sé par l'ISO. Avec JPEG, l'utilisateur dispose d'une grande latitude dans le choix du taux de compression. Ce taux dépend du compromis à trouver entre le résultat attendu et la taille du fichier. JPEG induit des pertes dans le contenu des images lorsqu'on applique de très forts taux de compression. Par expérience, ces pertes sont perceptibles au delà d'un ratio de 1/25 et varient en fonction de la nature de l'image. Il est également possible d'utiliser JPEG pour compresser une image couleur, sans dégradation mais en acceptant un taux de compression plus faible. Le comité ISO chargé de la normalisation de JPEG et de JBIG doit faire face à une évolution très rapide de ces techniques. Les experts du JPEG 2000 espèrent néanmoins donner naissance à une

solution multicodec capable de s'adapter à un large éventail de besoins.

Nous avons cité là les méthodes normalisées de compression des images fixes mais il en existe de nombreuses autres faisant appel à des algorithmes propres à une société et dont le résultat est parfois surprenant (fractales, ondelettes, etc.). Toutefois, dans un souci de pérennité et de transfert éventuel des documents vers d'autres programmes, la majorité des fournisseurs GED ont choisi d'utiliser des techniques normalisées et universellement employées. Cette largeur d'esprit n'est pas partagée par tous. C'est ainsi que l'on trouve souvent dans les logiciels de GED des entêtes de fichiers propriétaires qui ne permettent pas de porter les images vers un logiciel concurrent. Il convient donc de vérifier par des tests les possibilités d'exporter ou de convertir les images si l'on veut être sûr de pouvoir relire les documents numérisés au cas où l'on changerait de fournisseur de logiciel de GED. Les images numérisées peuvent être compressées soit par logiciel, soit à l'aide d'une carte accélératrice intégrant des processeurs de traitement conçus à cet effet. Aujourd'hui la puissance de calcul des micro-ordinateurs ou des stations de travail rend possible la compression des images à l'aide d'un logiciel intégré aux programmes GED. Cette méthode est bien adaptée à des traitements quotidiens peu nombreux. Lorsqu'on utilise des numériseurs rapides ou recto/verso, il faut faire appel à une ou plusieurs cartes additionnelles. En général, ces cartes ont plusieurs fonctions : interfacer le scanner au micro-ordinateur, compresser grâce aux processeurs de traitement intégrés qui sont capables de compresser en temps réel des images immédiatement après leur saisie et, enfin, réaliser des prétraitements d'amélioration du rendu des images (suppression de fond, etc.). Ces cartes existent en différentes versions et assurent la compression aux normes CCITT G3/G4 ou en JBIG. Les images couleurs nécessitent en général une carte dotée d'un processeur de traitement JPEG ou un simple logiciel. Dans les prochaines années, toutes ces fonctions de compression seront réunies sur une même carte en utilisant un processeur universel capable d'utiliser les différents algorithmes de compression.

### **Les autres traitements optionnels**

Les images peuvent être appelées à subir d'autres traitements avant la compression. Tous ceux-ci sont programmables par logiciel et peuvent se faire sans intervention humaine lorsqu'ils ont été préalablement paramétrés. Par exemple, l'amélioration du rendu des images numérisées se fait par application de plusieurs filtres. Ils ont pour effet d'augmenter le contraste entre les noirs et les blancs ou d'éliminer un fond gris résultant de la numérisation d'un document dont l'original est imprimé sur un papier de couleur. De même, des filtres permettent de réduire le nombre de défauts, des points noirs ou des rayures, présents sur un original de mauvaise qualité. Lorsque l'application prévoit la reconnaissance optique des caractères, on peut intégrer de nombreux logiciels complémentaires pour faciliter cette reconnaissance, par exemple un module éliminant les cadres ou les cases utilisées dans un formulaire pour délimiter les zones à remplir. Ces informations de forme, utiles lors de l'établissement du document original, sont inutiles, voire nuisibles, au traitement ultérieur et sont automatiquement éliminées par des logiciels

## INTRODUCTION A LA GEIDE

spéciaux. Tous ces filtres permettent au final une meilleure reconnaissance optique de caractères, qualité qui elle-même conditionne l'acquisition de données soit d'indexation, soit d'enrichissement des applicatifs ou de bases de données. Parmi les développements prometteurs, des ingénieurs ont mis au point un logiciel qui, couplé avec l'électronique du numériseur ou de sa carte d'interfaçage, permettent d'optimiser automatiquement le résultat de la numérisation de documents saisis par lots. Ils procèdent en utilisant des fichiers de référence préalablement réalisés par l'opérateur afin d'adapter automatiquement les traitements à la nature des originaux.

### Les différentes solutions GED

L'offre en solutions GED est particulièrement bien fournie et répond à la plupart des besoins. Il existe des logiciels de GED pour la majeure partie des plates-formes informatiques et des systèmes d'exploitation du marché. Actuellement, les plus répandus utilisent l'architecture client/serveur dans laquelle les serveurs fonctionnent sous Unix, Windows-NT ou NetWare (Novell) et les postes clients sous Windows (3.X, 95 ou NT), MacOS, ou les différents Unix, etc. D'autres solutions fonctionnent sur des mini-ordinateurs, notamment AS/400, ou en liaison directe avec des ordinateurs centraux. Pour schématiser, cette offre se scinde en trois catégories: les logiciels d'entrée de gamme ou packagés, les programmes génériques utilisés en complément des applicatifs informatiques et les modules d'intégration entrant dans la catégorie GED par composants.

Les logiciels GED d'entrée de gamme ne représentent actuellement qu'une très infime partie du marché, même aux USA. Ces programmes recèlent néanmoins un potentiel commercial important du fait de leur faible prix et du large public qu'ils sont susceptibles de toucher. En général, il s'agit de solutions mono-poste destinées à des entreprises artisanales ou individuelles ou même à des applications personnelles. La plupart de ces programmes ne fonctionnent qu'avec des numériseurs de bureau et ne permettent de saisir qu'un nombre limité de documents par jour. Ils contiennent une base de données où sont classées les données d'indexation des images numérisées et autorisent l'importation de télécopies ou de fichiers déjà convertis. En général, ils stockent ou archivent les documents sur le disque dur de l'ordinateur mais certains acceptent un support amovible tel qu'un disque optique ou une cartouche magnétique. Depuis un an, il existe aussi des solutions combinant le préformatage des informations à la norme ISO-9660 et l'enregistrement sur un CD-R et CD-RW en utilisant le format logique UDF. On pourra demain les coupler à des enregistreurs de disques optiques enregistrables tels que les DVD-R. Aux USA, ces logiciels de GED personnelle ont fait leur apparition sur le marché en 1995 et sont vendus au prix des logiciels de bureautique, traitements de texte ou tableurs. Certains programmes possèdent des fonctions plus avancées comme la reconnaissance optique de caractères et une indexation en texte intégral. C'est un marché en émergence.

### La GED générique ou spécifique

La deuxième catégorie, celle des logiciels de GED dits génériques, se compose de solutions s'intégrant en complément des applications bureautiques ou transactionnelles des entreprises et des administrations. Ils sont disponibles pour de nombreux systèmes d'exploitation et s'intègrent généralement dans une architecture client/serveur en connexion ou non avec l'informatique centrale de l'entreprise. Ils peuvent être personnalisés et adaptés aux exigences d'applications particulières et ils peuvent être interfacés à des gestionnaires de bases de données. La raison



Exemple de traitement par OCR de document numérisé

## INTRODUCTION A LA GEIDE

d'être de ces logiciels est d'ajouter des fonctions d'acquisition, de traitement, de recherche, de consultation et d'archivage de documents numérisés ou de fichiers numériques dans l'environnement bureautique tout en jouant le rôle de fédérateur au sein d'un système de gestion d'informations. Cette catégorie constitue la majeure partie de l'offre actuelle et est étoffée par des applicatifs GED répondant à des besoins spécifiques pour la gestion de dossiers d'assurances, de crédit, de l'état civil, etc.

### Le GED modulaire ou "à composants"

Outre les programmes de gestion électronique de documents et d'informations packagés ou personnalisables, on trouve sur le marché la GED dite modulaire ou par composants. Cette catégorie recouvre des modules logiciels intégrables à des applications au travers d'outils de développement comme Visual Basic, Visual C++ ou PowerBuilder. Ils sont en général vendus regroupés dans une "boîte à outils" aussi appelée "toolkit" et s'adressent aux programmeurs ou aux sociétés de services informatiques désireux d'ajouter des fonctions de gestion électronique de documents à leurs logiciels. Disponible depuis quelques années, la GED par composants offre des possibilités très larges d'intégration. On y trouve des modules de numérisation et de traitement de documents, d'extraction d'informations par reconnaissance optique de caractères ou de marques, d'indexation automatique par lecture de codes à barres ou autres, d'exportation d'index vers des bases de données, de gestion de stockage des images numérisées. Elle s'adapte aux volumes à traiter, de quelques centaines à plusieurs dizaines de milliers d'originaux; il suffit d'en équiper des stations d'acquisition réparties, connectées à un serveur au travers d'un réseau local. L'acquisition des documents est généralement réalisée par lots, de manière continue, avec répartition des tâches entre plusieurs micro-ordinateurs. Certaines solutions de GED modulaire sont ainsi capables de capturer, de traiter et d'indexer quotidiennement des dizaines de milliers de documents sans intervention humaine, si ce n'est l'alimentation des numériseurs et le contrôle du bon fonctionnement des périphériques et des ordinateurs. Cette catégorie de GED est disponible pour les environnements informatiques les plus utilisés. Elle se conforme aux normes de l'industrie et tourne sur des plates-formes banalisées. Outre ces modules intégrables, plusieurs éditeurs ont conçu des programmes de saisie et d'indexation packagés et paramétrables pouvant être utilisés pour réaliser toutes les opérations de numérisation, d'indexation et d'extraction d'informations dans un environnement client/serveur. A ceux-ci s'ajoutent désormais des modules d'archivage et de gestion de périphériques de stockage tels que des juke-boxes.

### Le workflow pour la circulation de l'information

Sous le terme workflow sont regroupés les logiciels de circulation sous forme électronique de documents ou d'informations et d'attribution de tâches administratives. Ces programmes vont plus loin que la simple distribution; ils permettent de créer des procédures complexes de suivi de dossier et de contrôle des traitements réalisés. Les logiciels de workflow forment un excellent tandem avec les programmes de GED dans toutes les applications de trai-

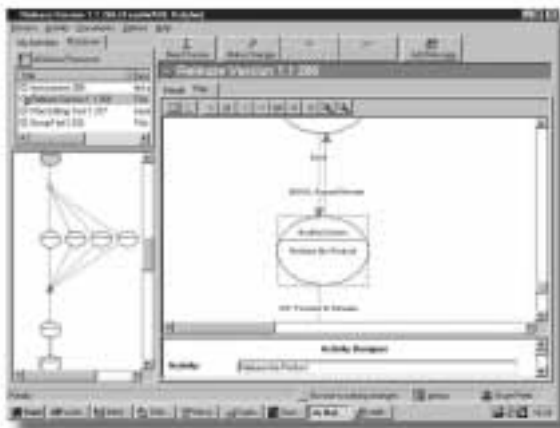


*Copie d'écran d'une application de GED générique avec recherche en texte intégral Zylab.*



*Maxim de Creativ'Systems pour l'ajout de fonctions GED/GEIDE à des applicatifs*

## INTRODUCTION A LA GEIDE

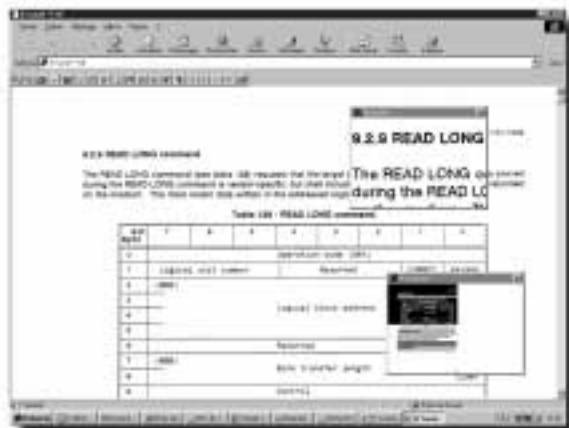


Copie d'écran d'une application workflow

tement de dossiers où interviennent plusieurs personnes. Grâce aux nouvelles techniques de programmation, la plupart sont paramétrables par l'administrateur du système, par le biais d'une interface graphique de circulation des documents et d'attribution des tâches. Ils sont dotés de protocoles de vérification puissants qui permettent de vérifier le circuit effectué par un dossier. Les logiciels de workflow fonctionnent également de concert avec les programmes bureautiques et s'appuient en général sur les messageries électroniques les plus répandues. Depuis deux ans, il existe également des logiciels de workflow dont la partie "client" permet d'accéder à un serveur au travers de réseaux de type intranet/internet; ce qui ouvre des perspectives intéressantes aux entreprises qui ont des filiales ou des équipes décentralisées.

### L'ouverture de la GED et du workflow sur internet et intranet

Comme bon nombre d'applications informatiques, la GED et le workflow ne restent pas à l'écart des bouleversements provoqués par la montée en puissance des réseaux internet ou intranet (voire extranet). La majorité des éditeurs de programmes GED ou de workflow ont conçu des passerelles entre leurs logiciels et la consultation via ces réseaux. La première étape consiste à donner à l'utilisateur la possibilité d'interroger à partir d'un micro-ordinateur muni d'une connexion internet/intranet les bases de données ou de références des documents ou des fichiers gérés par le programme GED, soit COLD, soit "image". Quand la confidentialité l'exige, certains répertoires ou groupes de documents sont protégés par des mots de passe. Le résultat des recherches est converti au format HTML et est affiché sur le "browser" ou feuilleté banalisé du poste de travail. L'utilisateur peut ainsi les parcourir. Les images elles-mêmes de documents numérisés peuvent être récupérées et affichées, soit après conversion du mode CCITT G3/G4 au format GIF par le serveur, soit directement sur l'ordinateur au travers d'un "plug-in" adapté. Il existe actuellement sur le marché plusieurs logiciels de ce type qui fonctionnent en association avec les deux principaux programmes de consultation sur internet. Ces plug-ins intègrent un module de décompression des images qui est aussi un module de manipulation puisqu'il possède des fonctions de type agrandissement sélectif, amélioration du rendu par ajout de niveaux de gris ("scale-to-gray" en anglais), impression, etc.



Visualisation d'images GED avec Navigator de Netscape en utilisant un «plug-in» adapté.

De même, les logiciels de workflow disposent d'extensions qui permettent d'étendre la circulation des documents ou d'informations ainsi que l'attribution de tâches à des postes de travail distants connectés au travers d'un réseau internet ou intranet. Certains développeurs de logiciels GED ou de workflow envisagent d'aller plus loin dans cette démarche. Ils prévoient de proposer dans les années à venir de délocaliser certaines tâches ou de déporter des fonctions spécifiques, du moins lorsque les nouveaux outils et réseaux de télécommunications auront apporté une solution aux problèmes de débit. En 1998, la technologie Java a fait son apparition dans les parties "serveur" des solutions GED et workflow pour intranet/internet.

**INTRODUCTION A LA GEIDE****Quels supports de stockage pour la GED ?**

Il n'y a pas de supports de stockage d'informations répondant spécifiquement aux besoins des solutions GED; celles-ci acceptent plusieurs catégories de médias, le choix étant d'abord fonction des besoins de l'utilisateur. Par support de stockage d'informations d'une solution GED, il faut entendre à la fois le ou les médias sur lesquels vont être enregistrés et archivés les images numérisées, les fichiers spools ou informatiques (de type bureautique, son, vidéo, etc.) gérés par le système mais également les supports où sont stockés les index du gestionnaire de base des données ou le support du progiciel documentaire utilisé.

Ces supports sont dépositaires d'une information qui devra être conservée plus ou moins longtemps mais parfois plusieurs dizaines d'années. Certaines informations ont un statut spécifique et leur authenticité ne doit pas remise en doute. Il est donc des circonstances où les supports doivent répondre au double impératif d'une conservation dans le temps et d'une conservation sans modification possible. Les médias les plus aptes à répondre à ces exigences sont certainement les disques optiques numériques mais ils ne sont pas les seuls. Dans certains cas, un sous-système RAID à base de disques durs magnétiques est aussi une solution financièrement viable et assez sûre. Dans d'autres cas, lorsque la consultation des documents est peu fréquente, ceux-ci pourront être archivés sur des bandes magnétiques, voire enregistrés sur des micro-films 16 ou 35 mm.

**Francis Pelletier**

© Copyright 1998 - Mos Magazine - ARCA éditions - Tous droits réservés.

**Pour en savoir plus  
sur les différents supports de stockage et d'archivage :**

**Les différents disques optiques numériques**

**Pour en savoir plus et être informé  
sur les technologies, les produits  
et les applications**

**ABONNEZ-VOUS à**

**BP 303 - 56008 VANNES CEDEX - FRANCE****Tél : (33) 02 97 47 83 06 - Fax : (33) 02 97 47 49 46 - E-mail: info@mosarca.com**