

accès
libre

Préserver son patrimoine numérique

Classer et archiver ses e-mails, photos,
vidéos et documents administratifs

Guide à l'usage des particuliers et des entrepreneurs individuels

L'informatique libre à la portée de tous !

L'auteur

Claude Huc, ingénieur au Centre national d'études spatiales pendant plus de 30 ans, puis consultant indépendant, travaille depuis 20 ans sur la préservation des documents numériques. Expert auprès de la Commission européenne, il a fondé, puis animé pendant près de dix ans, le groupe *Pérennisation des informations numériques* qui rassemble nombre d'acteurs publics et privés confrontés à ce problème. Co-auteur, en 2009, avec L. Duploux et F. Banat-Berger de l'ouvrage *L'archivage numérique : les débuts de la maturité*, il a perçu peu à peu l'importance critique de ce sujet pour les particuliers, les entreprises individuelles et les petites structures.

Préserver son patrimoine numérique

Votre patrimoine numérique est en danger :

adoptez les bonnes pratiques et méthodes pour classer tous vos documents et les conserver à moyen ou long terme.

- **Identifiez** les périls qui guettent vos fichiers : changements de format et d'encodage, dégradation du support de stockage, virus...
- **Distinguez** les documents de travail de ceux à conserver durablement
- **Organisez** votre patrimoine numérique avec un arbre de référence
- **Préservez** vos courriels à partir d'un webmail ou d'un client de messagerie
- **Choisissez** votre stratégie de stockage et les supports appropriés à vos besoins
- **Constituez** et sécurisez votre patrimoine numérique
- **Mettez à jour** votre patrimoine selon sa croissance
- **Procédez** à des vérifications régulières et renouvelez vos supports
- **Sauvegardez** votre site web, blog, vidéos en ligne... ainsi que vos données présentes sur les réseaux sociaux

Préserver son patrimoine numérique

Classer et archiver ses e-mails, photos,
vidéos, documents administratifs

Collection « Accès libre »

Pour que l'informatique soit un outil, pas un ennemi !

WordPress 3 pour le blogueur efficace.

F.-X. ET L. BOIS.
N°12829, 2011, 358 pages.

Concevoir et déployer ses sites web avec Drupal 6 et 7.

Y. BRAULT, préface d'Edwy PLENEL.
N°12780, 2^e édition, 2010, 420 pages.

Se protéger sur Internet. Conseils pour la vie en ligne.

X. TANNIER.
N°12774, 2010, 218 pages.

Réussir son blog professionnel. Image, communication et influence à la portée de tous.

T. PARISOT.
N°12768, 2^e édition, 2010, 312 pages.

Bien rédiger pour le Web et améliorer son référencement naturel.

I. CANIVET.
N°12496, 2^e édition, à paraître 2011, 412 pages.

Lisibilité des sites web. Des choix typographiques au design d'information.

M.-V. BLOND, O. MARCELLIN, M. ZERBIB.
N°12426, 2009, 326 pages.

Ergonomie web. Pour des sites web efficaces.

A. BOUCHER.
N°12479, 2^e édition, 2009, 458 pages.

Conversion web. Améliorer ses taux de conversion web.

S. ROUKINE, préface de P. KOSCIUSKO-MORIZET.
N°12499, 2009, 270 pages.

Réussir un projet de site web.

N. CHU.
N°12742, 6^e édition, 2010, 256 pages.

Réussir son site web avec XHTML et CSS.

M. NEBRA.
N°12485, 3^e édition, 2010, 300 pages environ.

Premiers pas en CSS et XHTML. CSS 3 et HTML 5.

F. DRAILLARD.
N°12724, 3^e édition, 2010, 342 pages.

Magento. Réussir son site e-commerce.

M. BLANCHARD.
N°12515, 2010, 352 pages.

Joomla et Virtuemart. Réussir sa boutique en ligne.

V. ISAKSEN, T. TARDIF.
N°12487, 2^e édition, 2009, 316 pages.

Réussir un site web d'association... avec des outils libres !

A.-L. QUATRAVAUX ET D. QUATRAVAUX.
N°12000, 2^e édition, 2007, 372 pages.

SPIP 2 – Premiers pas pour créer son site avec Spip 2.0.3.

A.-L. QUATRAVAUX, D. QUATRAVAUX.
N°12502, 2009, 300 pages.

Boostez votre efficacité avec FreeMind, Freeplane et XMind. Bien débiter avec le Mind Mapping.

X. DELENGAIGNE, P. MONGIN.
N°12696, 2^e édition, 2010, 332 pages.

Montage vidéo et audio libre. Du caméscope au DVD : prise de vue, montage, post-production.

L. BELLEGARDE.
N°12148, 2010, 412 pages avec DVD-Rom.

Scribus. Premières mises en page professionnelles.

C. GÉMY.
N°12547, 2009, 300 pages.

La 3D libre avec Blender 2.49b.

O. SARAJA.
N°12497, 4^e édition, 2010, 486 pages avec DVD-Rom.

Gimp 2.6. Débiter en retouche photo et graphisme libre.

D. ROBERT.
N°12480, 4^e édition, 2009, 352 pages.

Inkscape efficace. Réussir ses dessins vectoriels.

C. GÉMY.
N°12425, 2009, 280 pages

OpenOffice.org 3.2 efficace.

S. GAUTIER, G. BIGNEBAT, C. HARDY, M. PINQUIER.
N°12755, 2010, 412 pages avec CD-Rom.

Linux aux petits oignons. Les meilleures recettes pour bien débiter !

K. NOVAK.
N°12424, 2009, 524 pages avec DVD-Rom.

Ubuntu efficace.

L. DRICOT, K. NOVAK.
N°12362, 3^e édition, 2009, 360 pages avec CD-Rom.

MediaWiki efficace. Installer, utiliser et administrer un wiki.

D. BARRETT, adapté par S. BLONDEEL.
N°12466, 2009, 374 pages.

Richard Stallman et la révolution du logiciel libre.

R. M. STALLMAN, S. WILLIAMS, C. MASUTTI.
N°12609, 2010, 300 pages.

Économie du logiciel libre.

F. ELIE.
N°12463, 2009, 195 pages.

Chez le même éditeur

- G. Gete. – À la découverte de son iPad. N°12823, 2010, 120 pages.
B. Morin. – L'iPhone OS 4 maîtrisé. Pour l'utilisateur avancé. N°12746, 2010, 256 pages
O. Andrieu. – Réussir son référencement web. N°12868, 3^e édition, 2011, 462 pages.
A. Boucher. – Mémento Ergonomie web. N°12698, 2^e édition, à paraître 2011, 14 pages.
M. Lavant. – À la découverte d'Internet. N°12797, 2010, 152 pages.
G. Gete. – Mac OS X Snow Leopard efficace. N°12586, 2010, 370 pages.

Claude **Huc**



Préserver son patrimoine numérique

Classer et archiver ses e-mails, photos,
vidéos, documents administratifs

Guide à l'usage des particuliers et des entrepreneurs individuels

Préface de Françoise **Banat-Berger**

EYROLLES

ÉDITIONS EYROLLES
61, bd Saint-Germain
75240 Paris Cedex 05
www.editions-eyrolles.com



Le code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée notamment dans les établissements d'enseignement, provoquant une baisse brutale des achats de livres, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

En application de la loi du 11 mars 1957, il est interdit de reproduire intégralement ou partiellement le présent ouvrage, sur quelque support que ce soit, sans autorisation de l'éditeur ou du Centre Français d'Exploitation du Droit de Copie, 20, rue des Grands-Augustins, 75006 Paris.

© Groupe Eyrolles, 2011, ISBN : 978-2-212-12789-8

À ma femme Renée, à Gilles, à Ivan, à Nicolas

Préface

C'est avec une grande joie que j'ai accepté d'écrire la préface de cet ouvrage rédigé par Claude Huc. En effet, le thème développé ici répond à un besoin essentiel, sur lequel je reviendrai plus tard. Par ailleurs, j'ai eu le plaisir et l'honneur de travailler avec son auteur dans le cadre du groupe Pérennisation des informations numériques (PIN)¹ et plus largement dans celui de mes fonctions au sein du service interministériel des archives de France.

La question de la conservation pérenne des données et documents numériques devient une question cruciale au fur et à mesure que se développe l'univers numérique.

Cette croissance exponentielle du numérique se manifeste de plus en plus, et c'est évidemment patent dans le domaine de la photographie. Toutefois, au-delà de cet exemple particulièrement évident, nous sommes soumis à une pression de plus en plus forte, visant à nous faire abandonner le support papier, présenté comme inconmode, nuisible à l'environnement et finalement rétrograde.

Certains résistent et continuent à imprimer les photographies qu'ils préfèrent, à refuser obstinément tout envoi de document numérique. Viendra pourtant le jour où il faudra payer pour avoir le privilège de continuer à recevoir du papier, puis celui où cela ne sera plus possible. Il s'agit d'un mouvement profond, que l'on devine irréversible. Pourtant nous n'en mesurons pas bien les conséquences, notamment dans le domaine de la conservation du patrimoine numérique en cours de constitution. Nous sommes nombreux à avoir perdu des données et documents numériques et, dans certaines circonstances, ces pertes se sont révélées très lourdes de conséquences.

1. <http://pin.association-aristote.fr/>

Quelques réponses sont apportées concernant cette conservation du numérique, avec l'apparition d'offres de stockage en ligne. Ces dernières portent des appellations variées, démontrant la difficulté des concepteurs à appréhender ce que recouvre réellement le processus de l'archivage numérique. Ces offres ne prennent en compte que l'aspect stockage, sans offrir de réponse à la véritable question d'enjeu qui est la suivante : comment conserver durablement une information numérique, au-delà de l'obsolescence technologique des matériels, logiciels, périphériques et systèmes d'information sur la base desquels elle a été produite ?

L'initiative de Claude Huc est, par conséquent, particulièrement pertinente et indispensable. L'ouvrage est séduisant car il est écrit par un professionnel extrêmement aguerri sur le sujet. En effet, Claude Huc a joué, pendant quinze ans jusqu'en 2007, un rôle moteur dans la pérennisation des données spatiales au CNES. Il a, dans ce cadre-là, participé à de nombreux travaux de normalisation au niveau international, qui se sont révélés fondamentaux, et il a animé durant dix ans le groupe PIN. Il a su mettre cette grande expérience au service d'un ouvrage pédagogique, en suivant un parcours très didactique. Ainsi, le lecteur se sent-il soutenu sans être assisté.

Le cheminement est logique : l'ouvrage débute par des considérations générales sur le document numérique et sa spécificité, ainsi que sur la définition de ce que l'on appelle l'archivage numérique (dont le champ est bien plus vaste que la simple sauvegarde ou encore le stockage, ou les solutions de coffre-fort numérique). Il continue par un panorama de ce que recouvre ce patrimoine numérique, avant d'aborder les différents champs de l'archivage numérique. Les chapitres suivants sont essentiels car ils donnent des pistes concrètes de stratégies à mettre en œuvre, suivant son activité et le volume de son patrimoine numérique, ainsi que les étapes pré-cises à suivre sous forme de tableaux synthétiques.

La question des coûts n'est pas oubliée et l'on découvre que loin d'être hors de portée de chacun de nous, la conservation de notre patrimoine numérique est avant tout une question d'organisation et de méthode. À la fin de l'ouvrage, nous serons tous convaincus par la pertinence et la faisabilité de la démarche.

Françoise Banat-Berger

Conservateur général du Patrimoine

Avant-propos

Pourquoi cet ouvrage ?

Depuis 20 ans, les technologies numériques envahissent inéluctablement toutes les activités humaines et éliminent, parfois brutalement, les technologies qui les précédaient : les CD audio ont déclenché la disparition progressive des disques vinyle ; les appareils photographiques numériques ont entraîné la mort des procédés argentiques ; les GPS remplacent peu à peu les plans et cartes routières ; la correspondance manuscrite entre les personnes ne résiste pas beaucoup face à l'omniprésence du courrier électronique, des SMS et du téléphone. Ne parlons pas de la télévision, de la radio, des baladeurs, organiseurs, etc., tous ces moyens étant potentiellement interconnectés.

Ce mouvement nous apporte sans cesse de nouvelles possibilités pour créer, traiter, échanger, partager, stocker nos documents numériques. Il est à l'origine de la création et de la croissance de ce que nous pouvons appeler notre patrimoine numérique personnel : nos photographies et vidéos numériques, nos correspondances électroniques, les produits de nos passions d'artistes amateurs, mais aussi de plus en plus souvent, des factures, des relevés de banque ou d'assurance maladie, et bientôt, les bulletins de paie numériques. Il ne s'agit là que de quelques exemples. Ce patrimoine, nous souhaitons le conserver durablement pour nous ou nos enfants lorsqu'il s'agit des documents et témoignages de notre vie privée. Nous devons le conserver aussi longtemps que nécessaire lorsqu'il s'agit de documents à caractère administratif, fiscal, commercial, juridique, médical...

Malheureusement, la puissance fantastique du numérique n'a d'égale que la courte durée de vie des technologies matérielles et logicielles sur lesquelles il s'appuie. Nous sommes ici en plein paradoxe entre cette montée de l'éphémère et des préoccupations mémorielles grandissantes. Pour parler concrètement : comment conserver durablement des documents numériques avec des technologies qui changent radicalement tous les cinq ans, avec des supports de stockage, disques optiques, disques durs, qui vieillissent tous plus ou moins vite jusqu'à ce qu'ils deviennent illisibles, avec des logiciels qui ne savent plus lire les documents produits par leurs anciennes versions ? Tel est le problème posé.

À qui nous adressons-nous ?

La pérennisation d'un patrimoine numérique est un problème général qui affecte toutes les activités de la société, des archives patrimoniales à l'industrie, de la culture à la science, de l'administration à l'université.

Les grandes institutions de mémoire, Archives de France, Bibliothèque nationale de France (BNF) et Institut national de l'audiovisuel (INA) en tête, élaborent des méthodologies et mettent en place des infrastructures techniques lourdes et complexes pour faire face à cet enjeu. Elles sont suivies sur ce terrain par de nombreuses organisations publiques ou privées et par de grandes industries comme l'aéronautique.

Cependant, qu'en est-il de tous ceux qui n'ont à leur disposition ni les moyens financiers, ni la maîtrise technique ? C'est à eux que cet ouvrage s'adresse : il tente de proposer des solutions qui minimisent les risques. Nous nous adressons ici aux particuliers, aux familles, aux citoyens ordinaires, aux étudiants, auxquels nous suggérons quelques principes et méthodes, qui, appliqués avec sérieux et régularité, devraient mettre leur patrimoine numérique à l'abri de pertes inéluctables.

Cet ouvrage pourra aussi être utile aux enseignants, aux chercheurs, aux consultants, aux cabinets médicaux, aux architectes, à tous les professionnels individuels. Il pourra intéresser enfin les structures de petite taille, municipalités, associations qui ont besoin de conserver durablement des documents indispensables à leur activité.

N'oublions pas non plus que les écrivains, les compositeurs et autres artistes ont fait de l'ordinateur leur instrument de travail quotidien. Là aussi, la question de la préservation de leur création se pose.

DOCUMENTS PERDUS

Combien de scénarios de pertes parfois dramatiques de documents ?

2004 : Gilles à Paris voit son ordinateur infecté par un virus ravageur. L'écran devient bleu et plus rien ne fonctionne. Il porte son ordinateur au service de sécurité informatique de son entreprise qui procède à un examen du disque dur, et qui conseille de ne plus rebrancher ce disque dur sur un quelconque ordinateur. Perte sèche et définitive des documents et des quelques centaines de photos présentes sur le disque.

2008 : un ami pianiste de jazz et compositeur amateur, également caricaturiste de talent, après un long et fructueux travail de composition, procède à la recopie de son travail sur un second disque dur. L'opération est mal maîtrisée, l'application utilisée pour composer s'arrête en erreur au moment de la recopie. Tout est perdu. Rage et déception, l'inspiration ne reviendra plus sur ce thème.

2009 : Charlotte à Paris : « pour mon expérience personnelle, j'ai eu à me frotter à la question de la conservation des documents numériques un peu brutalement puisqu'on m'a volé, en mars 2009, mon ordinateur personnel. Je n'avais pas de sauvegarde. Plus tard, j'ai reçu un CD sur lequel le voleur avait recopié les données de mon ordinateur volé, mais pendant 10 mois, j'ai dû vivre avec l'idée qu'une partie de mes documents personnels étaient irrécupérables. »

2010 : un autre ami, retraité, écrivain et poète. En son absence, un petit incendie se déclenche dans sa maison. Ordinateur et disque dur externes sont hors d'usage.

2010 : un cas très classique dont a été victime la fille d'un collègue à Toulouse. C'est un cambriolage, vol de l'ordinateur et en même temps, du disque dur de sauvegarde posé juste à côté. Perte de toutes les photos enregistrées depuis 6 ans. Reconstitution très partielle à partir de celles qui avaient été distribuées à la famille.

Mais que dire aussi :

- des CD mal stockés qui deviennent illisibles car, sur leur étagère, ils ont été un peu trop exposés à la lumière ;
- de l'étudiant qui perd le fichier contenant son mémoire ;
- de l'entrepreneur individuel qui perd son carnet d'adresses ?

Ces histoires se comptent aujourd'hui par milliers.

Relire nos documents dans 20 ou 50 ans ?

Nous créons aujourd'hui des documents numériques avec un logiciel fonctionnant sur un système d'exploitation, lui-même installé sur un ordinateur. Il s'agit de savoir préserver ces documents numériques pour une durée de 5 ans, 10 ans, 20 ans, 50 ans ou plus encore, de façon à ce que leur contenu demeure accessible, à l'aide de logiciels, de systèmes d'exploitation et d'ordinateurs qui n'existent pas encore aujourd'hui.

Ceci nous amènera à une démarche de non-dépendance des documents par rapport aux logiciels et systèmes utilisés pour leur création. Chaque fois que cela sera possible, nous utiliserons les approches ouvertes, les logiciels libres (c'est-à-dire les logiciels gratuits et dont l'utilisation est libre de droit). Surtout, nous éviterons, voire proscrire, toutes les technologies propriétaires fermées qui nous rendent totalement dépendants d'un éditeur de logiciels susceptible de disparaître sans préavis.

Nous ne sommes pas dans un discours théorique ou partisan, mais dans une approche pragmatique, argumentée, qui doit être largement illustrée par des exemples concrets, des expériences vécues, positives ou catastrophiques.

Un objectif raisonnable : l'ambition de cet ouvrage est d'abord de convaincre le lecteur que son patrimoine numérique est presque toujours en danger. Il est ensuite de donner à tous ceux qui n'ont pas de connaissance technique une confiance suffisante dans la sécurité et la pérennité des documents numériques les plus courants, avec juste un minimum de frais et de temps, un peu de méthode et de volonté.

La structure de l'ouvrage

Le **chapitre 1** vise à sensibiliser le lecteur sur la fragilité des documents numériques de toute nature et à remettre en question quelques idées reçues tant sur la durée de vie des supports que sur notre capacité à pouvoir utiliser indéfiniment un document qui a été convenablement stocké.

Le **chapitre 2** constitue une première exploration du sujet : qu'est-ce qu'un document numérique ? Quelles sont ses particularités ? Quels sont

les différents problèmes à résoudre pour le préserver dans des conditions de risque raisonnables ?

Le **chapitre 3** permet de prendre conscience de la richesse et de l'importance croissante du patrimoine numérique des particuliers. Entre les photographies numériques et les documents administratifs, entre les documents liés au travail et aux études et ceux provenant d'activités sociales ou artistiques, ce patrimoine ne cesse de grandir.

Le **chapitre 4** aborde les premières questions pratiques : comment rassembler et hiérarchiser les documents à pérenniser ? Il faut avant tout commencer par savoir où ces documents se trouvent sur le disque dur, avoir une idée sur la meilleure façon de les nommer et enfin regrouper sous une forme organisée tout ce que nous ne voulons absolument pas perdre.

Le **chapitre 5** traite des formats de représentation de l'information, formats d'image, formats bureautiques, formats audio et vidéo, etc. Quels sont ceux qu'il vaut mieux utiliser pour conserver les documents et pourquoi ?

Le **chapitre 6** s'intéresse au courrier électronique (e-mails). Les e-mails sont des documents d'un type tout à fait particulier. Si certains doivent être préservés, comment s'y prendre en fonction du type de messagerie utilisée ?

Le **chapitre 7** présente et analyse les supports de stockage à la portée des particuliers : CD, DVD, Blu-ray, disques durs, mémoires flash. Que faut-il savoir sur ces supports pour s'en servir au mieux ? Les solutions de stockage en ligne, gratuites ou payantes, qui fleurissent sur Internet sont-elles une alternative possible et raisonnable ?

Le **chapitre 8** définit les principes de base à respecter en matière de stockage. Il analyse deux stratégies de stockage possibles entre lesquelles il faudra choisir : stratégie préventive ou stratégie curative. Il propose également des repères pour choisir les supports adaptés au volume des documents.

Le **chapitre 9** entre dans le détail de la suite des opérations : la création d'une version initiale du patrimoine numérique personnel, sa vérification, sa surveillance ainsi que sa mise à jour. Les opérations de migration visant à renouveler les supports de stockage sont décidées en fonction de la stratégie retenue. Enfin, quelques recommandations permettent de gérer au mieux les situations catastrophiques pour lesquelles on croit avoir tout perdu.

Le **chapitre 10** détaille un plan d'action pour savoir comment s'y prendre et ne rien oublier d'important.

Pour aller plus loin, le **chapitre 11** aborde, sans rentrer dans le détail, la préservation des contenus en ligne et des objets complexes pour lesquels il n'y a pas de solutions simples.

Les **annexes** en fin d'ouvrage fournissent quelques compléments sur les formats de fichiers, sur les questions juridiques et sur diverses sources pour approfondir le sujet et se tenir informé.

Nous avons essayé de couvrir les principaux types de problèmes qui seront à résoudre. En fait, chacun d'entre nous n'aura à faire face qu'à une partie de ces problèmes. Pour disposer d'une idée d'ensemble, il est utile de lire l'ouvrage en entier, mais pour répondre au besoin de conservation d'une catégorie particulière de documents, nous pourrons limiter la lecture à ce qui sera directement utile.

Remerciements

De nombreuses personnes ont contribué à la rédaction de cet ouvrage, parfois directement et bénévolement, parfois indirectement. Il n'est pas possible de les citer toutes. Je tiens cependant à remercier tous ceux qui ont passé un temps certain à une relecture attentive et critique du projet, Françoise Banat-Berger et Michel Jacobson (SIAF¹), Benoît Habert (EDF R&D), Danièle Boucon (CNES²), Laurent Duploux (BnF³), Charlotte Fabre (Atos Origin) et Nathalie Regagnon (Archives départementales de la Haute-Garonne).

J'adresse un grand merci également à ceux qui ont apporté une contribution et une expérience irremplaçable sur des domaines spécifiques : Jean-Marc Fontaine (Ministère de la Culture UPMC/IJ DA⁴), Franck Laloë (GIS

-
1. Service interministériel des archives de France.
 2. Centre national d'études spatiales.
 3. Bibliothèque nationale de France.
 4. Université Pierre et Marie Curie/Institut Jean le Rond d'Alembert - équipe LAM (Lutheries – Acoustique – Musique).

DON¹), Jacques Perdereau (LNE²) pour les technologies de stockage, Marc Batllo (CINES³) et Nicolas Larrousse (RISC-CNRS⁴) pour les données vidéo, Émilie Masson (INIST⁵) pour les questions juridiques, Fabrice Andrieux (EGMT⁶) pour la généalogie, Clément Oury (BnF) pour les données en ligne, Xavier Passot (CNES) pour la photographie, Jérôme Barthélémy (IRCAM⁷), Patrick Mazal, Denis Minguillon (CNES) pour la musique, Dominique Heulet (CNES) pour nombre de questions techniques et bien d'autres encore sur tel ou tel sujet pointu.

Enfin, je n'oublie pas que c'est au sein du groupe PIN (Pérennisation des informations numériques, <http://pin.association-aristote.fr/>) de l'association Aristote (<http://www.association-aristote.fr/>), véritable creuset de réflexions, d'expériences, de projets et de partage d'idées nouvelles, que l'idée de cet ouvrage a pu prendre corps.

Claude Huc
claude.huc@laposte.net

-
1. Groupement d'intérêt scientifique sur les disques optiques numériques.
 2. Laboratoire national de métrologie et d'essais.
 3. Centre informatique national de l'enseignement supérieur.
 4. Relais d'information sur les sciences de la cognition – CNRS.
 5. Institut de l'information scientifique et technique.
 6. Entraide Généalogique du Midi Toulousain.
 7. Institut de recherche et coordination acoustique/musique.

Table des matières

1. QUELQUES IDÉES REÇUES	1
Croire que nos CD et DVD sont durables • 2	
Croire qu'il y aura toujours un logiciel capable de lire nos documents ! • 7	
2. LES ENJEUX DE LA CONSERVATION DU NUMÉRIQUE	17
Le temps et la technologie • 18	
L'analogique en quelques mots • 20	
Un document numérique, ce n'est pas magique • 21	
Au départ, on a le codage • 21	
Du codage au format • 25	
Documents et fichiers • 27	
Quatre questions essentielles • 31	
Savoir ce que l'on souhaite conserver • 31	
Retrouver un document, telle une aiguille dans une botte de foin • 32	
Assurer la préservation des séquences de bits • 35	
Pouvoir restituer le contenu à partir des bits • 37	
Conserver pour prouver notre bon droit • 38	
Conventions et éléments terminologiques essentiels • 39	
3. PRENDRE CONSCIENCE DE NOTRE PATRIMOINE NUMÉRIQUE	43
Un patrimoine numérique à conserver • 44	
Les documents personnels et familiaux • 45	
Les photos numériques • 46	
La musique • 48	
La vidéo • 50	
L'écrit • 52	
La généalogie • 52	
Les documents liés aux études et à la carrière professionnelle • 55	

Le cas particulier des données de santé • 56

La gestion des biens et des affaires familiales • 58

Le courrier électronique, cet objet aux usages multiples • 63

Les données professionnelles • 65

Les entreprises individuelles • 65

Les enseignants et les chercheurs • 68

Les données des associations • 69

Savoir ce que l'on veut conserver et pourquoi • 70

4. ORGANISER ET STRUCTURER NOS DOCUMENTS 73

Savoir où sont nos documents sur le disque dur ! • 74

Que faire avec les dossiers prédéfinis ? • 74

Maîtriser la localisation des fichiers • 78

Comment nommer les fichiers et les dossiers ? • 79

Règle 1 - Quels caractères utiliser ? • 81

Règle 2 - Choisir des noms qui traverseront le temps • 83

Règle 3 - N'utiliser que des minuscules • 84

Règle 4 - Codage des dates et des nombres dans les noms de fichiers et de dossiers • 84

Distinguer ce qui est à conserver de tous les documents de travail éphémères • 86

Pour y voir clair, organiser le patrimoine numérique • 88

Définir et construire l'arbre de référence • 88

Une structure d'arbre cohérente et équilibrée • 88

Bien construire son arbre de référence • 91

Le cas des photographies • 93

Transférer les documents dans l'arbre de référence • 95

Rechercher et retrouver les documents • 97

5. LA JUNGLE DES FORMATS 101

Le codage : un principe simple, une mise en œuvre complexe • 102

La tour de Babel • 102

Unicode, le codage universel • 104

L'hexadécimal : une représentation pratique • 106

L'exemple des fichiers de type texte • 107

Pourquoi les formats de fichiers posent-ils problème ? • 112

Le rôle du format de fichier • 112

Savoir de quoi on parle • 113

Les formats sont très nombreux • 115

Formats ouverts et formats secrets ! • 117

Les formats sont comme des poupées russes • 118

On ne peut pas toujours utiliser librement les formats • 119

Autres petits ennuis pour les généalogistes et les musiciens • 120

Le cas de GEDCOM • 120

MIDI • 121

Comment choisir ses formats ? • 122

- Opter pour les formats ouverts • 122
- Limiter le nombre de formats utilisés • 123
- Utiliser des formats largement répandus • 126
- Vérifier l'existence de plusieurs logiciels de lecture • 126

Petit tour d'horizon des formats généraux • 126

- Les documents texte • 127
- Les images • 132
- Les graphiques vectoriels • 132
- La vidéo et l'audio • 133
- Les feuilles de calcul • 140
- Les présentations • 140

Créer le tableau de bord des formats utilisés • 141**Quand lancer une migration de format ? • 144**

- Effectuer une migration de format • 144
- Prendre ses précautions pour les migrations • 145

6. CONSERVER SON COURRIER ÉLECTRONIQUE 149**A-t-on vraiment besoin de préserver les courriels ? • 150**

- Pour pérenniser les messages : bien comprendre les enjeux techniques • 151
 - Le webmail : gestion des e-mails à distance • 152
 - La gestion locale des e-mails avec le protocole POP • 155
 - La gestion synchronisée des courriels avec le protocole IMAP • 158
- Anatomie d'un message • 161
 - Une expérience simple • 161
 - Nous ne maîtrisons pas le codage utilisé dans les messages reçus • 164
 - Comment sont transportées les pièces jointes ? • 167

Préserver les messages à partir d'un client de messagerie • 168

- Préserver quelques messages seulement • 168
- Préserver un certain nombre de messages • 169
- Préserver un grand nombre de messages • 171
 - Où sont stockés les messages ? • 171
 - Dossiers de messages avec Outlook Express et Outlook • 173
 - Dossier de messages sous Thunderbird • 173
 - Changer l'emplacement de stockage des messages à préserver • 174
 - Le processus de préservation des courriels • 175
 - Que faire avec d'autres clients de messagerie ? • 176
- À propos des carnets d'adresses • 176

Préserver des messages à partir du webmail • 177**7. STOCKER OUI, MAIS OÙ ? LES SUPPORTS DISPONIBLES 181****L'information numérique : les ordres de grandeur • 182**

L'incroyable évolution des supports de stockage et ses conséquences • 184

Que deviennent nos documents sur les supports de stockage ? • 186

Séquence de bits : de la vision immatérielle au support matériel • 186

La dégradation des supports au cours du temps • 188

Des possibilités limitées en matière de correction des erreurs de bits • 190

Au niveau des octets • 190

Que se passe-t-il en cas d'erreur de bit non corrigible ? • 192

Durée de vie des supports de stockage aujourd'hui : savoir à quoi s'en tenir ? • 193

Les supports de stockage à notre disposition • 195

Les disques optiques • 196

Les disques durs • 203

Les mémoires Flash • 205

Récapitulatif des durées de vie des supports • 207

Quel usage peut-on faire des offres de stockage en ligne sur le réseau Internet ? • 208

Les offres gratuites : attention aux pièges ! • 208

Les offres payantes : restons très vigilants sur les conditions ! • 210

8. CHOISIR LA STRATÉGIE ET LES SUPPORTS DE STOCKAGE 217

Quelle approche retenir pour le stockage des fichiers ? • 218

La préservation physique • 218

Trois principes majeurs à respecter • 219

Diversifier les technologies utilisées • 219

Minimiser le nombre de supports • 219

Délocaliser • 220

La stratégie préventive : renouveler les supports avant la panne • 221

La stratégie curative : augmenter le nombre de copies et attendre la panne • 225

Choisir les supports • 227

9. CRÉER, VÉRIFIER, SURVEILLER, METTRE À JOUR, SAUVER 233

Créer la version initiale du patrimoine numérique • 234

Étape 1 : version initiale temporaire du patrimoine de référence • 235

Étape 2 : créer la version de référence • 236

Étape 3 : dupliquer la version de référence • 236

Cas des disques durs • 238

Cas des disques optiques, CD, DVD, Blu-ray • 238

Procéder à des vérifications périodiques • 243

Lectures et copies de contrôle • 244

Vérifier la cohérence • 245

Surveiller l'état des supports : est-ce possible ? • 248

SMART et les disques durs • 248

Le cas des disques optiques • 251

Le renouvellement des supports : lancer des opérations de migration • 252**Bien s'y prendre pour les mises à jour • 253**

Les opérations à mettre en œuvre • 254

Simplifier la mise à jour • 256

Cas 1 : un patrimoine à croissance rapide • 256

Cas 2 : un patrimoine photographique à croissance lente • 258

Prendre soin des supports de stockage • 259

Attention aux étiquettes • 259

Les manipulations et le rangement • 260

Créer et renseigner le tableau de bord des supports • 261**Les scénarios catastrophes • 264**

Problèmes avec un disque dur externe ? • 265

Problème avec un disque optique ? • 266

En dernier recours • 267

10. PLAN D'ACTION : CONSTITUER ET CONSERVER LE PATRIMOINE NUMÉRIQUE 269**Constituer et sécuriser le patrimoine numérique • 270**

Étape 1 : réfléchir, investiguer et décider • 270

Étape 2 : s'assurer que l'ordinateur est sain et convenablement sécurisé • 271

Étape 3 : regroupement • 272

Étape 4 : examen des formats et des logiciels • 273

Étape 5 : renommage des fichiers • 274

Étape 6 : le courrier électronique • 276

Étape 7 : vérifier • 278

Étape 8 : choisir les supports et la stratégie de stockage • 278

Étape 9 : constituer la version initiale du patrimoine de référence • 280

Un cas simplifié • 280**Les vérifications périodiques et la surveillance • 281****La mise à jour du patrimoine • 282****Combien tout cela va-t-il coûter ? • 284****11. NOS CONTENUS EN LIGNE ET AUTRES OBJETS COMPLEXES 289****Les sites web • 290**

Créer le site, deux méthodes • 290

Archiver le site • 294

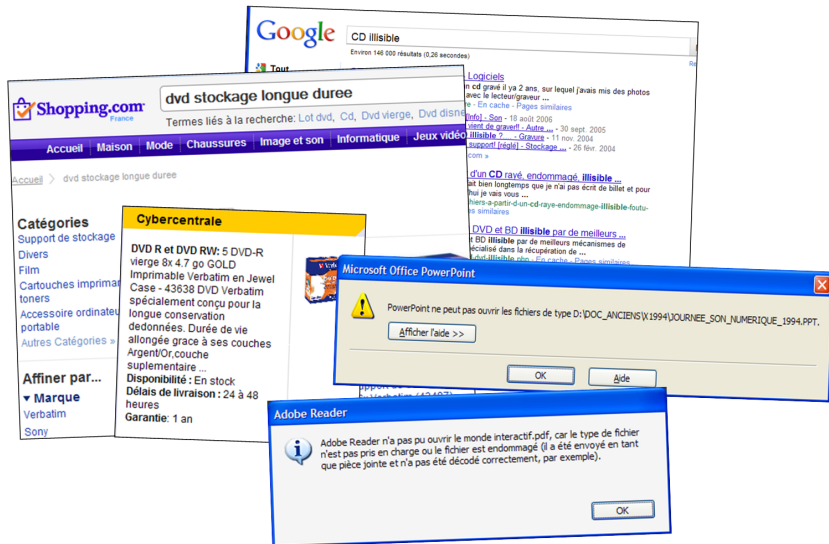
Les blogs • 295**Les réseaux sociaux et les sites de diffusion de vidéos • 297****Les jeux vidéo • 298****Les bases de données • 300**

A. ÉLÉMENTS DE DESCRIPTION DES PRINCIPAUX FORMATS DE DONNÉES	303
Les images • 303	
Les documents textuels et les suites bureautiques • 304	
La vidéo • 306	
L'audio • 309	
Les normes de codage de la série ISO 8859 • 310	
B. ASPECTS JURIDIQUES À CONNAÎTRE	313
Un conflit à propos de la propriété intellectuelle d'un document ? • 314	
La valeur probante d'un document • 315	
Un cas plus simple : disposer d'une copie fidèle • 317	
Nos contrats et achats sur Internet • 319	
C. QUELQUES PISTES POUR ALLER PLUS LOIN	321
Les sites web • 321	
Le groupe PIN • 321	
Les Archives de France • 322	
Une norme essentielle : le modèle OAIS • 322	
Deux ouvrages récents • 323	
INDEX	325



chapitre

1



Quelques idées reçues

Certaines idées reçues nous conduisent à accorder une confiance irraisonnée à la technologie. Elles vont mettre en péril nos documents les plus précieux plus vite qu'on ne le croit.

SOMMAIRE :

- ▶ CD et DVD : quelle longévité ?
- ▶ Il y aura toujours un logiciel capable de lire mes documents !

Il importe de remettre en cause un certain nombre de fausses idées largement répandues, idées fausses mais aussi dangereuses puisqu'elles nous masquent les problèmes essentiels qui seront rencontrés. Deux principales idées reçues sont à mettre en exergue par rapport à la perspective de ce chapitre :

- 1 Croire que les disques optiques numériques (les DON), en particulier les CD et DVD enregistrables, ont toujours une durée de vie suffisamment longue pour qu'on ne s'en préoccupe pas.
- 2 Croire qu'un document créé avec un logiciel pourra, sur le long terme, être relu avec ce même logiciel ou un autre.

D'autres idées reçues, engendrant de multiples risques, seront décortiquées au fil des chapitres.

Croire que nos CD et DVD sont durables

Une certaine confusion entre le Compact Disc audio et le Compact Disc enregistrable (CD-R), mais aussi une argumentation marketing sans fondement scientifique sérieux des grandes marques et une publicité trompeuse des revendeurs sont à l'origine d'une conviction encore tenace dans le grand public : les CD ou les DVD permettraient la conservation des documents enregistrés pendant des dizaines, voire des centaines d'années, comme le laissent croire les annonces présentées sur les figures 1-1 et 1-2. Il ne s'agit là que de quelques exemples parmi beaucoup d'autres.

MALENTENDU Autour du fameux Century Disc

La première annonce (figure 1-1) se double d'un autre malentendu. Dès les années 1990, la société française Digipress (aujourd'hui Plasmon OMS) avait mis au point un CD-R en verre trempé d'une résistance et d'une longévité exceptionnelles. Son coût élevé en faisait un produit réservé aux professionnels. La Bibliothèque nationale de France avait expérimenté ce type de disque avec succès et ne l'a abandonné qu'en raison d'une capacité de stockage insuffisante par rapport à ses immenses besoins. Aujourd'hui, il apparaît sur le marché un produit du même nom, le Century Disc, fabriqué par Mitsubishi, mais dont les caractéristiques techniques n'ont pas grand-chose à voir avec le disque en verre.



FIGURE 1-1 100 ans pour le Century Disc !

Certes, les experts du domaine savent à quoi s'en tenir et ils auront l'impression que nous enfonçons une porte ouverte. Certes, on trouve des informations suffisamment claires à ce propos sur les sites Internet d'un certain nombre d'institutions, mais la croyance est tenace. Nombre de particuliers et d'entreprises pratiquent encore l'enregistrement sur n'importe quel CD ou DVD comme s'il s'agissait d'un acte de conservation définitif, sans s'interroger sur la durée de vie et sur la qualité des disques utilisés.

Kodak CD-R Preservation Gold 5 pack (580138)

12165

[more information](#)

Gold Archive quality 700Mb 52X CD-R 5 pack.

- ◆ For use in long term archiving of images, video & data
- ◆ Licensed Mitsui CDR GOLD technology
- ◆ Rated at up to 300 year lifetime (with proper care)
- ◆ 700Mb / 80 minutes audio capacity.

More about longevity & archiving...

Similar.. Goes with..

Keywords: , Media, CD, 12, cm, 12cm, 120mm, 120, CD-R, CDR, 80min, min, 700, 700MB, MB, 52X, gold, mitusi, GD, archive, master, disc, branded, archive, mam, mam-e, mam-a, kodak, 12165, 580138, single, 1pk,

FIGURE 1-2 Jusqu'à 300 ans pour le Preservation Gold de Kodak !

Un certain nombre d'études sur ce sujet ont été conduites en France par le Laboratoire national de métrologie et d'essais (LNE). Ces études, menées par un laboratoire public, sont neutres et indépendantes de tout intérêt marchand. Elles montrent, sur la base d'analyses statistiques sur des CD et des DVD entreposés dans des sites différents, que la durée de vie de ces disques est souvent réduite à deux, trois ou cinq ans. Elles montrent aussi

que tous les disques ne sont pas équivalents, mais qu'il n'est pas facile de connaître a priori ceux qui sont plutôt fiables et ceux qui ne le sont pas.

À SAVOIR

À propos de la qualité et de la durée de vie des CD et DVD

Le LNE a conduit de nombreux tests de vieillissement accéléré des CD et DVD. Il est intéressant de savoir que :

- Les meilleurs résultats obtenus ne dépendent pas de la marque ou du label. Toutes les marques ont des disques qui résistent mal à ces tests, certaines ont un modèle qui se comporte mieux.
- Les meilleurs résultats obtenus ne dépendent pas de la fameuse couche en or que les fabricants mettent en avant sur les disques haut de gamme. Certains disques de ce type se comportent beaucoup plus mal que d'autres relevant d'une fabrication standard.
- Les meilleurs résultats obtenus ne correspondent pas aux disques vendus spécifiquement pour l'archivage. Certains disques *Archival Gold* se comportent plus mal que d'autres n'ayant pas ces caractéristiques.

Nous reviendrons plus en détail sur ces questions dans les chapitres 7, 8 et 9 consacrés au stockage.

TX-WEA propose un CD enregistrable, le TX Gold CD, garanti 30 ans, ce qui est déjà plus raisonnable que 300 ans ! Après enquête, il apparaît en fait que ce CD est fabriqué par la société française MPO et qu'il est identique au CD MPO Gold recommandé par le LNE. Il s'agit donc ici d'un CD qui a subi des tests de vieillissement accéléré approfondis et objectifs. Ces tests montrent une bonne résistance au vieillissement, mais ne permettent pas pour autant d'évaluer la durée de vie probable !

GARANTIES **Ce que proposent les fabricants**

Il peut être rassurant de lire, sur la notice des supports que nous achetons, une mention concernant la garantie. Il faut savoir qu'il y a très peu de chances de pouvoir être indemnisé à la hauteur de la valeur des données en cas de perte. Au mieux, le support sera remplacé à l'identique (lorsque la preuve d'une gravure de bonne qualité et d'un stockage dans des conditions conformes aux recommandations du fabricant pourra être apportée).



TX Gold CD

TX offers you the best archival media to save your pictures, data, movies, music for a long time.

TX Gold discs preserves much longer data from dye oxidization than classical discs.

TX Gold CD-R 80mn specifications :

- 700 Mb
- 32X Speed
- 24K Gold reflective layer
- High quality phtalocyanine dye
- **30 years warranty with proper handling**
- Made in France

FIGURE 1-3 30 ans pour le TX Gold.

Cette idée reçue sur la durée de vie de tous les CD et DVD est tenace. Elle est aussi mise en avant par de nombreuses sociétés de services de conversion de cassettes vidéo vers des DVD (figure 1-4).

La situation reste cependant complexe, voire confuse, en raison d'une très grande disparité de comportements des disques, avec des CD ou DVD de trois ans illisibles et des CD vieux de 15 ans, des DVD vieux de 10 ans ou plus qui restent parfaitement lisibles. En outre, lorsqu'un disque est illisible, il n'est pas toujours possible de distinguer ce qui est dû à ses matériaux constitutants et à la qualité de sa fabrication, ce qui provient de la gravure et ce qui peut être expliqué par les conditions de conservation. Cette situation confuse permet d'entretenir le doute dans les esprits et facilite les annonces fantaisistes.

NE PAS CONFONDRE CD pressés et CD gravés

Les CD audio sont fabriqués depuis le début des années 1980 et tous ceux qui possèdent des disques achetés il y a 25 ans peuvent encore les lire sur leur lecteur audio s'ils ont été conservés dans de bonnes conditions. Ils devraient pouvoir durer encore beaucoup plus longtemps. Il y a deux raisons principales à cette durée :

- Les CD audio sont issus d'un processus de fabrication incluant la création d'une matrice et un pressage, radicalement différent de celui des CD-R que nous pouvons graver avec notre ordinateur.
- S'il rencontre une petite erreur dans le codage binaire de l'information audio, le lecteur procède généralement à une interpolation qui n'est pas perçue par l'oreille. Une interpolation est une opération mathématique qui permet, par exemple, sur la base d'un certain nombre de points d'une courbe, d'en déduire d'autres. Une telle interpolation ne saurait être envisagée sur un CD contenant des données dont le lecteur ignore la nature.
- Les CD-Rom vendus dans le commerce sont aussi des CD pressés. Ils contiennent des données et des programmes. Ils ne bénéficient pas de l'interpolation, mais possèdent un mécanisme de correction d'erreur supplémentaire qui les protège nettement mieux encore que les CD audio.



FIGURE 1-4 Comment les idées reçues sont parfois véhiculées !
(Photo Jean-Marc Fontaine)

ATTENTION Tous les disques ne sont pas équivalents

Prendre du recul et ne pas croire n'importe quel argument des fabricants fait partie des réflexes de base de tout consommateur averti. Il ne faut pas pour autant en déduire que tous les disques optiques, toutes les marques, tous les modèles sont équivalents. Il a été effectivement et objectivement mesuré que certains se comportaient mieux que d'autres et présentaient de meilleures garanties de conservation. Quelles que soient ces garanties, une surveillance régulière s'impose. Nous reviendrons sur ce sujet dans les chapitres consacrés aux supports de stockage et nous préciserons les modèles de CD ou de DVD qui peuvent être recommandés.

Croire qu'il y aura toujours un logiciel capable de lire nos documents !

Aussi appelé programme ou application, un logiciel est un outil qui permet de traiter les informations comme nous le souhaitons : éditer un document avec un logiciel bureautique, retoucher les photos avec un logiciel de traitement d'images, dessiner, afficher une vidéo sur l'écran, etc.

Il existe un logiciel, appelé « système d'exploitation » (par exemple Windows, Unix, Linux, Mac OS), qui joue un rôle particulier : c'est lui qui assure toutes les fonctions de base de l'ordinateur, qui permet l'utilisation des composants matériels (mémoire, périphériques...) et qui sert de plateforme commune pour le fonctionnement de tous les autres logiciels.

Chacun sait que les ordinateurs fonctionnent sur une logique binaire. Cela veut dire concrètement que tout document numérique se présente dans la machine et également sur les divers supports de stockage que nous utilisons, comme une séquence constituée uniquement de 0 et de 1. La figure 1-5 illustre cette réalité. La notation en 0 et 1 est purement conventionnelle. Ces deux valeurs pourront par exemple se matérialiser suivant la technologie du support comme des creux et des bosses, des orientations différentes des particules magnétiques ou des orientations différentes de réflecteurs de lumière.

```

00101110001100000010000001001101011000010110001101101
00101101110011101000110111101110011011010000000000000
11001000110000001100000011010100111010001100010011000
10011101000110010001100100010000000110001001101100011
10100011001100110010001110100011010000110100000000001
11111111011000111111111110000000000000000100000100101
0010001100100100101000110000000000000000100000001000
00001000000000100100000000000010010000000000000000000
011111111111111100000000000001100010000010111000001110
00001101100011001010100110101100001011100100110101100
00101011111111110110110000000010000100000000000000101
1000010000000100000001010000010000000011100001011000
01010000010010000101011.....

```

FIGURE 1-5 Voici ce qu'est en réalité un document numérique.

Cependant, ce qui nous intéresse véritablement, ce n'est pas cette séquence de 0 et de 1, mais l'information qu'elle véhicule, c'est-à-dire notre photographie, notre texte, notre vidéo sous une forme visible, notre enregistrement audio sous une forme claire et audible !

Pour que cela soit possible, il faut que notre logiciel sache comment notre information est organisée au sein de cette séquence de 0 et de 1, comment les lettres de l'alphabet, les chiffres, les pixels de l'image ou les sons sont codés, rangés, de manière à ce qu'on puisse s'y retrouver. C'est ce qu'on appelle la question des formats de représentation de l'information ou plus simplement des formats.

Ces formats évoluent au cours du temps, plus ou moins vite au gré des changements de technologie, et il en résulte que nous pouvons ne plus disposer de logiciels capables de lire nos documents. Regardons ce problème de plus près.

À SAVOIR Norme et standard

Ces deux termes français sont traduits en anglais par le mot *standard*. Pour distinguer l'un de l'autre quand cela est utile, les Anglo-Saxons utilisent les termes de *standard de jure* (pour norme) et *standard de facto* (pour standard).

FORMATS Deux grandes catégories, les formats normalisés ou standardisés et les formats propriétaires

Les formats normalisés sont définis par un organisme officiel de normalisation : l'Afnor en France (Association française de normalisation), l'ISO au niveau international (Organisation internationale de normalisation). Les formats standardisés sont définis par un groupe collégial (comme le *World Wide Web consortium* qui gère le Web). Dans les deux cas, le format évolue très lentement car le processus de normalisation ou de standardisation implique de trouver un consensus entre les partenaires concernés. Cette évolution est accompagnée de la mise en œuvre d'une vraie compatibilité avec les versions antérieures. Le format d'image *Portable Network Graphics* (PNG), largement utilisé sur Internet, a par exemple été standardisé en 1996 et ne semble pas près d'évoluer.

Les formats « propriétaires » sont, eux, définis par une entreprise privée. Ils évoluent au gré de la politique commerciale de cette entreprise. C'est évidemment dans ce cas que les dangers nous guettent. Les exemples sont nombreux, c'est le cas de la suite logicielle Office de Microsoft (Word, Excel, PowerPoint), de Photoshop fréquemment utilisé pour le traitement des photographies numériques ou encore d'AutoCAD pour la réalisation des plans.

Le danger, on le comprend bien, vient plus souvent des formats propriétaires. Examinons le scénario véridique suivant :

- 1994 – Nous disposons d'un ordinateur doté du système Windows, version 3.0. Nous créons et enregistrons un premier ensemble de documents avec le logiciel de traitement de texte de Microsoft Word, version 2.0.
- 1996 – Nous faisons l'acquisition de la suite Microsoft Office 95, dotée notamment de Word 95 et de PowerPoint 95 pour les présentations.
- 1997-2000 – Nous enregistrons un ensemble de documents textuels avec Word 95 et de présentations avec PowerPoint 95.
- 2000 – Nous installons le système d'exploitation Windows 2000.
- 2000-2008 – Les années passent, la vie continue, nous changeons pour un ordinateur doté du système d'exploitation Windows XP. Nous installons aussi Office 2007.
- 2008 – Nous constatons que nos logiciels Word et PowerPoint d'Office 2007 sont totalement incapables de lire les fichiers enregistrés avec Word 2.0, PowerPoint 95 et Word 95. Les figures 1-6 et 1-7 présen-