

## LA BIBLIOTHÈQUE NUMÉRIQUE ET L'ENTREPRISE

Jacques CHAUMIER

Conseiller auprès du Président

du Bureau van Dijk

Ingénieurs Conseils Paris

En 1965, un chercheur du M.I.T, J.C LICKLIDER, écrivait dans son ouvrage *Libraries of the future* (Licklider, J.C, 1965) :

« Actuellement, pour qu'il y ait transfert de la mémoire ou du stock au lecteur ou usager, il faut déplacer le livre ou le lecteur ou les deux. L'un des processus futurs sera de supprimer cela, l'information seule se déplaçant... L'homme moyen aura une console, sa Cadillac intellectuelle comme il achète maintenant une automobile. » Cette prévision technologique est devenue une réalité avec la bibliothèque numérique ou bibliothèque virtuelle. La bibliothèque virtuelle est, comme l'a écrit Corrado PETTENATI, un « concept organisationnel basé sur l'alliance de l'informatisation de la bibliothèque, de l'accès aux télécommunications et d'une nouvelle gamme d'outils directement utilisables par le lecteur, le dispensant de recourir à un intermédiaire humain. » Nous avons là tous les ingrédients de la bibliothèque virtuelle mais il faut y ajouter la technologie la plus importante de toute la panoplie des technologies de l'information et de la communication, les TIC (Technologies de l'information et de la communication), la numérisation. À l'ère de la bibliothèque virtuelle, le document papier se fait de plus en plus rare et les auteurs préfèrent soumettre leurs travaux sous forme électronique. L'information, le document, sont devenus numériques ou digitaux. C'est la numérisation et donc la dématérialisation du document qui a permis la constitution de bibliothèques virtuelles ou, pourrait-on dire aussi en référence au musée imaginaire d'André Malraux, de bibliothèques imaginaires. C'est parce que le document est numérique que l'on peut en stocker un grand nombre dans les mémoires magnétiques ou numériques, que l'on peut le transporter sur les réseaux de télécommunications. L'information électronique est devenue une matière première stratégique, dont le coût est certes parfois élevé, mais dont la valorisation est essentielle pour l'entreprise. Ainsi, selon une étude du GFII (Groupement français de l'industrie de l'information), le marché de l'information électronique professionnelle qui était de 1,12 milliards d'euros en 2001, devrait atteindre 2,36 milliards d'euros en 2005. L'information financière, l'information de solvabilité et l'information marketing sont actuellement les trois principaux segments de ce marché puisque représentant près de 70%. Une autre étude du cabinet britannique IRN Research, *The European Online Information Market 2003*, évalue le marché européen (Europe des 15) de l'information professionnelle électronique, à 3,45 milliards d'euros en 2002, hors information financière temps réel et hors dépêches de presse, ce marché affichant une croissance de 12%.

### Fonctions de la bibliothèque numérique

Une bibliothèque, traditionnellement, était organisée autour de quatre principes :

- Une organisation en fonction d'un public déterminé, on pourrait même dire en fonction d'une « clientèle »,
- Une sélection de documents en fonction d'une identification des besoins et d'un ajustement permanent à ceux-ci,
- Une organisation des connaissances : préparation du catalogue, indexation des documents (en fonction de la clientèle et de ses besoins), etc....
- Une offre de service : aide à la recherche, accès aux documents, etc....

La bibliothèque numérique reste basée sur ces principes mais va élargir ses fonctions et offrir de nouveaux services à un lecteur qui ne se déplace plus et qui trouvera un document toujours disponible du fait de sa dématérialisation. L'un des principes essentiels de la bibliothèque numérique dans l'entreprise est celui de *the right information for the right man*. Mais on ajoutera aussi *in the right time*. Ce dernier point est sans doute l'un des plus importants dans le fonctionnement de la bibliothèque numérique. L'information doit être délivrée au bon moment et doit pouvoir atteindre son destinataire quel que soit le lieu où il se trouve. Le fait qu'à tout moment et en tout lieu et ceci sans frontière, le cadre puisse se connecter à la bibliothèque numérique de son entreprise ou de son organisation est un élément déterminant dans la valorisation de l'information. La disponibilité immédiate du document numérique dès sa création est aussi un facteur d'efficacité très important. Il y a instantanéité entre création et consultation. Comme on l'a dit précédemment les fonctions de la bibliothèque numérique se sont considérablement enrichies. La première fonction reste celle de l'identification des documents. C'est la traditionnelle fonction de consultation de catalogue. Mais sur la bibliothèque virtuelle, l'offre en matière de consultation de catalogues pourra être très riche. Cette richesse devra cependant être raisonnée. Il ne sert à rien de pouvoir accéder à n'importe quel catalogue dont on ne connaît pas l'intérêt et la valeur. La mise à disposition de l'utilisateur, de bookmarks ou signets, c'est-à-dire une sélection de sites ou catalogues à consulter, sites choisis et évalués en fonction de leur intérêt et des besoins des utilisateurs valorise l'offre disponible. Cette fonction de consultation de catalogues, de recherche documentaire, est largement facilitée par les logiciels d'interrogation, à partir d'outils banalisés tels que les navigateurs classiques d'Internet. Cette interrogation pourra être faite, en fonction des fichiers et bases de données accessibles sur la bibliothèque virtuelle, soit à partir de mots-clés, soit à partir du texte intégral des documents. Une nouvelle catégorie d'outils informatiques est apparue avec les infologiciels, outils de recherche et d'analyse de l'information textuelle. Selon une étude d'IDC, le marché des logiciels de recherche et de classification de l'information devrait atteindre 2,630 millions de dollars en 2005. La seconde fonction est, bien entendu, la visualisation du document sur l'écran de la station de travail ou du micro-ordinateur portable. Le document peut-être affiché, selon les cas, en mode page restituant le document en fac-similé, véritable photographie de l'original, ou en mode texte afin de permettre un accès dans le texte. Certaines applications prévoient d'ailleurs le double affichage. Le système peut également proposer un système d'alerte dans ce que l'on appelle le mode push. Ce n'est plus le lecteur qui va chercher (mode pull) les documents de sa propre initiative, mais le système qui lui propose, sur base d'une définition d'un besoin permanent (le profil) les documents répondant à ce besoin et susceptible de l'intéresser. Le signalement des documents peut être envoyé directement dans la boîte-aux-lettres électronique de l'utilisateur. L'appropriation du document par l'utilisateur est également un élément fondamental dans le concept de bibliothèque numérique. En effet, le document n'étant plus unique mais, du fait de la dématérialisation, toujours disponible, l'utilisateur peut profiter du document à sa guise. Il peut avoir une copie électronique du document par télédownload et obtenir une copie papier sur son imprimante. D'ailleurs, maintenant, un certain nombre d'éditeurs n'imprime plus les ouvrages qu'ils éditent mais propose

une impression à la demande à partir d'une copie numérique du document. L'utilisateur de la bibliothèque numérique va, en outre, avoir à sa disposition un certain nombre de possibilités de manipulation du document. Parmi ces diverses possibilités, la fonction d'annotation est très intéressante pour l'utilisateur. Celui-ci va pouvoir annoter le texte, le marquer (vieux rêve d'enfant que de pouvoir écrire sur les livres), ajouter ses commentaires. Et ce qui ne pouvait être fait avec le document papier, le document électronique peut, lui, être restructuré selon les pratiques de l'utilisateur. C'est ainsi que chez un grand fabricant de pneumatiques, les fascicules de brevets numérisés ont été restructurés selon les habitudes de consultation des ingénieurs en propriété industrielle. Les brevets, au lieu de la mise en page habituelle avec d'abord le texte du brevet en premier, présentent immédiatement le point essentiel d'un brevet aux yeux des spécialistes que sont ce que l'on appelle les revendications, c'est-à-dire l'objet du brevet. La création de dossiers électroniques offre également de nombreux avantages à l'utilisateur. Les dossiers pourront être classés par l'utilisateur selon son propre mode de classement, un document pouvant en outre être affecté à plusieurs dossiers sans que celui-ci ne soit physiquement dupliqué. L'épuration des dossiers pourra également être automatique si l'utilisateur prévoit des dates de péremption sur les documents. Le document lui-même pourra être modifié ou mis à jour, les diverses versions de travail pouvant être archivées dans les dossiers personnels avant que le document définitif ne soit versé dans ce qui constitue la mémoire de l'entreprise. L'utilisation de procédures de workflow dans le cadre de la bibliothèque numérique enrichira également l'utilisation de celle-ci avec la mémorisation des différentes étapes de la création d'un document et les diverses versions de celui-ci. Mais si les manipulations sur le contenant, c'est-à-dire la structure du document, sont intéressantes, l'exploitation du contenu de ceux-ci, dans le cadre des bibliothèques numériques, offre encore plus d'intérêt. Parmi les possibilités d'exploitation et de valorisation, l'analyse bibliométrique est une des voies les plus intéressantes qu'apporte la constitution de fonds numérisés. Sur un corpus, ensemble de documents défini par l'utilisateur, celui-ci, à l'aide d'un logiciel spécifique, va pouvoir faire une analyse sur certains critères permettant de mettre en évidence certaines tendances ou de faire apparaître des phénomènes aberrants. Ce type d'analyse statistique est régulièrement pratiqué dans un certain nombre d'entreprises dans les domaines chimique ou pharmaceutique, en particulier sur les brevets. Une surveillance d'activité de la concurrence dans certains domaines ou dans certains pays peut aussi être exercée. Cette mesure de l'activité documentaire est également importante sur les communications dans les congrès, celles-ci étant un des indicateurs de l'activité dans un centre de recherche ou dans un laboratoire. L'analyse de contenu peut également être une analyse lexico-sémantique permettant, à partir de l'apparition de termes dans un corpus, de détecter des tendances technologiques ou économiques à travers les signes avant-coureurs de ces tendances. Ces tendances peuvent être visualisées à l'aide de logiciels de cartographie des connaissances. Ces cartes de la connaissance permettent de suivre l'évolution d'une technologie dans le temps par comparaisons successives. D'autres outils sémantiques peuvent également être mis, avec la bibliothèque numérique, au service des utilisateurs. C'est le cas, par exemple, des bases de données terminologiques ou des bases de données de suivi des noms de sociétés. La base de données terminologiques, réalisée chez un manufacturier français, permet à l'ensemble de la communauté d'utilisateurs, sur tous les sites dispersés dans le monde, de parler, si ce n'est la même langue, au moins le même langage technique. La base de données du suivi des noms de société, dans une grande banque belge, va permettre aux divers services de voir que la Bayerische Hypotheken Bank et la Vereinsbank ont fusionné pour devenir la Bayerische Hypo Vereinsbank, et qu'il ne faut pas confondre son sigle BHV avec celui du grand magasin, le Bazar de l'Hôtel de Ville. Enfin, dernières fonctions dans une bibliothèque numérique, l'établissement des liens. Cette fonction hypertexte donne la possibilité à

l'utilisateur de créer des liens à l'intérieur d'un document pouvant créer ainsi des cheminements de lecture. Il peut aussi créer des liens, dans un corpus, entre différents documents. C'est à partir de fonctions de ce type qu'un grand serveur a construit sa bibliothèque de clauses pour la préparation des contrats avec ses clients. Le concept de bibliothèque numérique, en particulier lorsque son accès est possible sur l'intranet de l'entreprises, rencontre une forte adhésion des utilisateurs et décuple le recours à l'information. Ainsi les AGF (Salleras, B, 2000) avaient deux objectifs dans la mise en place de leur bibliothèque numérique sur l'intranet : offrir aux utilisateurs un accès immédiat au texte intégral des documents, et non plus simplement à des mots clés ou des résumés, et valoriser l'information.

La bibliothèque numérique est organisée autour de trois bases de données : une base "généraliste" de 20.000 documents, articles, ouvrages et études sur les domaines intéressants les AGF ; une base "concurrence" comportant 10.000 articles et dépêches de presse, et enfin, l'ensemble des notes et circulaires, en texte intégral, éditées par la Fédération française des sociétés d'assurances, constituant les documents de références de la profession. Le passage d'un fonds documentaire de références à une bibliothèque de documents numérisés a fait passer le nombre de connexions mensuelles sur le site de 250 à 1.300.

### **Les difficultés de la bibliothèque numérique**

La création et la mise à disposition d'une bibliothèque numérique au sein de l'entreprise ne sont pas exemptes de difficultés, difficultés dues, pour beaucoup, à la jeunesse des technologies mises en œuvre. La normalisation est un point important car la bibliothèque numérique doit être un outil de communication et, qui dit communication dit normalisation. Mais c'est sans doute le point sur lequel les difficultés sont résolues ou en voie de résolution. La normalisation est maintenant acquise dans le codage des caractères avec UNICODE qui se substitue peu à peu au codage ASCII. UNICODE permet l'utilisation d'un grand nombre de jeux de caractères et présente ainsi un caractère international avec l'utilisation des langues non latines. Le balisage et la structuration des documents numériques sont désormais eux aussi normalisés avec l'adoption de la norme XML qui se substitue, elle aussi, à la norme SGML. Une large part de la documentation technique est désormais créée sous XML. Les métadonnées ont également leur norme avec le Dublin Core. Le terme de meta-données désigne de manière générique, les données créées pour fournir des informations sur les ressources électroniques d'information (Lupovici, C, 2000). Le Dublin Core définit les méta-données (titre, auteur, éditeur, date, etc...) caractérisant un document. L'utilisation conjointe d'XML et des méta-données normalisées devrait permettre un catalogage automatique des documents dans les bibliothèques numériques.

Enfin, PDF (Portable document format) est une norme largement reconnue et utilisée pour la diffusion des documents électroniques, de même que l'adoption de la norme Groupe IV du fax pour la compression des documents en noir et blanc et des normes JPEG (Joint Photographic Export Group – ISO 109181) et MPEG (Motion Picture Expert Group – ISO 11172) pour la compression des images fixes et des images animées. Les problèmes de l'identification des documents à travers les bibliothèques numériques commencent à trouver un début de solution avec la création d'un identifiant unique pour les documents numériques, le DOI (Digital Object Identifier) proposé par l'Association of American Publishers. Cet identifiant comporte deux parties : un préfixe identifiant l'éditeur, un suffixe identifiant l'objet chez l'éditeur. Ces identifiants sont stockés dans un répertoire central avec leur traduction dans les URL des systèmes de détenteurs de droit. Dans le domaine de

la normalisation du document numérique, des travaux sont actuellement menés par un consortium réunissant à la fois des professionnels du livre (Book Industry Study Group), des éditeurs d'ouvrages (Harper Collins), des éditeurs de logiciels (Adobe Systems, Microsoft), des sociétés développant des livres électroniques (Softbook Press, Glassbook) et des constructeurs (Compaq, Philips, Xerox), afin de définir The Electronic Book Exchange System (EBX – [www.ebxwg.org](http://www.ebxwg.org)). L'EBX doit établir des spécifications pouvant prendre en compte tous les intervenants de la chaîne, de l'éditeur à l'utilisateur en passant par les libraires, les distributeurs, les bibliothèques. Ceci nous amène à évoquer un autre point sensible en matière de document numérique, celui du copyright. Le problème du copyright et du droit de copie dans son ensemble n'est pas nouveau et se posait déjà avec le document papier, mais les facilités de duplication et de transfert des documents dématérialisés ont rendu la question encore plus cruciale. Les problèmes juridiques sont au cœur de la problématique du document numérique. Lors de la mise en place de sa bibliothèque numérique, l'entreprise, pour tous les documents externes, devra s'assurer qu'elle dispose des droits de reproduction des documents qu'elle veut utiliser. Pour cela, l'entreprise devra, ou passer un accord avec les éditeurs détenteurs des droits ou signer un accord avec une société de gestion collective des droits de reproduction. En France, la perception et la répartition des droits de copie de la presse et du livre sont gérées par le Centre français d'exploitation du droit de copie (CFC). Mais celui-ci n'est agréé que pour les copies papier. De même, les problèmes de protection contre le piratage, de paiement en ligne, de suivi des accès aux documents numériques sont des questions encore difficiles à résoudre.

Des premiers éléments de réponse sur la protection contre le piratage sont données avec les techniques de tatouage ou watermarking. Cette technique consiste à glisser dans le document numérique des bits d'identification qui n'apparaissent pas à la consultation du document mais qui peuvent être révélés avec un programme spécifique. Les éditeurs de contenu numérique commencent à mettre en place des systèmes de gestion des droits des contenus numériques ou DRMS (Data Rights Management Systems), leur permettant d'assurer protection, suivi et paiement des droits sur les contenus numériques dont ils sont détenteurs. Les documents numériques de l'entreprise considérés comme sensibles en terme de confidentialité peuvent également être protégés par des algorithmes de cryptographie. Enfin, dernière question soulevée par les bibliothèques numériques, celle de la pérennité du document numérique. Cette pérennité se pose à trois niveaux : Le premier est celui de la conservation. Que faut-il conserver ? Il est certain qu'une large part des documents qui sont passés sur le Web a désormais disparue entièrement. Par ailleurs, il est hors de question de vouloir conserver ad vitam aeternam les millions ou milliards de pages Web. En l'absence aujourd'hui d'un dépôt légal des documents numériques virtuels (à l'exception des documents numériques sur un support physique tel que le cédérom), les pratiques sont encore bien hésitantes en ce domaine. Le second niveau est celui de la durée de vie des nouveaux supports, support magnétique et surtout support optique. En l'absence de toute certitude en la matière, certitude ne pouvant être apportée que par le temps, il convient de recopier régulièrement les documents numériques sur les supports jugés les plus sûrs. Enfin, et cela est sans doute le plus important, il ne faut pas oublier que le document dématérialisé nécessite un système de lecture sans lequel le document ne peut prendre vie. Ce système de lecture se décompose en deux éléments : un appareil de lecture et un logiciel d'accès. Or, si les premiers cédéroms, apparus maintenant il y a une quinzaine d'années, sont en parfait état de conservation, l'on ne dispose plus ni du lecteur de cédéroms nécessaire, ni du logiciel (logiciel d'accès et système d'exploitation). Deux possibilités s'offrent alors : la migration, c'est-à-dire la conversion d'un document dans des formats plus récents, au fur et à mesure qu'ils apparaissent, et l'émulation du

matériel, en sauvegardant, à l'aide de métadonnées, le logiciel d'application et le système d'exploitation. Ces problèmes d'archivage à long terme commencent à être pris en compte au niveau international avec la norme OAIS (Open Archival Information System), qui définit un modèle de référence d'un système d'archivage ouvert.

## Conclusion

Comme on l'a vu, la bibliothèque numérique offre de nombreuses possibilités à l'utilisateur, possibilités que le document papier, même s'il garde tout son attrait, ne peut lui apporter. Bien entendu, la mise en œuvre d'une bibliothèque numérique peut se heurter encore à un certain nombre de difficultés et même parfois de pièges. Mais, au fur et à mesure du développement des technologies, et avec l'expérience acquise en ce domaine, la maîtrise du document numérique est de plus en plus importante. La bibliothèque numérique, une fois les fonctions maîtrisées et les difficultés aplanies, est un formidable outil de management pour l'entreprise. La bibliothèque numérique est l'instrument de capitalisation du savoir de l'entreprise. Elle est la véritable mémoire de l'entreprise et constitue un patrimoine inestimable. Les fonds patrimoniaux des entreprises sont souvent d'une très grande richesse et méritent d'être portés à la connaissance du public, professionnel et grand public. Le document numérique peut être partagé par tous. Il n'y a plus la notion si importante de propriété du document qui freinent tant la circulation de l'information. A partir du moment où toute personne, selon bien sûr ses droits d'accès, peut accéder à tout document, la célèbre phrase l'information c'est le pouvoir perdra son sens. De plus, en matière d'information, l'intérêt ne réside pas dans la détention ou l'acquisition de celle-ci, mais dans sa valorisation. Combien de bibliothèques traditionnelles ne sont-elles pas de véritables tombeaux du savoir. La bibliothèque numérique, nous l'avons détaillé, offre des possibilités multiples pour valoriser le contenu du document. La bibliothèque virtuelle est aussi un outil d'ouverture, pour l'entreprise, sur le monde extérieur, cette ouverture pouvant d'ailleurs se faire dans les deux sens, du et vers ce monde extérieur. Enfin, on aura garde d'oublier que derrière la bibliothèque électronique, il y a des hommes et que la valeur de l'outil sera, en large part, fonction de la valeur des hommes qui l'animent. La bibliothèque numérique est le point de rencontre du manager qui définit les orientations de celle-ci en fonction de la stratégie de l'entreprise, du spécialiste de l'information (documentaliste, webmestre) à qui revient le choix des sources et le contrôle qualité, et l'informaticien responsable du choix des outils et de leur mise en œuvre. La bibliothèque numérique doit être une équipe au service d'une équipe plus importante.

## BIBLIOGRAPHIE

- ANDRE, J., HUDRISIER, H., dir, 2002, "Unicode, écriture du monde ?", Document numérique, vol. 6, n°3-4
  - BENSOUSSAN, A., LE ROUX, Y., dir, 1999, Cryptologie et signature électronique, Hermès
  - CHAUMIER, J., 1996, La gestion électronique de documents, P.U.F
  - CHAUMIER, J., DEJEAN, M., 2003, "Recherche et analyse de l'information textuelle. Tendances des outils linguistiques", Documentaliste- Sciences de l'information, vol. 40, n°1, pp14-24
- Document numérique, 1998, "Les bibliothèques numériques", vol. 2, n°3-4

- JACQUEMIN, C., dir, 2000, "Traitement automatique des langues pour la recherche d'information", Traitement automatique des langues, vol. 41
- JACQUESSON, A., RIVIER, A., 1999, Bibliothèque et documents numériques. Concepts, composantes, techniques et enjeux, Éditions du Cercle de la librairie
- LAFOUGE, T., LE COADIC, Y.F., MICHEL, C., 2002, Éléments de statistique et de mathématique de l'information : infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie, Presses de l'ENSSIB
- LE LOARER, P., 2000, "Les livres électroniques ou le passage", Documentaliste- Sciences de l'information, vol. 37, n°5-6, pp 298-311
- LE MOAL, J.C., HIDOINE, B., dir, 2000, Bibliothèques numériques, ADBS Éditions
- LICKLIDER, J.C.R., 1965, Libraries of the future, M.I.T. Press
- LUCAS, A., 1998, Droit d'auteur et numérique, LITEC
- LUPOVICI, C., 2000, "De la bibliothèque classique à la bibliothèque numérique : continuité et rupture", Documentaliste - Sciences de l'information, vol. 37, n°5-6, pp 286-297
- SAADOUN, M., 2000, Technologies de l'information et management, Hermès
- SALLERAS, B., DALBIN, S., 2000, "La nouvelle conception de l'intranet documentaire des AGF", Documentaliste - Sciences de l'information, vol. 37, n°3-4, pp 200-204
- STACKPOLE, L.E., KING, R.J., 1999, "Electronic journals as a component of the digital library", Science and Technology Librarianship, n°22
- SUTTER, E., 1998, Bibliothèques virtuelles et éducation : étude d'impact, Parlement Européen Terminologies nouvelles, 1996, "Banques de terminologie", n°15
- VAN SLYPE, G., 1997, "Les systèmes électroniques de gestion des droits d'auteurs : définition, objectifs, méthode d'analyse des besoins", Document numérique, vol. 1, n°3, pp 283-294 |cc| Jacques Chaumier - avril 2004