

Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives – Louvain-la-Neuve 20-21 Mai 2007

Notre apport à ces journées sera peut-être très pragmatique et factuel, en tout cas dans la première partie où nous nous concentrerons sur la gestion courante des archives numériques et des archives que l'on numérise. Dans la seconde partie, nous tenterons quelques réflexions, fondées notamment sur les expériences décrites dans la première partie et qui portent sur 35 années de pratique du métier d'application de l'informatique aux textes et aux documents: cette gestion informatisée apporte-t-elle une valeur ajoutée à l'archivage? et, dans l'affirmative, avec quels risques?

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples ►

1.2 Choix des supports ►

1.3 Procédures ►

1.4 Le microfilmage ►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►

R.-Ferdinand Poswick, o.s.b. (Informatique & Bible, Maredsous)

Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives - Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents

Nous commençons par déplacer et élargir l'objet traité informatiquement, car il nous semble que l'archivistique et les archives ne sont que l'une des branches dans lesquelles le traitement informatique de textes et de documents intervient. Peu importe, en effet, qu'il s'agisse d'un inventaire d'archives, du descriptif d'un document, de la préservation d'un texte ou d'une image, du traitement de supports visuels ou auditifs, voir de données déjà numérisées. On trouve tous ces types de données dans des dépôts d'Archives, et tous peuvent se voir appliquer avec une amplitude variée des traitements informatiques.

Une autre précision à donner concerne la notion-même d'erreur. Faire un choix pour constater à moyen terme qu'il eût mieux fallu en faire un autre, est-ce vraiment une erreur? Il y aurait vraiment erreur dans deux cas:

- a) si le choix a été fait sans information préalable et sans tests suffisants: la foi béate (qui est crédulité) mène à la faute!;
- b) si le choix, même s'il est réalisé avec une bonne information critique, ne répond pas aux objectifs désirés, ou encore lorsque ce choix est trop rapidement disqualifié par un développement technologique difficile à prévoir même avec une bonne habitude et une bonne méthode de veille technologique.

Les exemples qui suivent doivent être reçus en tenant compte de cette double précision sur l'extension du domaine envisagé (tout ce qui peut être mémorisé avec l'écriture électronique) et sur le caractère fragile et mouvant de la frontière entre choix (*airésis* en grec) et erreur (hérésie, en français!).

Pour ce dernier aspect, nous attirons l'attention sur le fait que le progrès technologique lui-même fait des choix que l'on peut juger "erronés" et qui, pourtant, sont imposés à la plus large partie des utilisateurs par cette volonté marchande (*merchandising*) sous l'empire de laquelle nous nous trouvons presque tous, souvent de mauvais gré! L'exemple le plus connu dans le domaine de l'audio-visuel est le choix, à l'époque, du standard VHS pour la distribution de documents visuels mobiles, choix effectué au détriment d'autres standards comme le BETACAM, bien meilleur, mais qui est resté dans le cercle des seuls professionnels du domaine (avec tous les problèmes – et l'industrie complémentaire en découlant! - du transfert d'un standard vers l'autre). Et, pour le standard VHS, la découpe, liée à des stratégies géopolitiques de marketing, entre les différentes normes: PAL, SECAM, NTSC. Un découpage que l'on a essayé de reproduire avec les différentes "zones" pour les DVD! Le bibliothécaire multimédia qui a décidé, dès son apparition dans nos pays, de se constituer une bibliothèque de documents audio-visuel en VHS-PAL, a-t-il fait une erreur? Non. Même s'il s'est privé d'une meilleure qualité de documents; même si, en attendant un lecteur multistandard, il n'a pu accéder à un certain nombre de documents de valeur; même si, aujourd'hui, toute sa bibliothèque VHS risque de devenir inutilisable à court terme, rendue obsolète du point de vue de la qualité d'image et de son par les productions purement numériques que le marché propose sur DVD et qui vont faire disparaître les bandes vidéo aussi rapidement que le disque vinyle au moment du passage au CD-audio1?

Dans les exemples qui suivront, il faut avoir ces réserves à l'esprit. Elles illustrent par elles-mêmes notre propos: l'évolution technologique peut, effectivement, être source

d'erreur dans la gestion, la conservation et la distribution de données mémorisées.

Encore un mot: nous tenterons de nous limiter à des exemples que nous avons effectivement rencontrés dans notre parcours professionnel, chez nous ou chez des clients avec lesquels nous avons travaillé.

1.1 Choix de matériels: deux exemples ►►

1.2 Choix des supports ►►

1.3 Procédures ►►

1.4 Le microfilmage ►►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►►

1. Voir: Derek G. Law, *Back to basics: A-Literacy, the Boolean gene, convergence and the long tail*, The electronic Library, 24/6, p. 73 "Those of a certain age will remember the war over video recorders where the consensus was that Betamax was technically far superior to VHS. But VHS was the popular choice and the financial success – except that both were overtaken by DVD".



Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives – Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples

– En 1981, nous avons déterminé que, pour diffuser des Bibles numérisées, il fallait offrir parallèlement des services informatiques adaptés aux besoins spécifiques de trois bassins de clientèles : les bibliothèques, les paroisses et les écoles. Pour ce faire, nous avons conçu et mis au point l'un des tout premiers programmes de catalogage sur P.C. pour des bibliothèques de taille moyenne ou petite. Apple venait de sortir son premier micro-ordinateur et cheminait vers l'Apple-II qui sera son premier vrai succès commercial. Mais plusieurs grandes firmes venant de l'informatique lourde (IBM, Bull, DEC, WANG, etc.) préparaient la riposte et IBM sortira effectivement son premier Personal Computer (P.C.) en 1982.

Face à cette bataille industrielle, comment choisir pour assurer un développement qui puisse se fonder sur un matériel et un système d'exploitation ayant un véritable avenir commercial?

Nous avons fait appel, à l'époque, à des chercheurs d'un célèbre Institut d'Informatique belge. Ils nous ont remis une étude, au demeurant fort bien documentée, qui concluait catégoriquement en faveur du système d'exploitation P/OS des PDP-11 de DEC (Digital Equipment) sous la forme du Professionnal 350, à cette époque compatible avec les programmations de haut niveau comme le COBOL que nous utilisions alors.

IBMistes convaincus puisque c'était dans cet environnement que nous avons travaillé jusque là depuis plus de dix ans, nous avons hésité jusqu'à l'extrême limite des délais de décision pour nous ranger, finalement, à l'avis universitaire qui, à l'époque, et faute de moyens pour effectuer nous-mêmes une veille technologique efficace, nous semblait critique et indépendant.

Or, expérience faite (dès 1984 nous faisons la première présentation de notre produit sur micro(?) DEC à tous les bibliothécaires ecclésiastiques de Rome réunis pour la première fois à cette occasion), nous avons dû constater que le micro-ordinateur de DEC ne se vendait pas, et que, par contre le PC d'IBM faisait sa percée sur tous les marchés, associé au premier DOS de Mr Bill Gates!

D'où la nécessité d'une nouvelle année de programmation pour transférer toute l'application dans l'environnement PC/DOS d'IBM.

Bref, beaucoup de temps et d'argent perdus ... et pour nous apercevoir, par la suite, qu'il entrainait une bonne dose d'idéologie anti-IBM – probablement partiellement justifiée – dans les conseils universitaires que nous avons reçus.

– Un second exemple est tout récent: un client avait constaté que certaines de ses archives étaient pauvrement stockées avec un risque de détérioration par l'humidité. Or il s'agissait, à ses yeux, d'archives importantes. Et, dans l'immédiat, il n'avait pas les moyens de relocaliser ce fonds.

Il a donc pris la décision de microfilmer ce fonds. Par une brève analyse et la persuasion d'un bon service de marketing, il s'est laissé convaincre d'acheter une machine performante permettant de créer des films non-développés d'images microfilmées à partir de documents en format A4. Une machine chère à l'achat (65.000 €uros) et chère à la maintenance (8.000 €uros par an).

Or, malgré une personne formée à l'utilisation de cet outil puissant, ce client n'a pu produire, en 4 ou 5 ans, qu'une très petite quantité de films (qu'il faut acheter en plus et faire développer auprès du fabricant de la machine). La raison: dans les boîtes d'archives, on trouve bien autre chose que des feuilles A4 uniformes dans leur format, d'un grammage constant, d'une lisibilité toujours égale et jamais attachées par une ferraille quelconque! Cela veut dire qu'au lieu d'utiliser l'avaleur de la machine pour suivre un rythme de production supposé par les inventeurs et vendeurs de cette machine, le responsable spécialisé local devait passer son temps à préparer des piles de documents, en rephotocopiant une à une des feuilles hors formats après avoir désagrafé des liasses (avec tous les risques de déclasserment possibles)... Rendement ridicule! Et, qui plus est, le résultat escompté sur microfilms est tellement médiocre (problèmes de contrastes différents pour chaque document) que l'étape suivante, espérée mais jamais testée, qui était de récupérer l'image du microfilm pour en faire une image électronique, est pratiquement irrécupérable pour cet objectif.

Sans commentaire!

1.2 Choix des supports ►

1.3 Procédures ►

1.4 Le microfilmage ►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►





La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives – Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples ►

1.2 Choix des supports

– Commençons par le support le plus traditionnel: le papier ! Depuis qu'il y a des fax et des imprimantes, on produit plus de papier que jamais. Qu'en restera-t-il demain?

Tout le monde connaît le phénomène d'effacement pratiquement complet au bout de 5 à 10 ans, suivant le mode de conservation, des papiers thermiques utilisés par la première génération de fax et de très petites imprimantes d'ordinateur. Archiver ce genre de support ne sert à rien. Il faut éliminer ou photo-copier ou -graphier ².

Le problème de la photocopie, comme des imprimantes à rayon laser (qui deviennent les plus répandues depuis quelques années), c'est que l'encre en poudre fixée sur le papier n'est pas réellement "imprimée" et comme accrochée aux fibres du papier. Elle est seulement "collée" avec plus ou moins de pression, et plus ou moins de qualité d'adhésion de la poudre, sur le support papier. Aussi, il peut se produire que des documents imprimés de cette façon soient retrouvés, dix ans après, complètement vides, toute la matière imprimée étant à nouveau réduite en poudre. Et, quand on le voit, il est trop tard généralement!

Impossible de vérifier chaque année toutes les feuilles stockées qui auraient pu l'être avec une technologie insuffisante!! Mais il y aura des surprises. Si un document est vraiment précieux et important, peut-être faut-il en faire une photographie avant de l'archiver aujourd'hui?

– Une publicité (mensongère?) de Philipps vente des DVDs sur lesquels on peut enregistrer des données numériques "pour l'éternité"! On vend aussi, à un prix nettement plus cher, des DVDs à "longue durée de vie".

Plusieurs niveaux d'erreur sont possibles dans le choix et l'utilisation des DVDs.

D'abord, il s'est avéré à l'usage d'une vie encore relativement courte (une dizaine d'années) que les CD-ROMs comme les DVDs peuvent s'abîmer, voire devenir illisible à assez courte échéance (après 5 ou 6 années d'utilisation selon l'environnement d'utilisation) ou peuvent être griffés s'ils ne sont manipulés avec une certaine prudence.

D'autre part, la guerre des "standards" entre les grands producteurs n'est pas terminée. Et l'on se trouve, parallèlement, avec une offre variée d'autres supports de stockage performants quant aux volumes et aux vitesses de traitement.

Quel choix faire pour un stockage de moyenne et longue durée?

Expérience faite, il n'y pas de choix exclusif et parfaitement fiable pour le stockage de l'écriture électronique. La copie régulière et la migration intelligente de données constituent la seule sécurité véritable.

Croire, comme pourraient le faire croire la publicité des constructeurs, que mettre des données sur CD ou DVD est une garantie de longévité, est une erreur. Mais on ne peut se passer de tels outils. D'où la tentation de certains de faire de l'archivage électronique en ligne, comme certains opérateurs le proposent. Là aussi ce peut être une lourde source d'erreur si l'on ne contrôle pas totalement (hard et soft) les serveurs sur les mémoires desquels on effectue un tel stockage. Sinon: où se trouvent réellement ces données sur la

planète? qui en permet l'accès?

– Un autre exemple où l'évolution technologique a surpris certains fabricants de stockage de données électroniques: la clef USB !

À cause des producteurs de supports musicaux, les clefs USB qui ont fait leur apparition commerciale il y a 4 ou 5 ans, et puis une percée très rapide, avec des capacités toujours plus importantes et une facilité d'utilisation "à chaud" sur n'importe quel support informatique doté d'une connexion USB.

Les disquettes, dont certains formats sont devenus définitivement illisibles (plus de lecteurs), vont rapidement devenir obsolètes et disparaître. Et il y a une logique à cette disparition dans un environnement cohérent tout électronique. En effet, tout ce qui est "disque" (y compris les disques durs qui restent performants actuellement – mais jusqu'à quand?), comporte une partie mécanique qui fait tourner le disque. D'où: frottements, possibilités d'usure du mécanisme de rotation, échauffements, etc. La logique du tout-électronique va donc vers le support de masse purement électronique (mémoire flash) n'impliquant pratiquement plus aucun mouvement mécanique.

Mais peut-on se fier aux clefs USB? Oui pour des stockages et transferts de courte durée. Non si l'on pense au moyen et long terme (ou en tout cas pas plus que jadis pour une disquette).

Demain ce sera probablement et tout simplement le téléphone portable qui servira de support de stockage de transition, avec, en plus la possibilité de transférer téléphoniquement les données stockées vers une mémoire de grosse capacité à travers un ordinateur de télécommunications.

1.3 Procédures ►

1.4 Le microfilm ►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►

2. L'archiviste d'une grande Congrégation religieuse à Rome s'étant rendu compte de ce problème, est actuellement en train de rouvrir tous les dossiers de ces époques pour re-photocopier des pièces qui lui sembleraient en danger. Mais en se demandant s'il ne fait pas un travail qui n'aura pas une durée de vie beaucoup plus longue que celle des documents qui sont en danger d'effacement!



Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives – Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples ►

1.2 Choix des supports ►

1.3 Procédures

– Pour faire le lien avec la question des supports, voici un exemple d'erreur d'archivistique majeure que nous avons pu observer chez un client;

Ce client avait un fonds d'archives précieux: autographes abondants d'un fondateur de Congrégation religieuse. Sur le conseil d'un "informaticien de la maison" (un ami qui vendait des ordinateurs et des logiciels), ils ont décidé de scanner tous ces documents. Et ils ont payé une personne durant quatre années pour scanner ces documents. Cette personne a consciencieusement fait un travail "bête" (aux deux sens du mot): plus de 5.000 images de documents, selon des formats et des agrandissement différents (pourquoi? aucun critère), selon des luminosités et contrastes différents, et, surtout, sans aucune référence à l'emplacement de conservation du document original dans une archive qui n'était pas répertoriée.

Tout cela était soigneusement stocké sur divers CD-ROMs simplement numérotés au marqueur de 1 à x.

Nous avons fait quelques essais pour sauver ce travail. Mais le seul temps nécessaire pour identifier les fragments scannés coûtait beaucoup plus cher que la création d'un inventaire avec collation de tous les originaux ainsi qu'une photographie numérique à neuf de tous ces documents.

Le client n'a pas aimé qu'on lui dise qu'il avait payé un travail inutilisable, même s'il a vu de ses yeux les exemples et les tests que nous avons faits...!

– La préservation de textes passe souvent aujourd'hui – avec toutes les réserves faites plus haut sur les supports de stockage – par la numérisation des textes qui permet, en outre, si le travail de saisie a été fait intelligemment de manipuler ultérieurement ces textes et de s'en servir pour de multiples usages.

Quand on parle de numérisation de textes, le réflexe de la majorité des personnes est: alors: vous "scannez"!

Mais on a de multiples preuves de la stupidité qu'il y a à se fier au scannage de textes, suivi, tout de go, d'un traitement par OCR (Optical Character Recognition): les résultats sont lamentables et ridicules!

Un exemple qui vaut également pour les anglo-saxons: faites une recherche sur les "belges" dans ce moteur de recherche sur les ouvrages "scannés" et transcrit au vol: vous avez bien quelques "belges" quand la typographie distingue parfaitement le "l" et le "i"; mais vous serez surpris de devoir chercher les "belges" sous le vocable les "beiges": y auriez-vous pensé?

Ne parlons pas des programmes de traduction automatique : ils vous permettent tout juste, si vous ne connaissez vraiment pas une langue, surtout avec une écriture que vous n'avez pas appris à lire, de vous rendre compte du sujet traité dans un article ou un livre!

Le même type de problème se retrouve dans la saisie vocale . Pour la constitution

d'une très grosse Base de données textuelle à partir d'originaux manuscrits en français du 19e siècle, un des "bénévoles" sollicité par le coordinateur du projet, avant que nous ne le reprenions en charge, avait imaginé de dicter sa lecture à l'ordinateur avec l'un des 2 ou 3 programmes de reconnaissance vocale qui existent sur le marché (et qui peuvent, bien utilisés par des personnes à visibilité limitée, être pour elles d'un grand secours). Il était canadien. L'accent joue, mais également la constance du niveau vocal, l'environnement sonore, etc.

Voici l'une ou l'autre perle trouvées dans les sections enregistrées de cette façon, sections qu'il a été moins coûteux de réenregistrer complètement car le nombre de fautes stupides et aléatoires rendait la correction beaucoup trop saccadées et donc trop onéreuse:

"Yi Zhu vis la face de Dieu panda un instar" comprenez: "Jésus vit la face de Dieu pendant un instant"!

"Tout l'ai jour on ses machines..." comprenez: "Tous les jours, on s' imagine..."

"Hystérectomie" dans un texte liturgique! Comprenez (du latin) "Iste Rex gloriae"!

Cela veut dire que de tels résultats, non seulement ne préservent rien, mais ils détruisent les données originelles et empêchent toute recherche sérieuse.

Nous avons fait beaucoup de tests en ce domaine: rapidité, précision, enrichissements minimum pour une recherche qui ne comporte pas trop ni de "bruit", ni de "silence", sur tous les mots d'un texte et sans "black list". Tout dépend de la perfection que l'on cherche. Par exemple, s'il s'agit d'enregistrer un important dictionnaire encyclopédique qui fait autorité dans son domaine et dans lequel tous les mots, les dates, les chiffres, les références ont été créés par des personnes compétentes dans leur domaine, modifier même une virgule pourrait mener à un contre-sens et modifier une date ou une valeur peut casser la fiabilité de tout le dictionnaire. Il y a donc lieu d'être rigoureux dans la saisie. Seule la double saisie (deux personnes différentes saisissent parallèlement le même texte; le résultat est ensuite comparé automatiquement et souvent confronté à une saisie par scanning, si le texte n'est pas un manuscrit), permet un taux d'erreur inférieur à 0,0009%. Mais cela coûte cher et l'on ne trouve plus guère ce genre de garanties que dans les saisies faites en Asie (Indes, Philippines, Taiwan, etc.), et de préférence par des personnes totalement ignorantes de la langue qu'elles ont à enregistrer. Même dans ce cas, il y a lieu d'encore assurer une relecture complète!

Aujourd'hui, seule la transformation des données saisies en format XML (Extended Markup Language) permet une conservation et une transmission à peu près indépendante des programmes et des machines, car ce format permet des migrations intelligentes dans différents environnements. Encore faut-il avoir le courage d'utiliser toutes les potentialités de l'XML, en développant des DTD conformes aux besoins spécifiques de l'application et en ne se limitant pas à la paresseuse utilisation de Schémas d'XML qui ne sont que des DTD prémâchées par des états-unis trop fainéants pour analyser correctement un problème ou trop avides de gains à court terme.

Encore un mot dans le domaine de la saisie des données: je viens de constater, dans plusieurs professions libérales dont les secrétariats utilisent maintenant abondamment l'ordinateur pour créer des textes (gestion sociale, études de notaires, médecins, architectes) que l'on prend soin de passer les textes au correcteur d'orthographe, souvent rapidement, pour refaire un acte ou un contrat dont on a modifié sur place certains éléments. Mais vous savez que ce peut être une source réelle d'erreurs, car le correcteur d'orthographe ne regarde que la forme, il ne se soucie ni des homographies, ni des accords entre les mots (masculin, féminin, singulier, pluriel, accord des participes, etc.). Si l'on n'a pas la possibilité de relire les textes passés au correcteur d'orthographe, il vaut parfois mieux ne pas l'appliquer afin d'éviter le risque de transformations de données erronées, ou d'oubli de fautes énormes liées à des cas d'accord.

– Une autre source d'erreur possible, se trouve dans l'attribution de noms aux fichiers que l'on crée: quelle systématique utilise-t-on dans leur création? quelle constance? peut-on se fier pour retrouver des fichiers à un Google-Search, par ailleurs plein de "spyware"? faut-il renommer tous les fichiers qui vont être archivés? Si oui, selon quels principes?

La façon de nommer les fichiers peut générer un chaos catastrophique le jour où il s'agit de

faire une migration intelligente et sélective ou encore le jour où, dans un cadre professionnel, c'est une (ou un) remplaçant qui doit prendre la relève du titulaire de tel ou tel poste de saisie de données. Où se trouvent les fichiers? sous quels noms et sous quels répertoires va-t-on les retrouver?

– UNICODE : Au-delà de l'ASCII (code à 8 bits), nous sommes en train de passer à l'UNICODE (code à 16 ou 32 bits) comme code de base de tout traitement informatisé³. Ce qui explique les énormes augmentations des capacités de mémoire nécessaires à tous les niveaux (également justifiées par le traitement massif des images électroniques). Mais qui a été voir de près les effets de l'utilisation d'UNICODE dans la transmission de données électroniques aujourd'hui?

Un caractère UNICODE peut, en effet, se présenter de plusieurs façons différentes dans un fichier électronique.

Comment les textes que j'ai enregistré sont-ils codés? et ceux que je reçois? n'y a-t-il pas un risque que j'aie, dans le même fichier plusieurs représentations différentes d'un même caractère? et, dans ce cas, toutes ces représentations vont-elles toujours être interprétées correctement par tous les programmes qui vont utiliser ces données? Les tris vont-ils produire sur tous les éléments les mêmes résultats? Quelles conséquences pour l'utilisation de différentes polices de caractères? sont-elles toutes "Unicode-compatible"? Ne faut-il pas ajouter un contrôle (et un long travail de correction, le cas échéant) sur tous les fichiers électroniques qui doivent être stockés pour une longue durée, indépendante de l'environnement de leur producteur? Quel risque de perdre de l'information est-il attaché à ce type d'accident de codage?

1.4 Le microfilmage ►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►

3. Voir: *The Unicode Standard 4.0* by The Unicode Consortium, Addison-Wesley, 1991-2003, XXXVIII + 1462pp.





La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives - Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples ►

1.2 Choix des supports ►

1.3 Procédures ►

1.4 Le microfilmage ⁴

Malgré le progrès massif de l'électronique et du numérique partout, les responsables sérieux de la préservation sur le long terme, finissent toujours par se mettre d'accord que la seule préservation qui est aujourd'hui assurée pour une durée de 200 à 400 ans, est celle qui se fait sur un support de film argentique traité aux polysulfides ⁵.

Face à cette conviction, voici que les producteurs d'appareils de photographie numérique ont mis sur le marché des produits de plus en plus performants pour la saisie professionnelle de données, et, depuis la Photokina de septembre 2002 à Cologne, la possibilité de piloter la photographie et ses paramétrages par ordinateur en stockant directement les images sur un disque dur. La qualité est excellente et l'image en couleur a un rendu parfois meilleur que les documents originaux eux-mêmes.

Mais peut-on se lancer dans un archivage fondé sur ce type de technique sans avoir testé l'ensemble de la chaîne de traitement? En effet, il n'est pas question, aujourd'hui, pour des questions de coûts, d'imaginer la fixation de données en couleur sur des microfilms pour une préservation de longue durée. Et, la prise de vue numérique ne donne jamais de vrais images en noir et blanc. Il faut donc tester si l'image couleur de qualité, transformée en nuances de gris, est suffisamment contrastée pour permettre une bonne reproduction sur microfilm noir/blanc. Le test final est alors celui de la réversibilité des données: est-il possible, à partir d'images numériques en couleur, transformées en nuances de gris et fixées sur un film argentique noir/blanc, de recréer une image électronique lisible des mêmes données?

Si l'on n'a pas eu l'opportunité ou la faculté technique d'effectuer l'ensemble de cette chaîne de traitement, on court des risques.

Pour un important projet de sauvetage d'archives, nous avons choisi cette filière nouvelle. Nous en avons testé toute la chaîne. Le choix a-t-il été le bon? sera-t-il confirmé par l'évolution technologique? Nous pensons qu'il répond, actuellement aux différents critères pour la préservation et l'accès à des fonds d'archives dans un environnement qui devient de plus en plus "tout informatique". Mais l'erreur dans l'évaluation de l'évolution technologique reste possible.

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques? ►

4. Voir R.-F. Poswick et Y. Juste, Le microfilm à l'heure de la photo numérique, dans *Interface*, e-96, sept. 2004.

5. Les recommandations de *The Micrographics Quarterly*, Vol.1/2, Denver, Colorado,

2000, résumé de façon simple les directives de l'État du Colorado (USA) sur cette question: "Although microfilm is an old format, it is the more reliable format for archiving both for longevity and accessibility. When properly stored, microfilm can be maintained for a very long time, much longer than any electronic format. A concern with new technology is that it can be replaced as quickly as it is introduced. Many agencies are gifted with old Wang floppy disks, 5.1/4 inch and 8 inch disks, tape drive units and other marvels of the computer age that were "the thing" when new but are virtually unusable now. Microfilm is permanent, eyereadable and not dependent on ever-changing software or technology for access."





La foi béate mène à la faute:
l'évolution technologique comme source d'erreur,

7e Journée des Archives - Louvain-la-Neuve 20-21 Mai 2007

1. Sources d'erreur dans la gestion électronique de textes et de documents ►

1.1 Choix de matériels: deux exemples ►

1.2 Choix des supports ►

1.3 Procédures ►

1.4 Le microfilmage ►

2. La gestion informatisée d'archives apporte-t-elle une valeur ajoutée propre et quels sont ses risques?

a) Nous commençons par les risques en rappelant la réponse pleine de bon sens et d'expérience que donneront tous ceux qui travaillent consciencieusement dans ce domaine avec l'aide de l'informatique: "Le problème avec les nouvelles technologies c'est qu'elles peuvent être remplacées aussi vite qu'elles ont été mises en œuvre... [différentes machines et différents supports] qui étaient considérés comme 'la chose' [à utiliser] au moment où ils étaient tout neufs, sont devenus virtuellement inutilisables aujourd'hui" ⁶ .

Tout est dit. Notre expérience peut le confirmer. Tant que l'écriture électronique n'aura pas trouvé la stabilité (et la rigidité) qu'avaient acquises les écritures alphabétiques avec l'imprimerie (donc près de 3000 ans après le passage du pictogramme vers cette écriture alphabétique qui est la matrice de notre civilisation et donc de nos problèmes d'archives), il faudra rester vigilant à utiliser, parallèlement au stockage électronique (quand il est possible), toutes les ressources pour la préservation physique des originaux quand ils sont précieux et pour la création et le stockage scientifique de copies sur microfilms argentiques.

Et, en toute occurrence, aucune crédulité n'est permise. Comme nous l'avons dit en exergue, la foi béate, surtout dans les progrès technologiques annoncés ou réels, est souvent une source d'erreur. Quand on entre dans l'informatisation, de quelque nature qu'elle soit, ce serait une grosse erreur de croire que quelque chose est réalisé "pour l'éternité"! D'important frais de maintenance sont à prévoir dès le début d'un projet quelconque d'informatisation, avec la présence (et les coûts économiques liés) de vrais informaticiens. On ne peut se contenter de mettre des supports magnétiques ou électroniques dans une armoire ignifuge ou un coffre de banque. Cela appelle une restructuration de l'équipe archivistique, en y incluant un informaticien conscient des problèmes de gestion dans la durée (migrations de données) et apte à une veille technologique prévoyante.

b) Quant à la valeur ajoutée propre à une gestion informatisée de fonds d'archives, elle doit se lire sur la toile de fonds globale des risques liés à un système d'écriture/lecture pas encore stabilisé. Aussi les quelques traits que nous retiendrons pourront paraître des évidences. Ils vont de soi. Ce n'est pas pour cela qu'il ne faut pas les redire, après tant d'autres, mais avec cette pointe de scepticisme absolument nécessaire pour ne pas qu'on absolutise les aspects positifs de l'électronique appliquée aux textes, documents et archives au détriment de beaucoup de méthodes traditionnelles qui seront encore, si pas toujours, nécessaires!

1 – La première valeur ajoutée nous semble être de l'ordre de la méthodologie: l'électronique oblige tous les secteurs qu'elle touche à remettre en question les habitudes de travail et les principes acquis et qui n'ont plus été vraiment mis en question, au moins

depuis Gutenberg. Le transfert d'informations traditionnellement gérées en fonction de l'écriture alphabétique, de l'imprimé et de l'accumulation physique des traces de mémoire, vers des traces volatiles de mémoire, liées à la présence d'un courant électrique ou à une magnétisation (en attendant les mémoires électroniques sur support biologique), oblige à repenser l'ensemble des opérations de l'archivistique, voire même certains aspects de ses objectifs. Nous pensons, notamment à la question de la sélection de l'information utile à conserver: la masse potentielle d'information conservée sous forme électronique est en croissance exponentielle. Que faut-il garder ? Et, par rapport aux archives sur d'autres supports antérieurs, que faut-il faire passer dans les mémoires électroniques ?

Cette question de la mémoire utile appelle une décision intellectuelle (philosophique?), stratégique et politique: pour qui conserve-t-on? pour quel usage?

Mais également, quelle est la façon utile de conserver électroniquement ?

Ou encore: l'archiviste ne doit-il pas redéfinir son rôle ou le redéployer en plusieurs fonctions nouvelles et différentes pour rendre les mêmes services ou de nouveaux services à travers des outils en écriture électronique ?

2 – Une seconde valeur ajoutée est évidemment le caractère universel ou planétaire de la mémoire électronique : une fois constituée, elle devient accessible au plus grand nombre. C'est une très grande révolution pour un domaine (les archives) où les accès étaient réputés difficiles et réservés à quelques rats d'archives n'ayant pas peur de la poussière.

Mais "monsieur tout-le-monde" est-il habilité à accéder même à l'image virtuelle de documents d'archive, sans être à même d'évaluer tout le contexte de la conservation d'un tel document ?

3 – Une troisième valeur ajoutée est l'accès à la seule forme virtuelle du document. Ceci permet d'éviter des manipulations directes, des éventuelles destructions ou disparitions de documents, etc. (avec le risque qu'on se désintéresse peu à peu de la conservation physique des vrais documents, préoccupé que l'on sera de plus en plus de la préservation électronique... mais ceci fait encore partie des risques liés à une valeur ajoutée certaine!)

4 – Une quatrième valeur ajoutée est la possibilité graduelle de mettre des données en écriture électronique en communication les unes avec les autres, soit grâce des programmes de recherche multicritères, voire multibases, soit grâce à des hyperliens pertinents. Cette possibilité enrichira l'utilisation de documents d'archives bruts et suppléera, en tout ou en partie, aux inconvénients d'une distribution universelle des données d'archive.

5 – Une cinquième valeur ajoutée est la possibilité, à mesure d'une stabilisation de l'écriture électronique et/ou de ses méthodes de conservation à moyen et long terme, de stocker d'énormes volumes d'information dans des espaces virtuels nanométriques. Cela supposera, bien sûr, un nouveau tri dans les archives traditionnelles pour savoir quels sont les documents du passé qu'il faut absolument conserver physiquement (et non seulement de façon virtuelle) parce qu'ils ont une valeur de "relique" dans un domaine spécifique. Cela voudra dire aussi une modification sensible des surfaces de stockage correspondant à des archives traditionnelles ⁷.

Peut-on et faut-il aller plus loin ?

Ces perspectives sont déjà vertigineuses, d'autant qu'il faut les imaginer en synergie par rapport à un être humain planétairement muni du stylo de l'écriture électronique qu'est en train de devenir le téléphone cellulaire portable: n'importe qui, de n'importe où, pourrait, théoriquement, interroger une pièce d'archive ou demander de mettre en archive un élément qu'il voudrait conserver afin de pouvoir le communiquer à d'autres ultérieurement.

Felix culpa ?

Bienheureuses erreurs occasionnées originellement par l'évolution technologique ?

À chacun d'en juger !

6. *The Micrographics Quarterly*, Vol. 1/2, op. cit., ibid.

7. À ce jour nous n'avons pas eu accès à des études qui auraient fait sérieusement le

calcul comparatif des coût récurrents de stockage physique dans la durée, par rapport aux coûts de saisie et de maintenance électronique du même ensemble de données. Il nous semble qu'au prix de m², la question pourrait être pertinente au niveau d'Archives importantes. Nous pensons, sans que cela ait été dit de cette façon, que la création du centre d'archives de St-Gobain, en France, procède d'un calcul de ce type qui va encore plus loin, puisqu'il va jusqu'à valoriser économiquement la gestion électronique d'archives (voir: Didier Bondue, *Les Archives de Saint-Gobain, un patrimoine valorisé et une plus-value pour l'entreprise* , 6ième Journée des Archives, Louvain-la-Neuve, 21 avril 2005).

