



Figure 1. Les idées reçues sont parfois véhiculées de façon surprenante (photo : © Jean-Marc Fontaine).

enjeux et ont mis en place des équipes et des moyens pour garantir la pérennité de leurs documents numériques. Un exemple frappant est celui de l'industrie aéronautique. La documentation, entièrement numérique, d'un avion de ligne contient la description de toutes les pièces constituant cet avion. La durée de vie de l'avion et les contraintes juridiques impliquent une conservation de cette documentation pendant plus de cinquante ans. En 2008, Airbus et Boeing, deux féroces concurrents, ont signé, par l'intermédiaire des associations aérospatiales américaines et européennes, un mémorandum sur l'archivage à long terme des données qu'ils produisent. Ce projet a abouti à la création du projet LOTAR (long term archiving and preservation of data) [3] qui vise à apporter des solutions aux problèmes de conservation des documents produits par ces industries.

Dans le domaine patrimonial, si l'Institut national de l'audiovisuel (Ina) n'avait pas mis en place son projet « plan de sauvegarde et d'archivage », 835 000 heures d'archives auraient disparu en 2015, soit près d'un tiers des archives audiovisuelles conservées [4]. De son côté, la Bibliothèque nationale de France (BnF) a mis en place en 2010 son très ambitieux système de préservation numérique réparti (SPAR) [5]. Ce système constitue aujourd'hui une véritable base de données en tous points conforme aux normes essentielles et en particulier à la norme ISO 14721 « Modèle de référence pour un système ouvert d'archivage d'information » [6, 7]. Enfin, le Service interministériel des archives de France (SIAF) [8] mène une politique active visant à rationaliser l'activité de préservation numérique dans les différents ministères.

De leur côté, les établissements hospitaliers mènent de nombreuses opérations d'envergure :
 € la transition vers un dossier médical entièrement numérique. L'AP-HP (Assistance publique - Hôpitaux de Paris), l'imagerie médicale presque totalement dématérialisée fin 2010. Une dématérialisation totale, incluant résultats d'analyses, comptes-rendus des consultations etc., est planifiée pour 2015 ;
 € une réorganisation du dispositif informatique impliquant une forte

réduction du nombre des applications logicielles utilisées ;
 € la mise en place de procédures de sauvegarde sécurisées avec un stockage des données sur plusieurs sites géographiquement distants en attendant la conduite d'une réflexion de fond sur la problématique « long terme ». À cet égard, les établissements de santé rencontrent une diversité et une complexité de documents tout aussi grande que celle des autres grands organismes. Cette problématique se pose dans les mêmes termes pour le dossier médical personnel (DMP) [9] ou pour les agences sanitaires [10].

Un partage des approches et des retours d'expérience en matière de préservation à long terme des documents numériques est organisé depuis plus de dix ans au sein du groupe PIN (pérennisation des informations numériques) [11]. La conservation de l'information numérique avait fait l'objet d'une analyse dans médecine/sciences en 2008 [12] et d'un développement beaucoup plus complet sous la forme d'un ouvrage destiné aux acteurs principaux de cette conservation en 2009 [13].

Le monde de la recherche scientifique cherche sa voie

Le Centre national d'études spatiales (CNES) et le Centre national de la recherche scientifique (CNRS) ont signé ensemble des protocoles d'accord visant à créer des centres de données en charge de la pérennisation des observations issues des expériences scientifiques réalisées. Le Centre de données de la physique des plasmas (CDPP) [14] en est un exemple.

Le Centre informatique de l'enseignement supérieur (CINES) a été vu confier une mission nationale d'archivage de documents numériques du patrimoine scientifique [15]. La mission du CINES pour l'archivage des thèses numériques a fait l'objet d'un arrêté officiel le 7 août 2006. Le CINES archive un certain nombre de publications du portail HAL (hyper article en ligne) des archives ouvertes [16, 20], et médecine/sciences avait consacré un article [21].

Le Service des initiatives du TGE (très grand équipement) du CNRS [17] vise à assurer la préservation des données numériques issues de la recherche dans les sciences humaines et sociales en se basant sur une infrastructure d'archivage mutualisée entre le CINES et le Centre de calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3). Ces initiatives pilotes ne couvrent qu'une partie réduite de la production scientifique française. L'archivage

¹ Adonis : accès unifié aux données et documents numériques des sciences humaines et sociales. <http://www.tge-adonis.fr>

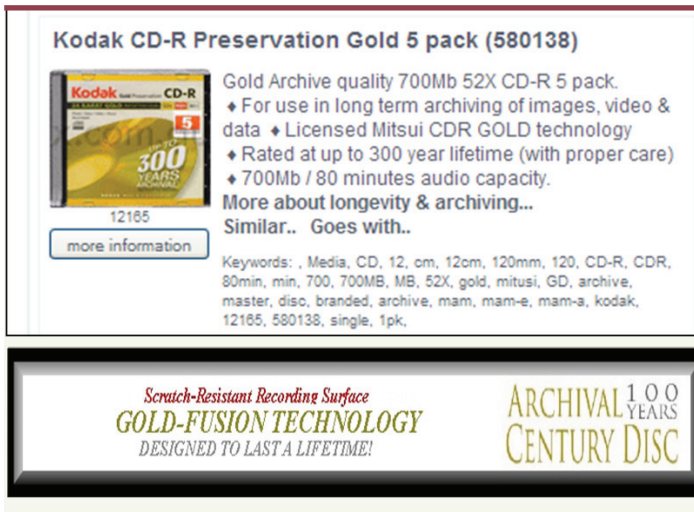


Figure 2. Quelques exemples de publicités trouvées sur internet. Elles annoncent avec sérieux des durées de vie des CD de 100 et 300 ans !

Le risque de tout perdre et quelques solutions pragmatiques pour l'éviter

Les dangers de tout perdre sont multiples et partiellement méconnus. Le vol de l'ordinateur et du disque dur de sauvegarde posé juste à côté, le sinistre, l'infection par un virus destructeur ne sont pas si rares. Mais, avec quelques précautions, il est possible de réduire les risques. Une situation plus insidieuse conduira à l'illisibilité de certains documents du fait de l'obsolescence des formats de fichiers ou de la dégradation des supports d'archivage les renfermant. Conscient de ces risques, chacun sera donc peu à peu conduit à appliquer d'une manière ou d'une autre un ensemble des recommandations présentées ci-après. Bien que fort limitées et peu détaillées, les actions proposées ci-après sont de nature à apporter une relative sécurité des documents numériques à conserver au cours du temps.

numérique de la production des chercheurs se heurte à la multiplication des unités et des équipes de recherche et à la grande autonomie des chercheurs. L'Institut national de la santé et de la recherche médicale (Inserm) a entrepris en 2009 une réflexion sur la mise en place d'un système d'archivage numérique. La situation n'a cependant guère avancé depuis.

On veut préserver ces premières actions à mener. Elle n'exige aucune connaissance particulière. La conservation de ce que nous appelons le patrimoine numérique ne pourra être convenablement assurée que si nous avons précisément identifié les documents à préserver.

Les besoins des structures les plus petites et des individus isolés

Les structures de taille réduite et les individus isolés ne disposent ni des moyens ni des compétences pour appréhender le problème de la même façon que les grands organismes. Pourtant, chacun, au sein de son équipe ou de son activité professionnelle individuelle, a besoin de conserver un certain nombre de rapports et mémoires, études, publications, etc. À ces s'ajoutent ceux qui relèvent d'obligations légales. Pour les cabinets libéraux, ce sera par exemple le dossier de suivi des patients, la comptabilité, les documents fiscaux, etc. Dans certains cas, ces petites structures peuvent s'adosser aux systèmes de plus grands organismes comme la BnF ou le CNRS qui jouent alors le rôle de tiers archiveur. Cependant, elles sont nombreuses à ne pas avoir cette possibilité. Comment assurer la conservation des documents avec un niveau de sécurité et sans dépenses significatives ?

ce qui aura une valeur ou une utilité dans le futur. La préservation à long terme ne concerne donc ni les multiples documents d'usage éphémère ni les versions intermédiaires de travail.

Les particuliers, dans leur vie privée, se posent la même question. Ce sont les photographies numériques qui ont déclenché la création d'un patrimoine numérique personnel. À cela s'ajoutent les courriers importants, les documents issus de nos passions, généalogie, musique, etc., c'est-à-dire tout ce qui constitue un patrimoine affectif ou artistique auquel s'ajoutent les documents numériques utiles pour faire valoir nos droits : factures, relevés de Sécurité sociale, déclarations fiscales.

Il faut nommer soigneusement les documents à être préservés. Le disque dur de l'ordinateur ressemble à une gigantesque bibliothèque dans laquelle nous pourrions ranger des centaines de milliers de documents. Si nous ne prenons pas la peine de regrouper au sein d'une organisation structurée de dossiers tout ce qui nous paraît important, nous n'aurons toutes les chances de retrouver dans le futur. La dispersion des documents sur le disque dur est en partie provoquée par les logiciels qui suggèrent d'enregistrer les documents dans les dossiers les plus variés. Cette dispersion peut aussi être aggravée par la négligence de l'utilisateur. Une fois les documents regroupés, il convient de leur constituer des noms explicites susceptibles de traverser le temps. Ce patrimoine numérique qui s'en tient sur des supports de stockage ne bénéficie pas des technologies réservées aux grands sites informatiques, nous avons le choix entre trois catégo-

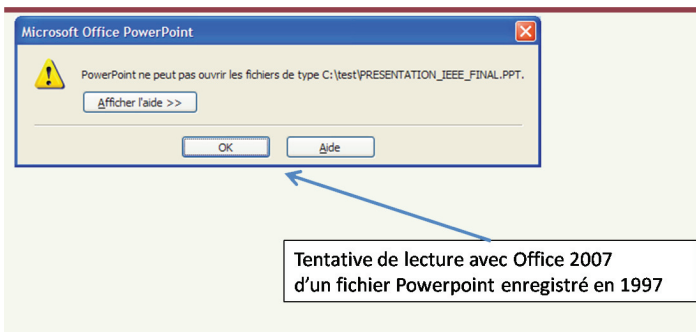


Figure 3. Les logiciels de la suite Microsoft Office 2007 ou 2010 ne savent plus reconnaître les fichiers créés 10 ou 12 ans plus tôt par Word ou PowerPoint de Microsoft !

ries : le disque dur, les disques optiques (CD, DVD, Blu-ray). On pourra par exemple recopier la totalité des données sur un espace tampon. Si un des supports est défaillant, il devra alors être remplacé par un support valide. La durée de vie des disques durs ne dépasse généralement pas cinq ans en moyenne et le disque peut cesser de fonctionner de façon prévisible. Le dysfonctionnement du disque dur constitue la première cause de panne des ordinateurs. Pour les disques optiques, les informations les plus contradictoires circulent. Ces informations sont polluées par le discours marketing de certains auteurs (Figure 2).

Nous devons nous en tenir aux travaux du Laboratoire national de métrologie (LNE), laboratoire public, neutre et indépendant de tout intérêt marchand. Le LNE a montré, sur la base de ses statistiques représentatives sur des CD et des DVD de différents sites, que la durée de vie de ces disques est réduite à deux, trois ou cinq ans. Tous les disques ne sont pas équivalents mais il n'est pas facile de connaître ceux qui sont plutôt fiables et ceux qui ne le sont pas. Ces analyses sont disponibles en ligne sur le site du pôle de recherche sur la conservation des données sur disques optiques numériques [18].

À propos des services de stockage en ligne

Ces services sont appelés à jouer un rôle important dans l'usage de moyens de stockage sécurisés. Cependant, la situation n'est pas encore claire : certains services de stockage gratuits sont à exclure totalement. Les autres se réservent le droit d'utiliser nos données. Les services de stockage payants sont à utiliser avec beaucoup de prudence. Lorsqu'on ne connaît pas la localisation du site de stockage, on peut s'interroger sur les garanties de confidentialité. En fonction des services, on observe aussi des écarts de pratiques.

Choisir les supports adaptés, disposer de plusieurs copies et les surveiller. La seule réponse crédible face à ces constats consiste à réaliser plusieurs copies des mêmes documents et de surveiller régulièrement la surveillance et le renouvellement. Le choix des types de support dépend du volume de données à stocker. Pour stocker 500 Go de données et de documents, il sera préférable de choisir un disque dur plutôt qu'une centaine de DVD et a fortiori un millier de CD. Un nombre réduit de supports de stockage facilitera leur surveillance. En l'absence d'information fiable sur le vieillissement des supports, cette surveillance consiste à effectuer, au moins annuellement, une relecture complète de l'ensemble des données. Cette méthode consiste à gérer un tableau de bord des supports et des opérations réalisées comme le Tableau I.

Les documents sans logiciel de lecture

Pour de multiples raisons, un éditeur de logiciel peut disparaître du jour au lendemain. Les logiciels qu'il développe ne seront alors plus mis à jour ni adaptés aux nouveaux systèmes d'exploitation. Tout aussi préjudiciable : certains logiciels mondialement utilisés comme la suite Microsoft Office, n'assurent que la compatibilité ascendante permettant d'ouvrir les anciens fichiers avec les nouvelles versions des logiciels (Figure 3).

Pour les documents très courants, nous pourrions avoir une relative confiance dans les formats PNG, JPEG, JPEG2000 pour les images et dans les versions normalisées de PDF : PDF/A (ISO 19005-1) et PDF 1.7 (ISO 32000-1). Pour les documents spécifiques à chaque discipline, on examinera la situation au cas par cas avec plusieurs principes : exclure les formats propriétaires, privilégier les formats ouverts et minimiser le nombre de formats différents utilisés.

² CD : compact disc ; DVD : digital versatile disc ; « le disque Blu-ray ou Blu-ray disc (abréviation officielle BD, autre dénomination B-RD) est un format de disque numérique breveté et commercialisé par l'industriel japonais Sony permettant de stocker et restituer des vidéogrammes en haute définition. Sa dénomination provient du type de rayon laser qu'il exploite, de couleur spectrale proche du bleu » (Wikipédia).

Date	Action	Version de référence	Copie 1	Copie 2	Copie 3
21-10-2007	Création version initiale	Disque dur 500Go de Western Digital Création et relecture OK 12 143 fichiers 211 Go			
22-10-2007	COPIE		Disque dur lomega Prestige Desk-Intenso memory top hard drive 500 Go	Disque dur Intenso memory house 320 Go	Disque dur Intenso memory house 320 Go
			OK	OK	OK
15-04-2008	Vérification : relecture totale	OK	OK	OK	OK
12-07-2008	Mise à jour version de référence avec nouveaux documents	13 665 fichiers 243 Go			
13-07-2008	Mise à jour copies		OK	OK	OK
09-12-2008	Vérification	OK	OK	OK	OK
	Etc.				
22-10-2010	Vérification	OK	OK	Disque illisible - remplacement du disque dur Western Digital Elements (1 To)	OK
	Etc.				

Tableau I. Un exemple de ce que peut être un tableau de bord des supports de stockage.

Au-delà de cette réflexion

Conclusions

La société contemporaine a vu progressivement se développer des objets éphémères. Rien de ce qui se fabrique aujourd'hui est construit pour durer. Qu'il s'agisse de l'ordinateur, du téléphone ou du lave-linge, chaque appareil est conçu pour avoir une vie limitée. Ce processus cyclique au cours duquel chaque objet utilise, jette puis remplace ne concerne pas que les objets matériels. L'information - considérée de plus en plus souvent comme un autre - subit les mêmes lois avec les mêmes résultats. Emmanuel Hoog, dans une analyse pertinente, qualifie ce processus de « fleuve d'années » [19]. Cependant, et par opposition à ce processus comme pour tenter de le contrebalancer, les sociétés européennes et la société française tout particulièrement n'ont jamais été fébrilement attentives à leur passé et soucieuses de leur mémoire.

Nous le résumons fait que survoler la question et bien des aspects sont comme la préservation des courriels, celle des bases de données ou des sites web ou encore le sujet complexe des multiples formats vidéo n'ont pas été abordés. Les spécificités du secteur de la santé ne sont pas les données du problème. Si le lecteur est peu sensibilisé, cet article aura atteint son objectif. Ce protecteur pourra toujours recourir aux références pour disposer des solutions plus précises à ses besoins particuliers.

SUMMARY
 A key point of long term preservation of digital documents and how to cope with it

The development of digital technologies in all activities sectors of our society leads to a growing number of digital documents. A significant part of these documents needs to be durably preserved. This long-term preservation has to face the short life expectancy of the technologies and the digital storage media. Large national organizations already take this problem into account and set up teams or structures means to face this challenge. At the opposite, the small structures, doctor's offices, individuals, students, etc. are not generally aware of the problem or are stripped to face there. A certain number of simple actions, not requiring specific skills in data processing can nevertheless be undertaken. Without important expenditure, they increase to a significant degree, the security level of the documents over the long term.

CONFLIT D'INTÉRÊTS

L'auteur déclare n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

1. Huc C. Préserver son patrimoine numérique. Eyrolles, 2010 : 324 p.
2. Hourcade JC, Laloë F. Spécificité de l'information numérique. EDP Sciences, 2010 : 106 p.
3. <http://www.long-term-archiving-and-retrieval.org/>
4. <http://www.ina-entreprise.com/entreprise/activites/archives-sauvegarde-numerisation/index.html>
5. http://www.bnf.fr/fr/professionnels/conserver_spar.html

6. Organisation internationale de normalisation ISO. Open Archival Information System - reference model : ISO 2003 : 164 p.
7. Organisation internationale de normalisation ISO. Système d'archivage d'information - modèle de référence : ISO 2003 : 146 p.
8. <http://www.archivesdefrance.culture.gouv.fr/annuaire-services/direction/>
9. <http://esate.gouv.fr/dmp>
10. <http://www.sante.fr/>
11. <http://sin.association-aristote.fr/doku.php>
12. Huc C. La pérennisation des informations sous forme numérique : risques, enjeux et éléments de solution. Méd Sci (Paris) 2008 ; 24 : 653-7.
13. Banat-Berger F, Duplouy LL. L'archivage numérique long terme : les débuts de la maturité. Paris : La Documentation Française, 2009 : 284 p.
14. <http://dmp.cesr.fr/>
15. <http://www.cines.fr/spip.php?rubrique219>
16. <http://hal.archives-ouvertes.fr/>
17. <http://www.ige-adonis.fr/>
18. http://www.lne.fr/fr/r_et_d/gis-don/conserver-donnees-numeriques-gis-don.asp
19. Hoog M. Mémoire année 2008 : Seuil, 2009 : 208 p.
20. Duchange N, Autard D, Pinhas N. Le libre accès : une opportunité pour la recherche biomédicale. Méd Sci (Paris) 2008 ; 24 : 771-5.
21. Rouchon O. La préservation de l'information scientifique et technique. PAC, la plate-forme d'archivage pérenne de documents électroniques du Centre Informatique National de l'Enseignement Supérieur (Paris) 2008 ; 24 : 1099-102.

TIRÉS À PART
C. Huc

Apprenez à respirer à vos enfants

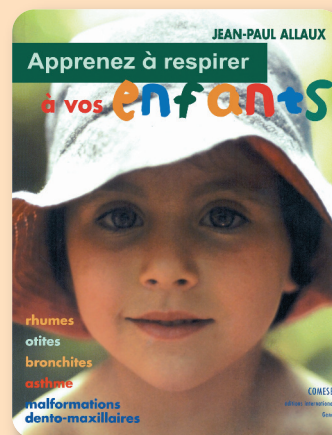
Jean-Paul Allaux

Les maladies ORL avec leur prolongement au niveau pulmonaire sont malheureusement de plus en plus répandues de nos jours et à tout âge.

La pollution atmosphérique, les appartements surchauffés ou enfumés par des parents tabagiques, les rues des villes saturées de gaz d'échappement... tous ces maux contemporains bien connus sont bien souvent responsables de l'installation de ces affections dès la petite enfance. Ils freinent par ailleurs le développement harmonieux du corps et n'installant parfois des déformations vertébrales mais aussi du palais et des mâchoires provoquant une mauvaise implantation des dents.

Fruit d'une longue expérience, ce livre se veut avant tout pratique. De nombreux exercices spécifiques à faire à la maison ou à l'école, seul ou en groupe, permettront aux éducateurs, parents et enseignants de trouver remède à ces différentes affections et d'aider efficacement les enfants à s'épanouir pleinement.

Un livre pour apprendre à respirer à nos enfants.



ISBN : 978-2-7598-0602-7
174 pages
25.00 €

Commandez directement sur : www.edition-sciences.com