

EXIGENCES NUMERIQUES ET BESOINS DOCUMENTAIRES

Marie-Anne CHABIN
Consultant en archive
pour la revue SOLARIS, 1999

Résumé

Le recours à la technologie numérique pour la production, la conservation et la communication des documents, convertissant tout texte, image ou son en suites de 0 et de 1, décuple les possibilités de création et d'échange. Cet état de fait exige des standards technologiques et des conventions d'utilisation entre les interlocuteurs. Le créateur comme l'utilisateur doivent se former au maniement des outils et s'impliquer davantage dans le processus technique du document.

Il s'ensuit un infléchissement des besoins documentaires et, par conséquent, une problématique nouvelle pour les spécialistes de l'information. D'une part, l'écrit est documenté au moment même de sa création et l'utilisateur peut effectuer tout seul un nombre croissant de recherches. D'autre part, la pérennité des supports ou des normes elles-mêmes et la fiabilité d'une information surabondante constituent de nouveaux défis pour les professionnels du document, bibliothécaires, documentalistes et archivistes.

Abstract

The use of new technology at all the steps of the life-cycle of the documents, by transforming every text, image or sound in series of 0 and 1, increases the possibilities of creation and exchange of information. This doesn't work without standards. By the way creator and user have to improve their computer skills.

As a consequence, needs of information are changing, and new professional issues are arising. On one hand, at the very moment of their creation, documents are documented by metadata and the user is able to get more and more information by himself, without being helped. On the other hand, durability of medium, durability of standards themselves and reliability of information, are the new challenges for both librarians and archivists or records managers.

Mots-clés: numérisation, normes, utilisateur, producteur, bibliothécaire, documentaliste, archiviste, enjeux professionnels, durabilité, fiabilité.

Key-words: digitalisation standards, creator, user, librarian, archivist, records manager, professional issues, durability, reliability.

Introduction

Les documents sont des objets qui enrichissent les relations entre les hommes. Ils leur permettent de fixer ou d'enregistrer leurs idées ou leurs discours, de les diffuser et de les conserver pour un usage ultérieur ou pour les générations futures. La capitalisation de l'écrit au cours des siècles, le nombre des personnes impliquées dans la production ou la réception des documents de toutes natures, la croissance toujours plus rapide des échanges, suscitent l'apparition de conventions entre les différents acteurs du monde documentaire : les fournisseurs de supports, les auteurs et producteurs d'archives ou de livres, les utilisateurs, les conservateurs et "communicateurs" de documents, qu'ils soient archivistes, bibliothécaires ou documentalistes.

Ces conventions ou normes sont d'autant plus élaborées qu'elles font appel à la technique et à la production industrielle : composition et formats des papiers, qualité de l'encre, normes de conservation en matière de température, d'hygrométrie ou de luminosité, standards de fabrication des films, bandes magnétiques et disques, mais aussi règles de présentation du discours et de mise en page et, enfin, critères de classement et de description des documents.

Or, la technologie numérique redistribue les points de repère du monde documentaire, basé depuis cinq siècles sur les valeurs essentielles de l'imprimerie : matérialité du support, opposition manuscrit unique/imprimé multiple, séparation texte/image. L'homme numérique succède à l'homme typographique¹.

La dématérialisation, la délinéarisation, le multimédia et les réseaux bousculent les tenants et les aboutissants du document, mais de façons différentes pour ceux qui utilisent l'écrit comme support de leur activité et pour ceux qui font profession de préserver les documents et les contenus en dépit du temps qui passe. En effet, pour la production, les standards des supports numériques et les normes de transmission représentent un point de départ nouveau, une exigence absolue. Pour la gestion documentaire, en revanche, le numérique ne se traduit pas seulement par une modification du support et de la diffusion ; il déstabilise toute la fonction de médiation du spécialiste de l'information entre un producteur et un utilisateur car les besoins documentaires sont relatifs aux nouvelles possibilités que leur offre directement la technologie.

Exigences numériques

Tendance irréversible

La première exigence du numérique est qu'on ne peut l'ignorer impunément. Le progrès technologique poursuit son chemin et s'impose aux hommes de manière irréversible. Ignorer ou dédaigner l'ordinateur et les réseaux, aujourd'hui dans le monde professionnel et demain dans l'environnement personnel, est un facteur d'exclusion ou du moins de marginalisation, comme ont pu l'être en leur temps le refus du téléphone ou de l'automobile, ce qui n'a pas empêché de continuer de s'écrire, de se rencontrer, d'aller à pied ou à cheval quand le contexte s'y prêtait.

Le document numérique s'immisce progressivement dans la vie quotidienne. Des documents de l'administration et des services publics qui étaient naguère

accessibles sous forme papier font place peu à peu à des documents consultables sur des sites électroniques, tels les annuaires et les horaires de train, et même la réglementation. Des rapports, des notes, des revues transférés sur un support électronique sont à la fois plus facilement accessibles par cette voie et plus rares en mode traditionnel de communication car on ne produit plus qu'un petit nombre d'exemplaires papier.

Les réseaux favorisent la rapidité des échanges, laquelle accélère à son tour la nécessité de mises à jour, facilitées par le traitement électronique des données, etc., à la manière de l'œuf et de la poule.

Producteurs et utilisateurs peuvent en revanche bénéficier des avantages associés à la dématérialisation du support : les avis de concours et autres notes affichées sur le site Intranet de l'entreprise évitent l'impression et la diffusion de multiples exemplaires ; télécharger un formulaire Cerfa disponible sur Internet économise le temps et la fatigue dépensés si l'on devait prendre le métro pour aller faire la queue dans le bureau administratif qui délivre l'imprimé papier du même formulaire ; la transmission de photographies numériques par les réseaux dispense d'un développement en laboratoire et d'un envoi postal.

Le numérique se révèle ainsi très concurrentiel à toutes les étapes de la vie du document : la rapidité d'exécution d'une table des matières ou d'un index sur un ordinateur ne se compare même plus avec la confection manuelle des mêmes documents ; la correction d'un rapport urgent, la mise à jour d'un plan de chantier, la présentation d'un curriculum vitae, l'envoi d'une information à un client sans le recours à la technologie numérique est très lourde (bien que les systèmes d'information mal conçus et les pannes informatiques démentent quelquefois cette assertion).

Par ailleurs, le numérique façonne les mentalités. La logique numérique ne s'accommode pas de certaines formes de raisonnement, elle est étrangère à la numérotation qui inclut des bis et des ter quand il lui est si facile de renuméroter des milliers de fiches lors d'un ajout dans une série ; ou encore elle affranchit du soin de rédiger et de concevoir linéairement un texte quand il est si facile de pétrir son écrit dans le sens que l'on veut.

La dépendance, revers du progrès

Le numérique introduit une dépendance du créateur de documents vis-à-vis de la machine, voire vis-à-vis de ceux qui savent la faire fonctionner. La liberté de production et la rapidité de transmission autorisées par les ordinateurs et les réseaux sont conditionnées à la disposition d'outils appropriés, notion très évolutive. Celui qui a pris le pli de travailler ou de se détendre avec l'informatique doit payer son dû en épousant les matériels et les logiciels, en acceptant les bogues, en sacrifiant à la nouvelle divinité.

La technologie galopante fait entrer le producteur, comme l'utilisateur, de documents numériques dans une course à la nouveauté qui n'est pas favorable aux projets à long terme. La baisse des coûts est compensée par la nécessité de mettre à niveau les performances des outils en matière de capacité de stockage, de temps d'accès à l'information et de débit, au point qu'il est quelquefois difficile de discerner les progrès technologiques des influences commerciales ou politiques sur la distribution des produits et les tendances au monopole.

Les techniques numériques confrontent leurs utilisateurs à deux problèmes qui leur sont propres et pour lesquels les solutions sont en constante évolution : le document lu sur un écran a (ou pas) le même aspect que le document imprimé, mais il ne peut être " pris en main " d'emblée comme un document papier classique. L'écran étant l'interface obligatoire, l'objet en devient totalement différent. En effet, on peut voir le document à l'écran, avec la même apparence que son impression papier mais on ne peut le toucher, c'est l'écran de l'ordinateur seul que l'on peut toucher. Lorsque l'on reçoit un message électronique, ce n'est pas " le " message écrit par son auteur, c'est son image, son clone.

La pérennité et l'exacte conservation du document électronique sont le deuxième problème. Dès lors que le numérique transforme toute lettre, tout trait, toute image et tout son en une kyrielle de 0 et de 1, tout document produit et transmis est tricoté, détricoté et reticoté : au bout du compte, peut-on dire qu'il s'agit du même document ? On n'envoie plus un gâteau mais sa recette, ou mieux sa formule chimique qui, à l'arrivée, sert à composer un gâteau identique et qui pourtant est autre. Et puis, une main preste n'aurait-elle pas, entre deux, ajouté un peu de fleur d'oranger ou de belladone ?

Même lorsqu'on ne veut pas transmettre le document mais simplement le stocker, comment être sûr qu'on le retrouvera intact ? Cela dépend de trois conditions : le souvenir de la formule, la stabilité de support physique où elle est enregistrée et la disponibilité de l'appareil de lecture (machine à reticoter).

C'est pourquoi le recours au numérique exige à la fois des garanties technologiques sur la qualité et la stabilité des outils et des modes opératoires sur la façon de s'en servir.

La durabilité des supports est le défi du numérique car c'est son grand inconvénient vis à vis du papier ou du film. Pour tous les documents qui méritent une conservation à long terme, le recours au numérique requiert des garanties de stabilité et d'intégrité des documents : il faut que le document puisse être consulté aussi longtemps que l'on en a besoin (un an, dix ans, un siècle ou plus) et qu'il se présente exactement sous la forme sous laquelle il a été produit, faute de quoi il y aurait déperdition d'information et altération du document initial. Il faut en outre que cette stabilité et cette intégrité puissent être prouvées.

Le grand argument des réseaux est de mieux communiquer grâce à la possibilité de faire passer toute information (texte, image ou son) dans un fil après l'avoir codée. Communiquer à l'échelle de la planète exige de s'accorder sur systèmes cohérents et compatibles si on ne veut pas construire une tour de Babel numérique. Le préalable de la communication numérique est donc la normalisation.

Standards technologiques et normes d'échange

Standardiser, normaliser, structurer

La normalisation de l'écrit numérique s'effectue donc, parallèlement, dans deux directions : la recherche des performances technologiques propres à assurer la qualité de forme et la durabilité des documents ; la recherche conjointe par les ingénieurs et les utilisateurs de conventions d'écriture et de lecture.

Les standards technologiques purs sont indépendants de la nature des documents et de leur contenu. Il s'agit de fabriquer des formats de conversion et de fixation

de l'écrit, quel qu'il soit, au meilleur rapport qualité/prix pour les besoins des clients.

Les formats de conversion concernent d'abord la représentation numérique des signes qui doit être partagée par l'outil qui crée le document et l'outil de consultation : codage binaire des lettres (codes ASCII, Unicode), des traits (graphiques vectoriels), des images fixes noir et blanc ou couleur (codage d'un pixel sur 8, 16 ou 24 bits permettant jusqu'à 16 millions de couleurs), codage des signaux sonores et vidéo. Le transfert d'une image en format numérique impose ensuite de choisir un niveau de résolution (décomposition du document d'origine en un nombre de points plus ou moins élevé à convertir). Enfin, la répétition de certains signaux (ligne blanche, tache de couleur, silence, image peu animée) conduit à développer des algorithmes de compression et de décompression de l'information, permettant de réduire le volume des fichiers et le temps d'accès ; l'efficacité de certains algorithmes sur ce plan s'accompagne toutefois d'une perte d'information par rapport au document d'origine.

Le rôle de conservation et de véhicule de l'information joué par le support traditionnel éclate en deux champs : d'un côté, la matière et le mode de fixation de l'information numérique autour de deux grandes familles de supports (magnétiques avec la polarisation de particules, et optiques avec la gravure des informations numériques sur un disque de verre) ; de l'autre, les protocoles de communication des informations sur les réseaux, les plus connus étant TCP/IP² et HTTP.

Ces solutions technologiques s'imposent de fait ou sont validées par les organismes nationaux et internationaux de normalisation. Les formats, presque tous propriétaires à l'origine, finissent par voir quelques-uns d'entre eux accéder au statut de norme. Plusieurs standards peuvent coexister pour une même fonction selon l'aspect privilégié par l'utilisateur (par exemple, gain de place contre qualité de l'information).

Les questions de normalisation de la structuration des textes et de fiabilité des documents se placent sur un autre plan parce qu'elles sont relatives au contenu et à la portée des documents. Leur élaboration associe donc des ingénieurs, des juristes et des spécialistes des documents concernés.

Coder le message exprimé est une chose ; coder l'articulation du discours, la hiérarchisation des mots et des images en est une autre. La numérisation de documents longs et riches destinés à être une source de travail et de recherche exige la signalisation de la structure du document voulue par l'auteur et qui n'est plus évidente comme elle d'était sur un support traditionnel. La structuration permet d'introduire un balisage plus précis pour faciliter la navigation dans le document ; elle suggère également l'ajout de données sur l'environnement de production du document que le numérique rend transparent.

La structuration des documents numériques comporte deux étapes : définir la structure logique du document, et coder les éléments descriptifs de cette structure, en harmonie avec les solutions technologiques appropriées. Depuis la fin des années 1970, plusieurs langages ont été mis au point, avec une durabilité plus ou moins développée. Le plus répandu, élevé au rang de standard est le SGML³ (*Standard Generalized Markup Language*) qui permet à la fois, grâce aux définitions de type de document (DTD) de structurer le document et de décrire son environnement à l'aide de métadonnées. Quelle que soit la qualité conceptuelle d'une norme de ce type, elle est tributaire de la technologie qui l'accompagne et

subit de ce fait le contrecoup de son vieillissement. C'est ce qui semble se passer avec XML⁴ (*Extensible Markup Language*) face au "vieux" SGML. Peut-être faut-il en conclure que l'évolutivité des normes se normalise... En tout cas, le monde numérique exige d'être à l'écoute de l'évolution de ces normes, sous peine de n'être plus dans la course.

Un autre volet des normes d'utilisation de l'écrit numérique en liaison avec les performances technologiques concerne la valeur probante des documents, domaine lui aussi très évolutif au gré des certitudes techniques et des concessions culturelles. Pendant des années, on a répété qu'un document numérique ne pouvait faire foi ; puis les échanges de factures et de documents comptables se sont développés dans le cadre de l'Edifact⁵ (*Electronic data interchange for administration, commerce and transport*) ; la jurisprudence avance doucement à reconnaître comme élément de preuve des documents électroniques ; on parle aujourd'hui de légaliser la signature électronique.

Implication de l'utilisateur

Au niveau de l'utilisateur qu'est le créateur ou le lecteur d'un document numérique, le constat qui s'impose est que les standards technologiques aussi bien que les normes d'échange l'obligent à s'impliquer dans le processus technique de production ou de recherche.

Plus un document numérique est riche, plus il a besoin de structuration pour faciliter le repérage de l'information. Or, contrairement à l'écrit traditionnel ou aux enregistrements analogiques dont l'articulation est visible et peut être relevée et détaillée après coup, posément, lors d'une lecture ou d'un visionnage du document, le document numérique, lui, exige d'être structuré dès sa création, faute de quoi l'information ne peut même pas être enregistrée. C'est pourquoi, les logiciels associent de plus en plus l'auteur du document à la formulation de ces éléments de structuration qui constituent dès lors également des éléments de description.

La rédaction d'un rapport sous une forme numérique sera accompagnée de la confection de la table des matières et d'un index, ainsi que d'une notice comportant la date, le nom de l'auteur, quelques lignes de résumé, des mots-clés. Le maniement d'une caméra numérique sollicite, pour chaque séquence individualisée, une indexation sommaire du contenu comme balise de cette partie du fichier.

De sorte que les fonctions de création, de fabrication et de documentation, bien différenciées dans l'environnement papier ou analogique, tendent à voir leurs frontières s'estomper. Le numérique tisse des liens vitaux et indissolubles entre le contenant et le contenu, introduit une interdépendance entre le technicien, le créateur et le gestionnaire de l'information.

Parallèlement, la multiplication des hyperliens, la réalité des performances des moteurs de recherche (même si leur qualité mérite d'être discutée), le développement des réseaux et l'accroissement des informations en ligne, permettent au chercheur d'accéder de façon autonome à un plus grand nombre de documents et, de plus en plus, incite l'utilisateur des réseaux à consulter directement l'information.

Cette implication des auteurs de documents et des chercheurs d'information (on est souvent l'un et l'autre) dans la structuration et la description de l'écrit

nécessite une formation spécifique aux outils qui, toutefois, entre peu à peu dans la formation générale, voire dans la culture générale.

Il est important de souligner que cette évolution concerne l'écrit sous toutes ses formes et quelle que soit sa fonction : base de données, texte de loi, contrat juridique, documentation commerciale, reportage audiovisuel, œuvre littéraire, statistiques, rapport scientifique, etc. La problématique numérique est commune à tous les types de document puisque tous les types de document se retrouvent sous un format numérique et de plus en plus dans une forme multimédia.

Besoins documentaires

Les cinq composantes du besoin documentaire

Quels que soient les outils, quels que soient les personnes impliquées ou les documents en cause, le besoin documentaire se définit toujours d'une manière relative comme la recherche, par un utilisateur, d'une information ou d'une pièce (complémentaire à la documentation qu'il possède déjà) au sein d'un ensemble de documents et de données existantes ou supposées exister, ensemble qui, s'il est important, est géré par une personne spécialisée.

Il y a donc cinq composantes à l'expression du besoin documentaire : l'utilisateur, la question, le fonds dans lequel est dirigée la recherche, les outils, et le spécialiste de l'information qui joue le rôle de médiateur.

Dans l'environnement qualifié de traditionnel, l'utilisateur dispose d'un lot de dossiers et d'ouvrages les plus récents ou les plus usuels conservés dans le bureau ou dans le couloir voisin et dans lequel il se repère grâce à sa mémoire, un fichier manuel ou une base de données personnelle. Si l'information recherchée ne se trouve pas là ou qu'il n'ait pas le temps ou le savoir-faire pour la retrouver, il s'adresse au centre de documentation, au service d'archives ou à la bibliothèque et à ses professionnels.

L'environnement numérique recompose le cadre d'expression de ce besoin, ou plus exactement provoque une translation de ce cadre sans véritablement en changer les proportions. Le changement se caractérise par : la production d'informations plus nombreuses et plus fugaces, la participation des producteurs à la description des documents, la recherche d'informations plus nombreuses et plus urgentes, la mise en ligne sur l'Intranet de l'entreprise ou sur les réseaux mondiaux de millions de pages d'informations plus ou moins denses et plus ou moins fiables, des possibilités très simples d'interrogation en texte intégral.

L'équilibre entre les composantes du besoin documentaire se trouve donc modifié mais pas nécessairement rompu. L'utilisateur dispose de plus d'informations mais en cherche encore davantage. Les fonds documentaires s'enrichissent et s'ouvrent à une allure vertigineuse. Les outils vont très vite et pointent sur une information, même si ce n'est pas celle que l'on recherche ; ils permettent aussi de contacter dans l'instant quelqu'un qui sait, à des milliers de kilomètres, ou de contacter plusieurs centaines de personnes à la fois par le biais des forums.

Les questions de leur côté se multiplient mais, facteur de permanence, sont toujours de deux types : ou bien on recherche un document précis parce qu'on l'a déjà vu, que l'on a sa référence ou que l'on sait par ailleurs qu'il doit exister (une facture donnée, un texte de loi, les vœux 1995 du président, le dernier livre de poèmes d'un tel) ; ou bien on recherche une information, un contenu, sans

préjuger du document qui la contient (une photographie d'Edith Piaf, une description de sous-marin, des témoignages sur le travail du fer forgé, une radio de foie atteint de cirrhose, etc.).

Enfin, le spécialiste de l'information, en bout de chaîne en quelque sorte, subit cette grande mutation ; une partie de son rôle traditionnel lui échappe puisque les producteurs se mettent à documenter leurs écrits et que les utilisateurs s'efforcent de chercher tous seuls. Il est secoué dans ses habitudes et poussé à réussir la mutation de son savoir-faire.

Maintenance du support et de l'information

Pour trouver l'information recherchée, il faut que les fonds soient accessibles. La première condition à leur accessibilité, c'est qu'ils aient été conservés. Là, l'instabilité et la fugacité du document numérique tirent le besoin documentaire en amont, soulignent la subordination du contenu par rapport au support, élargissent la responsabilité du médiateur documentaire.

On a pu considérer un temps que la tâche de l'expert ès documents était d'éliminer les documents inutiles pour ne pas être submergé par une production exponentielle. La production est telle aujourd'hui qu'il est de plus en plus manifeste que s'employer à éliminer une information inutile, est moins efficace que de s'attacher à sélectionner et à sauvegarder (*capture* en anglais) l'information pérenne, car l'information numérique est condamnée à périr d'elle-même

La maintenance des documents (supports et informations) susceptibles de répondre aux besoins des utilisateurs appelle une intervention sur trois fronts. Primo, le développement et la normalisation du document numérique ne doit pas faire oublier que toute l'information n'est pas systématiquement numérique, numérisée, numérisable. Secundo, la durée de vie de l'information est souvent courte et il faut agir avant sa mort. Tertio, le support doit être entretenu pour rester lisible.

On entend dire que 80 % de l'information électronique dans l'entreprise existe aussi sous forme papier, ce qui conduit à s'interroger sur les relations d'antériorité et de supériorité qu'entretiennent les deux supports concurrents sur le plan de la longévité, de la disponibilité, de la fiabilité. Tous les cas de figure se rencontrent : brochure papier éditée à partir de la maquette électronique enregistrée sur une disquette peu durable ; registre pluriséculaire en très mauvais état sauvegardé sur un cédérom ; statistiques à jour sur un site Web mais plus fiables dans une publication, etc. Les passerelles entre les deux familles de supports méritent elles-mêmes d'être normalisées en fonction de l'usage qui est fait des contenus et surtout du déroulement dans le temps de cet usage.

Les documents numériques présents sur les réseaux offrent une problématique particulière. La facilité de production de l'information accroît le nombre de documents mis en ligne, lesquels se superposent aux documents précédents. Il est facile de trouver n'importe quoi sur Internet, il est plus difficile de le retrouver une seconde fois. La nouvelle information chasse rapidement l'ancienne sans que sa valeur d'information pour l'utilisateur soit nécessairement éteinte. D'après une enquête réalisée en juin 1999 par l'OCLC⁶, (Online Computer Library Center), la fugacité (*volatility*) des informations est de 44 %, c'est-à-dire qu'une information affichée sur le Web à 44 % de chance de ne plus s'y trouver un an plus tard⁷. Une autre étude, menée par Brewster Kahle, promoteur du projet *Archiving the net*⁸,

concluait en 1996 à une durée de vie moyenne des documents Web de 75 jours et constatait que de plus en plus d'utilisateurs recherchant une référence Internet restaient bredouilles, le site ayant disparu ou ayant été modifié. Brewster Kahle lança alors une campagne de sauvegarde des sites ou des pages en danger de disparition en dépit de la validité de leur intérêt documentaire.

La préservation de ces documents numériques, basée sur la compétence scientifique des professionnels de l'information, est tributaire de la compétence technologique de traitement des formats et des standards. Il faut donc conserver non seulement les données que l'on voit à l'écran, mais encore leur environnement technologique et les métadonnées de provenance, de date, de droits ou autres.

En résumé, l'expert documentaire ou médiateur de l'information, doit développer sa connaissance des supports, au sens large que la technologie numérique peut conférer à ce mot, à savoir la chaîne physico-numérique qui permet de matérialiser le message : matière, format d'enregistrement et outil de lecture. Plus exactement, il doit maîtriser les normes, du point de vue de leur performance bien sûr, mais surtout du point de vue de leur durabilité et de leur adéquation aux besoins des utilisateurs.

L'abondance d'information n'élargit pas les possibilités de la mémoire humaine, même si les outils permettent de l'assister efficacement ; la compétence de l'expert démarre là où s'arrête la mémoire outillée de l'utilisateur, frontière non absolue mais relative à l'environnement

La description pérennisée et la notion de greffe

La réponse au besoin documentaire passe par les instruments de recherche dont la qualité technique et l'adéquation aux utilisateurs autorisent un accès plus ou moins rapide au document recherché ou au document pertinent pour la recherche.

Les notices descriptives des instruments de recherche sont élaborées par les professionnels de la chose écrite à partir des informations portées par le créateur du document sur ce dernier ou sur son conditionnement (annotations sur la chemise d'un dossier, étiquette sur une cassette) et de l'analyse du document lui-même, avec l'aide, le cas échéant, de dictionnaires et de documentation complémentaire.

Le document numérique normalisé intègre des données sur son contexte de création ainsi que les mots-clés sollicités par le logiciel pour le balisage des fichiers. Ces éléments de description intégrés par l'auteur lors de la fabrication du document sont, par définition, plus nombreux que pour un document sur support traditionnel. Ils sont liés à l'action que le document supporte à ce moment-là, à sa valeur de preuve ou valeur primaire ; par ailleurs ils sont d'abord destinés à prolonger la mémoire de l'auteur et ne tiennent pas compte d'un autre besoin ou d'une autre utilisation. C'est dire qu'ils sont sujets à se périmer s'ils ne sont pas consolidés et pérennisés par une description professionnelle. Cette description professionnelle a pour but, à la fois, de compléter les informations utiles à l'utilisateur-non-auteur et d'harmoniser cette description, par le biais d'un thesaurus notamment, avec le reste du fonds.

À partir du moment où ces éléments de description initiaux existent, la question est de savoir s'ils sont récupérables, au plan technologique d'une part, au plan documentaire de l'autre. Car peut-on raisonnablement décrire un document, écrit textuel ou audiovisuel, comme s'il s'agissait d'un document vierge de toute

indexation, de toute métadonnée, de toute annotation intégrée au document lui-même, alors que la technologie permet de capitaliser ces informations et qu'elles représentent une valeur ajoutée pour la compréhension du document ?

Sur le plan documentaire, deux attitudes sont possibles : ou bien on considère que ces métadonnées sont parties intégrantes du document, auquel cas elles doivent être prises en compte dans la conservation du document ; ou bien on considère qu'il s'agit là de données périphériques et provisoires qu'il faut reconstituer entièrement. S'il est vrai que les mots-clés notés par un journaliste sur son reportage ne sont pas ceux qui conviennent à une exploitation ultérieure, patrimoniale ou commerciale, des archives audiovisuelles, ils présentent malgré tout deux avantages : permettre une recherche approximative en l'absence d'une notice scientifique, et donner des indices sur la manière dont l'auteur voit son document et l'utilisation qu'il peut en faire, ce qui, en soi, représente aussi une information.

La technologie numérique permet et suggère d'introduire dans l'analyse documentaire la notion de *greffe* : préserver comme moyen d'accès transitoire et comme témoignage les données initiales formulées par le créateur de document, et y ajouter, y greffer les éléments d'une analyse professionnelle propres à garantir une exploitation plus large et plus durable de l'information. Le travail du professionnel du document se fait alors en deux temps : valider la première zone de description, rédiger la seconde.

Critique et fiabilité de l'information

Dans de nombreux centres de documentation, les premières bases de données n'étaient pas en accès libre et étaient manipulées par des spécialistes de la recherche, du fait de la technicité des opérations auxquelles l'utilisateur n'était pas formé et, peut-être aussi, en raison d'informations potentiellement confidentielles. À partir du moment où les outils le permettent et que l'utilisateur sait effectuer une recherche par mot-clé ou en texte intégral dans une base de données, dans un corpus de textes ou une base d'images, au nom de quoi freinerait-on le libre accès à ces informations ? L'argument protectionniste de préservation de l'activité de ceux qui font ce travail d'intermédiaire depuis dix, vingt ou trente ans ne tient pas. Il y a tant de choses à faire ailleurs !

Il ne fait aucun doute que, si l'accès à une information ou à un document est désormais l'affaire de tous, la critique de la pertinence de cette information et l'évaluation de sa fiabilité restent du domaine des professionnels du document. C'est là qu'il faut renouveler les principes d'analyse et de gestion documentaire, et développer des méthodes.

Une nouvelle ligne de partage apparaît dans la satisfaction du besoin d'information. D'une part, les réponses à une recherche documentaire qui peuvent être absolues, qui se suffisent à elles-mêmes ; d'autre part, les réponses qui doivent être relativisées, qui requièrent une évaluation de leur pertinence découlant à la fois de la fiabilité de l'information et de sa représentativité dans le champ de toutes les réponses possibles.

La recherche peut être terminée dès que la question obtient un résultat (même un seul). Soit il existe une seule réponse possible (exemple : le texte de la loi 83-177), soit la question n'exige pas un support documentaire particulier et se contente d'une réponse de type informatif (une notice un peu détaillée sur les

farines animales, une image de jardin en hiver offrant telle et telle caractéristique technique et esthétique). Toute personne un peu rompue au maniement d'Internet peut obtenir une réponse à ce genre de question. Il est d'ailleurs à remarquer que l'offre dans ce domaine alimentaire la demande et crée le besoin.

De l'autre côté, l'objectif professionnel qui a motivé le besoin documentaire, ou encore la formulation de la question exige une recherche exhaustive et une vérification des sources : ce texte sur le droit d'auteur est-il le seul sur ce sujet, est-il le plus récent, est-il complet ? Ces images d'accident d'avion sont-elles les seules qui aient été filmées et qui soient accessibles, ont-elles pu être truquées ? Ces données sur la population de patients atteints par telle pathologie au sein de notre établissement hospitalier sont-elles exhaustives ?

La réponse à ces questions assez pointues nécessite une technicité que la technologie facilite mais ne remplace pas. La rapidité d'une recherche dans une pléthore d'informations non organisées donne souvent une fausse impression d'exhaustivité qui ne résiste toutefois pas longtemps à une démarche d'ordre juridique (vérité de l'information) ou économique (rentabilité ou concurrence).

Le spécialiste de l'information a donc fort à faire pour maîtriser l'étendue et la provenance des informations. On a évoqué le phénomène de la masse produite trop imposante pour souffrir un tri et une élimination de l'inutile, cédant la place à la démarche inverse de sélection de l'utile. Ce phénomène se retrouve au niveau de la recherche documentaire : il s'avère souvent trop lourd de trier les trop nombreuses réponses possibles à une demande de recherche documentaire et de choisir dans la liste celle qui répond le mieux au besoin de l'utilisateur. Il ne s'agit plus de trouver des documents sur le sujet, il s'agit de trouver le ou les plus pertinents pour l'objectif final, en sélectionnant la source de l'information par rapport à sa réputation de fiabilité et un profil de document plus précis qu'un simple thème de recherche (des documents sur les chevaux, des vues de Paris). Il faut introduire en amont des critères de recherche plus subtils relatifs à l'auteur, à la forme ou au mobile du document, ce que la normalisation permet de plus en plus.

Les travaux en cours sur la normalisation des métadonnées vont dans ce sens, que ce soit les propositions de 15 éléments de description du Dublin Core⁹ pour le catalogage des documents électroniques ou les expérimentations conduites dans le cadre de la problématique posée par le records management dans un environnement numérique.

Conclusion : pour une convergence des métiers

Le fait d'avoir parlé jusqu'ici d'expert documentaire ou de médiateur de l'information, sans préciser davantage s'il s'agissait de bibliothécaires, de documentalistes ou d'archivistes, est délibéré. C'est que, technologie aidant, il s'agit dans la plupart des cas des trois professions. La transversalité du numérique et le poids qu'il donne aux normes techniques dans la production et le traitement scientifique des documents contribuent à gommer les distinctions traditionnelles entre l'imprimé et le manuscrit, entre le "document d'archives" unique et original et le "document de documentation" multiple et non original.

Les normes d'enregistrement, de reproduction, de structuration et de description des informations numériques visent tout écrit, au sens large du mot, c'est-à-dire

toute fixation ou enregistrement de l'expression humaine au moyen d'un langage, combiné ou non avec la capture du réel (son, image fixe ou image animée).

Que le document soit destiné à la diffusion comme un livre ou une œuvre audiovisuelle, ou qu'il supporte une activité administrative ou technique, il bénéficie et subit les mêmes normes, en fonction de sa longueur, de sa taille, du rapport texte/image/son, de sa valeur de preuve ou de témoignage.

Ce rapprochement entre les documents conservés dans les bibliothèques, les centres de documentation ou les services d'archives, voire dans les musées, est accéléré par la technologie numérique mais on en trouve des signes avant-coureurs depuis une vingtaine d'années : les documents figurés ou audiovisuels et la littérature grise en donnent deux exemples.

Où trouve-t-on les brochures publicitaires, les rapports annuels, les prépublications, les rapports d'étude, les tirés-à-part, désignés collectivement par l'expression de littérature grise ? Aussi bien dans les dossiers de documentation archivés par les services administratifs ou les équipes de recherche, que dans des collections constituées aussi régulièrement que possible par des bibliothèques, ou que dans des dossiers thématiques alimentés au gré des informations disponibles par des documentalistes¹⁰. De même, les cartes postales, les photographies, les diapositives, les affiches, les cassettes audiovisuelles (qu'elles proviennent d'opérations de communication ou accompagnent des projets de construction), sont conservées tantôt dans un centre de documentation, tantôt aux archives ou encore à la bibliothèque ou au musée, parfois dans plusieurs lieux. L'utilisateur, lui, ne saisit pas toujours la logique de cet éclatement.

La convergence technologique s'accompagnera nécessairement d'une convergence des métiers, quitte à voir se développer demain d'autres spécialisations. On parle aujourd'hui de l'utilisation de SGML pour la conservation des archives électroniques¹¹ ; il n'y aurait eu aucune pertinence jusqu'ici à envisager l'application du format MARC¹² au catalogage des archives administratives ni celle de la série W au rangement des livres.

Les prises de position se multiplient pour prêcher des rapprochements professionnels entre les métiers de l'information. Dominique Arot, secrétaire général du Conseil supérieur des bibliothèques, évoquant la place des bibliothèques, archives et musées, institutions culturelles, institutions de mémoire dans les programmes de recherche de la Commission européenne, écrit : " Les points communs de ces institutions (importance de l'image, formation, normalisation, conservation) semblent encore renforcés, tout en disposant d'atouts inédits "¹³. Jean-Pierre Wallot, ancien directeur des Archives nationales du Canada et ancien président du Conseil international des archives, insiste : " Au moment où les autorités et les citoyens requièrent l'information (sans ce préoccuper, il faut le comprendre, des " territoires " des différentes disciplines), les archivistes dominant-ils leur univers documentaire et ont-ils créé les liens nécessaires avec les gestionnaires de documents, les bibliothécaires et les documentalistes pour exhiber l'ensemble de la documentation pertinente (archivistique ou publiée) "¹⁴.

Si la valeur probante des documents, le cercle de leur diffusion, leur caractère plus ou moins confidentiel, continuent de jouer un rôle non négligeable dans la gestion des documents et de l'information, la problématique de collecte, de conservation

et d'exploitation, infléchiée par la normalisation du document numérique, rend cette convergence professionnelle inéluctable. Mais il faut l'organiser.

Marie-Anne Chabin.

L'auteur vient de publier :

Je pense donc j'archive. L'archive dans la société de l'information.

L'Harmattan, Paris, France. 208 pages, 110 F ISBN 2-7384-8340-2

Liens pour poursuivre la lecture sur quelques thèmes : métadonnées, standards électroniques, conservation des documents électroniques. Les sites les plus riches sont généralement hébergés aux USA.

- *Metadata and Data Management Info Page*

http://www.llnl.gov/liv_comp/metadata/metadata.html

- *UKOLN* (The UK Office for Library and Information Networking) *Metadata*, GB.

<http://www.ukoln.ac.uk/metadata/>

- *The Dublin Core: A Simple Content Description Model for Electronic Resources.*

<http://purl.oclc.org/dc/>

- *LBNL* (Lawrence Berkeley National Laboratory) *EPA* (Environmental Protection Agency) *Scientific Metadata Standards Project*. Washington, USA.

<http://pueblo.lbl.gov/~olken/epa.html>

- *The Pittsburgh Project*. University of Pittsburgh School of Information Sciences, USA.

<http://www.lis.pitt.edu/~nhprc/>

- *CoOL* (Conservation OnLine. Resources for Conservation Professionals), a project of the Preservation Department of Stanford University Libraries. USA

<http://palimpsest.stanford.edu/>

- *Electronic Records Research*. Working Meeting. May 28-30, 1997. David Bearman and Jennifer Trant. Archives & Museum Informatics, Pittsburgh, Pennsylvania

<http://www.dlib.org/dlib/july97/07bearman.html>

Articles en ligne :

- *The Preservation of the Integrity of Electronic Records*. Luciana Duranti, Principal Investigator, Terry Eastwood, Co-Investigator, Heather MacNeil, Research Assistant. School of Library, Archival & Information Studies, University of British Columbia, Vancouver, B.C. USA.

<http://www.slais.ubc.ca/users/duranti/index.htm>

- Simon Pockley. Killing the Duck to Keep the Quack.

<http://www.cinemia.net/FOD/FOD0055.html>

- Stevan Harnad. Plusieurs articles de cet auteur prolifique, notamment sur les publications électroniques.

<http://www.princeton.edu/~harnad/intpub.html>

Sur les archives en général

Un site qui indexe l'ensemble des services d'archives.

<http://www.tulane.edu/~lmiller/ArchivesResources.html>

¹ McLuhan (Marschall), *La Galaxie Gutenberg, la genèse de l'homme typographique*, Éd. H.M.H. Ltée, 1967 [éd. originale 1962]

Négroponte (Nicholas), *L'Homme numérique*, Lafont, 1995.

² Voir pour plus de détails : <http://www.rtd.com/pcnfsfaq/faq.html> (les questions les plus fréquentes, "FAQ", sur TCP/IP pour Mac et PC, en anglais) ou "Introduction to the Internet Protocols" : <http://oac3.hsc.uth.tmc.edu/staff/snewton/tcp-tutorial/> ou, bien sûr, le site de la CNIL (Commission nationale de l'informatique et des libertés) : <http://www.cnil.fr/>

³ <http://www.oasis-open.org/cover/>. *The SGML/XML Web Page* is a comprehensive online database containing reference information and software pertaining to the Standard Generalized Markup Language (SGML) and its subset, the Extensible Markup Language (XML).

⁴ <http://www.xml.com/>

⁵ <http://www.mcis.duke.edu/standards/other-standards-organizations/edifact.htm>

Voir aussi le site ISO : <http://www.iso.ch/welcome.html>

⁶ <http://www.oclc.org/>

⁷ Signalé par Dean DeBolt sur le forum Archives & Archivists le 14 septembre 1999 : <http://listserv.muohio.edu/archives/archives.html>

⁸ "Internet Archive" de B. Kahle : <http://www.archive.org>

⁹ "The Dublin Core: A Simple Content Description Model for Electronic Resources."

<http://purl.oclc.org/dc/>

¹⁰ Voir notamment Jacquesson (Alain) et Rivier (Alexis), *Bibliothèques et documents numériques. Concepts, composantes, techniques et enjeux*, Ed. du Cercle de la Librairie, 1999, p 341

¹¹ Commission européenne, *Actes du DLM-Forum sur les données lisibles par machine, Bruxelles, 18-20 décembre 1996*, Luxembourg, 1997, 374 p.

Voir le site : <http://www.dlmforum.eu.org/>

¹² On peut trouver des informations sur MARC (Machine Readable Cataloguing) et UNIMARC sur le site de l'IFLA (International Federation of Library Associations) : <http://ifla.inist.fr/index.htm>

ou sur celui de la Bibliothèque du Congrès : <http://lcweb.loc.gov/marc/marc.html>

¹³ Arot (Dominique), "Les bibliothèques françaises à l'âge du numérique", dans "Bibliothèques numériques", numéros spécial de *Document numérique*, vol. 2, n° 3-4, 1998

¹⁴ Jean-Pierre Wallot, "L'accès à l'information : les défis technologiques, examen de la pratique archivistique", dans *Actes de la trente-troisième Conférence internationale de la table ronde des archives (CITRA)*, Stockholm, 1998, p 84.

Voir le site du Conseil International des Archives : http://data1.archives.ca/ica/cgi-bin/ica?05_f