Jean-Charles Hourcade, Franck Laloë et Erich Spitz

Longévité de l'information numérique

Les données que nous voulons garder vont-elles s'effacer ?

ACADÉMIE DES TECHNOLOGIES
POUR UN PROGRÈS RAISONNÉ, CHOISI ET PARTAGÉ

INSTITUT DE FRANCE
Académie des sciences

EDP SCIENCES

LONGÉVITÉ DE L'INFORMATION NUMÉRIQUE

Les données que nous voulons garder vont-elles s'effacer ?

Rapport du
groupe PSN (pérennité des supports numériques)
commun à l'Académie des sciences
et à l'Académie des technologies

Membres du groupe :

Erich Spitz,
Académie des sciences et Académie des technologies, président

Jean-Charles Hourcade,
Académie des technologies

Franck Laloë,
LKB/ENS, rapporteur

INSTITUT DE FRANCE
Académie des sciences

EDP SCIENCES

ACADÉMIE DES TECHNOLOGIES
POUR UN PROGRÈS RAISONNÉ, CHOISI ET PARTAGÉ

Conception de la couverture : Jérôme Lo Monaco

Maquette et mise en pages : Patrick Leleux PAO (Lisieux)

Imprimé en France

ISBN : 978-2-7598-0509-9

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays. La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (alinéa 1er de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

© EDP Sciences 2010

Résumé

Pourquoi s'intéresser à la préservation à long terme de l'information numérique, alors que les capacités de stockage numérique n'ont jamais été aussi vastes et bon marché ? C'est qu'il ne faut pas confondre deux notions très différentes, celle de **stockage** des données et celle de leur **archivage**. Les progrès spectaculaires des disques durs et la chute de leur prix permettent maintenant de stocker aisément de l'information, généralement en plusieurs exemplaires, pour s'affranchir des risques de pannes imprévisibles comme les « disk crash ». Mais **archiver** de cette façon sur des décennies ou un siècle pose un tout autre problème, du fait que les supports numériques n'ont qu'une durée de vie de 5 ou 10 ans environ. Dès qu'un disque dur arrive en fin de vie et risque de perdre définitivement les données, il est indispensable de les recopier vers un support neuf. L'évolution des supports étant difficile à prévoir, seul un suivi constant des données permet d'en assurer l'archivage, avec un coût d'organisation important.

Les disques optiques enregistrables ont quant à eux des capacités plus réduites mais sont crédités d'une meilleure durabilité,

comme le suggère le vocabulaire courant : on « grave » des données sur un disque enregistrable. Cette notion de « gravure » nous renvoie inévitablement à l'image des inscriptions antiques gravées dans la pierre et le marbre de vestiges millénaires. Ce sentiment de sécurité est malheureusement trompeur : aucun support actuellement commercialisé ne peut garantir une bonne conservation bien au-delà de 5 ou 10 ans environ !

Nos sociétés génèrent des masses toujours plus grandes d'informations, alors que la durée de vie des supports disponibles pour la conserver n'a jamais été aussi courte. Si ce problème est correctement pris en compte dans quelques organismes publics spécialisés, il est très largement ignoré du grand public ainsi que de la majorité des institutions ou entreprises. Beaucoup d'information, personnelle, médicale, scientifique, technique, administrative, etc. est en danger réel de disparition.

Le groupe PSN (Pérennité des supports numériques) a été créé à l'automne 2008 par les deux Académies, des sciences et des technologies, à la suite à la constatation de cette situation préoccupante, et avec l'ambition de faire un point sur le sujet.

Le présent rapport se donne un périmètre précis (chapitre 1), indispensable pour un sujet où les digressions possibles sont nombreuses. Il se concentre d'abord sur la fraction de l'information qui garde sa valeur à long terme : documents personnels – souvenirs familiaux, données médicales, etc. – ou documents publics – données scientifiques acquises lors d'expériences uniques, etc.

Ensuite sont discutées les stratégies possibles (chapitre 2) : « archive et oublie », dite parfois stratégie passive, la plus naturelle ; la stratégie active (migrations perpétuelles) ; la délégation à un prestataire de service ; le retour à l'analogique.

Les différents supports de stockage sont passés en revue dans un troisième chapitre (disques optiques enregistrables, bandes magnétiques, disques durs, mémoires flash, etc.), avec une brève discussion de leurs qualités et limitations.

Le quatrième chapitre évalue la possible généralisation de la stratégie active à l'ensemble des besoins de la société, qu'il s'agisse de documents personnels du grand public ou de ceux des établissements publics et des entreprises privées.

Le dernier chapitre se concentre sur les disques optiques numériques enregistrables, pour lesquels toute une série de mesures alarmantes ont été effectuées récemment. Il souligne le caractère non fondamental des problèmes rencontrés : si le vieillissement des disques optiques enregistrables est actuellement mal contrôlé, cela tient plus aux priorités qui ont été retenues dans les choix du marché qu'à des raisons essentielles.

Le rapport propose quelques pistes qui pourraient conduire à des disques enregistrables de bien meilleure longévité.

Une série de quatre recommandations est émise en fin de rapport.

This page intentionally left blank

Table des matières

Introduction 11

Chapitre 1. Périmètre du rapport 15

1.1 ■ Une sélection de l'information par le contenu 15
1.2 ■ Les supports, les formats, les logiciels 17
1.3 ■ Les métadonnées, les normes, la certification 18
1.4 ■ Une information personnalisable 19
1.5 ■ Une projection réaliste 20
1.6 ■ Objectif. 20

Chapitre 2. Stratégies possibles 23

2.1 ■ La stratégie passive : « archive et oublie » 24
2.2 ■ La stratégie active : migration perpétuelle 25
2.3 ■ La délégation à un prestataire de service 26
2.4 ■ Le retour à l'analogique 28

Chapitre 3. Les supports d'information 31

3.1 ■ Les disques optiques numériques enregistrables (DONE) 31
3.2 ■ Les bandes magnétiques. 32
3.3 ■ Les disques durs magnétiques 33
3.4 ■ Les « mémoires flash » 35
3.5 ■ Nouveaux dispositifs 36

Chapitre 4. Une stratégie active généralisée ? Évaluation quantitative 39

4.1 ■ Le grand public, volume de données 39
4.2 ■ Dépense par foyer 41
4.3 ■ Établissements et entreprises 42
4.4 ■ Conclusion 42

Chapitre 5. La stratégie passive, les disques optiques numériques. 45

5.1 ■ État de l'art, avantages et inconvénients des disques optiques numériques enregistrables 45
5.2 ■ Processus physicochimiques mis en jeu. 48
5.3 ■ Variantes des disques optiques numériques 50
5.4 ■ Quelques pistes vers un disque optique numérique enregistrable de bonne longévité 52

Conclusion et recommandations 55

Abstract. Conclusion and recommendations 59

Appendices 69

1. Charte de l'Unesco sur la conservation du patrimoine numérique 71
2. Quelques projets français 79
3. Schéma du processus d'enregistrement d'un disque optique numérique 83

4. Quelques images illustrant le vieillissement de disques optiques enregistrables 87
5. Une synthèse du LNE faite à l'occasion de l'audition au groupe PSN. 91
6. Mémoires à nanotubes de carbone 99
7. Quelques idées reçues 101

Liste des auditions du groupe PSN 105

This page intentionally left blank

Introduction

Le sujet du présent rapport est la préservation à long terme de l'information numérique. Chacun sait que cette information est produite journellement en quantités énormes depuis quelques années. Le numérique a maintenant remplacé l'analogique dans presque tous les domaines : l'immense majorité des documents scientifiques, médicaux, administratifs, ou encore les souvenirs personnels (photos, vidéos, etc.), sont directement créés en numérique. L'Unesco estime la production annuelle de l'humanité à plus d'un milliard de Gigabits, soit 10^{18} (un Exaoctet)[1], chiffre qui dépasse l'imagination. Ceci s'explique par la grande commodité de l'utilisation du numérique : flexibilité d'écriture, facilité de réutilisation, stockage compact, transmission à distance aisée et presque instantanée, etc. Le fait que le numérique permette des recopies sans erreur en nombre pratiquement illimité est déjà en soi une nouveauté extraordinaire : auparavant, toutes les copies impliquaient une accumulation progressive d'erreurs et

1. http://portal.unesco.org/fr/ev.php-URL_ID=4805&URL_DO=DO_PRINTPAGE&URL_SECTION=201.html

une dégradation de l'information analogique, qui finissait par disparaître. Pour le numérique, si l'humanité s'y prend bien, en principe rien n'empêche que les informations durent bien plus longtemps que celles écrites sur papier, voire aussi longtemps que les tablettes gravées de l'antiquité ! On comprend que le mouvement vers le numérique soit maintenant irréversible.

Mais il y a loin entre possibilités théoriques et pratique. Dans les faits, la conservation numérique de l'information se heurte à de nombreuses difficultés de nature assez diverse, plus ou moins faciles à résoudre. Ce rapport se concentre sur la principale d'entre elles, la longévité des supports d'information eux-mêmes ; comme nous le verrons, c'est véritablement elle qui est la clé pour la résolution des autres. La multiplicité des problèmes à résoudre entraîne parfois, dans les discussions sur ce sujet, une certaine confusion ; toutes les questions sont abordées à la fois. Dans ce rapport, nous tenterons d'éviter cet écueil et de bien sérier les questions ; c'est pourquoi nous commençons par en fixer précisément le cadre, quitte parfois à mentionner tel ou tel sujet pour nous contenter de dire qu'il ne sera pas traité.

Une confusion courante se produit entre deux notions pourtant très différentes ; celle de **stockage** (ou de sauvegarde) des données à cout terme et celle de leur **archivage à long terme**. La première pose de moins en moins de problèmes grâce aux progrès spectaculaires des supports numériques – chacun connaît ceux des disques durs en termes de capacité et de baisse de prix. Pour stocker sur une durée de quelques années, il suffit de copier les données à sauvegarder sur quelques disques durs (un seul ne suffirait pas à cause des risques de pannes soudaines et imprévisibles de type « *disk crash* ») pour être assuré de les conserver. Mais il existe des données importantes qui doivent être gardées sur des durées bien plus longues, des décennies ou des siècles, pour pouvoir être transmises aux générations futures. On dépasse alors de beaucoup la durée de vie de tous les supports d'information numérique (5 ans environ pour les disques durs, probablement moins pour les mémoires flash). La seule méthode possible est alors de transférer les données d'un support ancien vers un support neuf, avant que la détérioration naturelle du premier ne

rende la recopie impossible ; mais comme le vieillissement des supports n'est guère prévisible, seul un suivi constant des données permet d'effectuer l'opération au bon moment. Un tel processus est coûteux, non pas du fait du prix des supports, mais des interventions humaines et de l'environnement technique qu'il nécessite. L'information numérique dont personne ne s'occupe meurt au bout de quelques années.

This page intentionally left blank

Chapitre 1

Périmètre du rapport

1.1 ■ *Une sélection de l'information par le contenu*

Il ne sera pas question dans ce rapport de chercher à conserver à tout prix toutes les informations que chacun d'entre nous produit ou consulte, y compris des données d'intérêt momentané. Nous partirons du principe qu'une sélection est nécessaire dans le choix de celles qui ont réellement besoin d'être préservées à long terme. De façon très générale, on peut en simplifiant distinguer deux catégories d'informations :

– celle qui prend un intérêt croissant dans le temps, ou du moins garde un intérêt constant ;
– celle qui peut être importante pendant quelques temps, mais dont l'intérêt diminue et va s'effacer progressivement (ou même rapidement) au cours du temps.

Dans la première catégorie, chacun d'entre nous range évidemment les documents de sa mémoire familiale, en particulier tous les documents, photographies et souvenirs légués par nos ancêtres, ainsi que ceux que nous désirons transmettre à nos enfants. Cette catégorie inclut également les documents médicaux (en particulier les images médicales) qui doivent être conservés pendant 30 ans, les documents juridiques et légaux (actes de propriété par exemple) dont la valeur peut durer des siècles, les documents administratifs (nécessaires par exemple au calcul de la pension de retraite). En plus de ces informations plutôt personnelles, il existe également toute une catégorie d'informations publiques qui ont une grande valeur et sont évidemment à préserver : résultats scientifiques (données des satellites et sondes spatiales, des grands accélérateurs, bases de données biologiques, sociologiques, etc.). De plus, les sociétés privées doivent garantir l'accès à l'information concernant des produits complexes qu'elles fournissent, matériels et immatériels (plans de bâtiments, centrales, etc.) ; Dassault Aviation doit conserver les plans de ses avions pendant 70 ans ; les entreprises pétrolières doivent conserver des informations géologiques concernant l'exploitation des sites, qui peut parfois s'interrompre pendant des décennies. Ainsi, dans ce document, nous ne prendrons en compte que cette première catégorie d'information, celle dont la valeur dure longtemps, et sa conservation à long terme (plusieurs décennies, voire plus).

Nous supposerons également que, dès la création de l'information, un minimum de précautions élémentaires a été respecté ; c'est en principe le cas si, dès la saisie, l'utilité de la préservation à long terme était déjà claire. Nous ne nous intéresserons donc pas à la récupération de documents produits dans n'importe quelles conditions, avec un format et un logiciel pris selon les hasards du moment, et sans documentation puisque seul un objectif de court terme est visé. Par exemple, la récupération de l'intégralité des données du disque dur donné par une personnalité, à des fins d'archives historiques, pose parfois des problèmes redoutables « d'archéologie informatique ». La Direction des Archives de France sait à quel point les problèmes de classement de fichiers, en formats fluctuants, de systèmes d'exploitation changeants, etc.

peuvent être délicats, mais ils sortent du cadre de ce rapport. Heureusement il existe également des formats de stockage de l'information qui offrent de bonnes perspectives de pérennité, et nous supposons qu'un choix raisonnable sur ce plan a été effectué.

Dans la même veine, ce rapport ne traitera donc pas non plus des questions relatives aux projets comme la « photographie du web », la conservation de tous les messages électroniques, du dépôt légal des jeux électroniques, etc., non parce que ces questions sont sans intérêt, mais parce qu'elles sortent du périmètre défini.

1.2 ■ *Les supports, les formats, les logiciels*

Ce rapport se concentre sur la longévité des supports physiques sur lesquels est écrite l'information, et donc sur leur évolution dans le temps et la maîtrise des processus de vieillissement intrinsèques, même sans utilisation. Ainsi il ne traite pas :

- des problèmes liés à une mauvaise utilisation ou à une conservation négligente (CD laissé au soleil, rayures ou bris, etc.) ;
- des accidents (incendies, inondations...) et pannes soudaines ; leur prévention relève des techniques classiques de sauvegarde, alors que ce rapport se concentre sur celles de l'archivage à long terme ;
- de l'usure éventuelle lors de la lecture (usure des bandes magnétiques contre les têtes de lecture, blanchiment du colorant d'un CD-R par le faisceau laser, etc.) ;
- des questions de matériel de gravure ou de lecture[2] ;
- des questions de formats et de logiciels, dont la rapide évolution peut rapidement rendre illisibles certains documents. C'est l'aspect qui est le plus généralement saisi par tous, à tel point que c'est presque devenu une idée reçue, chacun

2. Nous disons quelques mots de la question de l'adéquation du couple support-graveur dans l'appendice 7.

s'empressant d'évoquer l'évolution rapide des matériels et des logiciels dès que l'on évoque la conservation des documents numériques. Le problème est certes très réel, mais relativement bien circonscrit dans la mesure où il est possible, à des fin d'archivage, de choisir des formats de fichiers de façon raisonnable ; cet unique aspect du problème ne doit donc pas en occulter d'autres, plus difficiles à résoudre.

Tous ces problèmes sont, certes, très importants, mais ils se situent en aval du problème principal, celui de la longévité de l'information physique sur le support lui-même. Si une information importante est toujours présente sur un support numérique, on saura toujours la relire dans 50 ans ; mieux, l'existence même de cette information sur des supports pérennes suscitera le maintien ou la ré-apparition sur le marché des appareils de lecture nécessaires, avec les formats adaptés. Mais, bien sûr, si l'information elle-même disparaît du support physique, il n'y a plus aucune raison que cette continuité technique soit assurée. C'est bien le support physique de l'information qui est la clé de tout le processus de conservation !

1.3 ■ *Les métadonnées, les normes, la certification*

Nous n'aborderons pas non plus la question du classement et de l'identification des informations (métadonnées) et normes associées. Ces questions sont essentielles pour l'organisation de la communication des données aux utilisateurs, ainsi que de leur collecte, qui font partie de la mission d'organismes tels que le CNES ou la BNF. Elles ne sont cependant pas spécifiques de l'information numérique, mais se posent également pour toute autre forme d'information : des données, quelle que soit leur forme, perdent rapidement tout intérêt si personne ne sait à quoi elles correspondent, ni comment y accéder.

Il existe également des normes de caractère organisationnel plus que technique ; le modèle conceptuel de gestion OAIS (*Open Archival Information System*) sera brièvement mentionné au paragraphe 2.2.

Enfin, il existe des normes techniques, par exemple celles définissant les conditions de vieillissement artificiel en étuve (température, humidité) des disques optiques numériques[3] (ISO, AFNOR). En tant que point de repère permettant de comparer entre elles les mesures différentes, elles sont fort utiles. L'étape délicate est l'extrapolation (d'un grand facteur) pour obtenir une estimation de durée de vie en conditions normales d'utilisation. Cette extrapolation est généralement basée sur une formule d'Arrhenius ou d'Eyring contenant une exponentielle et une seule énergie d'activation, avec tous les risques que cela comporte : dans un système complexe mettant en jeu des processus multiples avec des constantes de temps probablement très différentes, un tel calcul peut conduire à des résultats totalement erronés – nous y reviendrons plus en détail au paragraphe 5.2.

Un autre problème important, mais qui sort du cadre de ce rapport, est la certification : il est important de pouvoir garantir que le fichier restitué au bout de quelques décennies est strictement identique au fichier original. La certification d'intégrité du fichier permet de garantir l'absence de modifications dues à des erreurs techniques, la certification d'authenticité qu'il n'a pas subi de modifications intentionnelles. Cette dernière est, en histoire par exemple, absolument essentielle pour des raisons évidentes. Différents outils sont utilisés pour associer à des fichiers des empreintes numériques (« *hash coding* ») qui permettent de garantir leur intégrité.

1.4 ■ *Une information personnalisable*

Ce rapport ne prend en compte que l'information écrite sur des supports enregistrables, que chacun peut utiliser en quelques exemplaires pour préserver ses données personnelles. Il ne sera

3. La norme ISO/IEC 10995 est la plus récente : http://www.iso.org/iso/catalogue_detail?csnumber=46554 ; elle se limite aux effets de la température et de l'humidité, sans prendre en compte les contaminants chimiques, l'exposition à la lumière, etc.

pas question de la longévité des CD/DVD/BD[4] dits « pressés », tirés à des milliers d'exemplaires, que l'on trouve dans le commerce lorsque l'on achète de la musique ou un film de cinéma. La durée de vie de ces disques est relativement bonne, sans commune mesure avec celle des disques type CD-R ou DVD-R enregistrables à l'unité, mais ils ne sont pas adaptés aux usages qui sont l'objet de ce rapport.

Dans la même perspective, ce rapport n'abordera pas des questions de caractère économique concernant l'industrie audiovisuelle : le futur appartient-il uniquement à une distribution des données par Internet, ce qui signifie que la vente des disques optiques numériques est condamnée à terme ? Une grande partie du public continuera-t-elle à préférer acquérir des objets matériels, que l'on peut plus naturellement offrir en cadeau, ou conserver dans une bibliothèque ? La question est pertinente, mais sort de notre cadre.

1.5 ■ *Une projection réaliste*

Enfin, ce rapport n'est pas un rapport de prospective scientifique, comme pourrait le rédiger un établissement de recherche qui se projette à long terme. Il se limite à l'examen de solutions technologiques qui sont soit disponibles soit proches de l'être. On n'y cherchera donc pas une liste exhaustive de toutes les techniques qui pourraient apparaître et devenir disponibles dans les dix ans à venir. Notre discussion se limitera plutôt aux possibilités techniques réelles à relativement court terme (quelques années).

1.6 ■ *Objectif*

L'objectif général de ce rapport est de mettre en relation des besoins sociétaux essentiels, qu'ils soient culturels, scientifiques, médicaux, administratifs, financiers, etc., avec des solutions

4. La définition de ces abréviations est rappelée en note du paragraphe 3.1.

techniques accessibles. Il s'agit également d'alerter nos concitoyens, souvent mal informés sur le sujet, ainsi que les décideurs. Certes, il existe des méthodes professionnelles qui permettent la préservation des données numériques mais, comme nous le verrons aux paragraphes 2.2 et 2.4, elles demandent des moyens qui en limitent l'application à quelques institutions prestigieuses. Or l'objectif de ce rapport est plus général ; il a pour ambition de dégager des solutions qui soient également utilisables de façon générale, que ce soit au niveau des entreprises, de l'administration ou des individus, satisfaisant le désir légitime de tout citoyen de préserver son patrimoine numérique, familial par exemple. Le besoin de la société concernant la préservation de l'information sur des décennies est réel et profond ; l'espoir est qu'il pourra être pris en compte et que les programmes nécessaires seront lancés pour y répondre.

This page intentionally left blank

Chapitre 2

Stratégies possibles

Nous l'avons rappelé dans l'introduction : l'information numérique a ceci de merveilleux qu'elle peut être recopiée sans erreur à l'infini[5]. C'est la première fois que cela se produit dans l'histoire de l'humanité : auparavant, lors de la recopie de l'information analogique, les erreurs s'accumulaient à chaque copie, avec

5. Cette propriété extraordinaire de l'information numérique a été rendue possible grâce aux travaux fondamentaux de mathématiciens, en particulier ceux de Irving Reed et Gustave Solomon, eux-mêmes élaborés à partir de ceux d'Évariste Galois au XIXe siècle. Leur article « *Polynomial codes over certain finite fields* » a été publié en 1960 dans le *Journal of the Society for Industrial and Applied Mathematics*. C'est cet article qui a introduit les codes de correction d'erreurs travaillant au niveau des bytes. Cependant, à l'époque, la technologie n'était pas prête, de sorte que ces codes ont été considérés comme des curiosités mathématiques ; il a fallu attendre des années avant que leur utilisation devienne générale. Elle est maintenant universelle (disques durs, bandes, CD, DVD, télévision, etc.) ; sans ces codes, presqu'aucun appareil actuel ne fonctionnerait. Le lien avec les travaux de Shannon est également étroit. C'est, une fois de plus, un bel exemple d'une percée technologique de portée considérable permise par des travaux théoriques bien antérieurs.

Voir par exemple : B. Cipra, « *The Ubiquitous Reed-Solomon Codes* », SIAM News, vol. 26, jan. 1993 ; http://www.eccpage.com/reed_solomon_codes.html ; « Code warriors fought errors byte by byte », http://www.wired.com/culture/lifestyle/news/1997/07/5221 ; ou l'article de Gilles Lachaud et Serge Vladut dans La Recherche « Les codes correcteurs d'erreurs » vol. 278, pp. 778-782 (juillet-août 1995).

une croissance exponentielle finissant par noyer l'information dans le bruit. Mieux encore : les codes de correction d'erreur permettent de faire une copie d'un disque optique numérique sur lesquelles les erreurs qui sont apparues entretemps ont été corrigées, pourvu qu'elles ne soient pas trop nombreuses. Le CD recopié est meilleur que l'original qui avait vieilli ; c'est le phénix qui renaît de ses cendres !

Mais cette médaille a son revers : si l'on n'y prend garde, lors des lectures ces codes d'erreur peuvent également cacher l'accumulation progressive des défauts dans le support, et ceci tant que ces codes arrivent à les corriger à la volée. Mais dès qu'un certain seuil est dépassé, ils ne peuvent plus fonctionner du tout, et toute l'information est perdue d'un coup ! Ainsi disparaissent les signaux d'alertes progressives auxquels nous étions habitués avec l'information analogique, comme par exemple le jaunissement progressif des photographies ; ce sont eux qui permettaient de prendre conscience du problème et de prendre des mesures, mais ces garde-fous ont disparu avec le numérique.

C'est pourquoi, comme nous le verrons dans ce qui suit, pour survivre à long terme, l'information numérique a besoin d'un suivi. Avec les techniques actuelles, celle dont personne ne s'occupe meurt au bout de quelques années, comme déjà noté dans l'introduction ; la migration régulière est son seul cordon de vie. De façon générale, quelles stratégies peut-on envisager pour une conservation de l'information à long terme ?

2.1 ■ *La stratégie passive : « archive et oublie »*

C'est la stratégie la plus naturelle, et ceci depuis des siècles : on conserve ses précieux souvenirs dans une boîte à chaussures au fond d'une armoire normande. C'est celle des bibliothèques du monde entier, qui conservent bien à l'abri sur leurs rayons les précieux documents du passé. C'est donc cette stratégie qu'instinctivement chacun d'entre nous a tendance à utiliser, celle par laquelle nous sont parvenus les souvenirs familiaux à travers les générations.

Les fabricants de CD-R ont bien senti ce besoin instinctif, et orienté leur publicité en ce sens. Ainsi, en 2003, on pouvait lire en gros dans une publicité en pleine page du journal *Le Monde* la phrase « vous êtes assurés de conserver pour l'éternité des enregistrements parfaits et inaltérables ». Hélas, il y a bien loin entre ces affirmations et la réalité : les mesures qui ont été réalisées depuis (cf. paragraphe 4.1) montrent à quel point il s'agit d'une illusion. Mais ce besoin de conserver un objet est tellement ancré chez chacun que l'illusion persiste, en particulier chez le grand public, qui de plus ne discerne pas toujours parfaitement la différence entre disques enregistrables et disques pressés (dont la durée de vie est effectivement bien supérieure).

2.2 ■ *La stratégie active : migration perpétuelle*

La « stratégie active » demande un suivi constant des données pour bien assurer leur « migration » régulière sur de nouveaux supports. Cette recopie de l'information d'un support vieux vers un support neuf doit être faite à temps et régulièrement, sans limite dans le temps ; si par exemple ce sont des bandes magnétiques, le délai entre recopies est typiquement de 5 ans. Un suivi technique a pour rôle de s'assurer que la migration d'un support vieux vers un support neuf est faite avant que le premier ne se soit trop détérioré, tant que les codes de correction d'erreur fonctionnent encore parfaitement. Il convient également de dupliquer l'information sur plusieurs supports matériels différents (si possible de marques différentes) qui soient rangés en des lieux distincts (pour éviter les pertes d'information du type catastrophe naturelle : incendie, inondation, etc.).

La méthode active est celle de grands acteurs institutionnels dont la mission comprend explicitement une composante de préservation à long terme : pour la France, BNF, CNES et INA par exemple. C'est également celle des banques et d'autres établissements, publics ou privés. Elle demande des équipes spécialisées, et met souvent en jeu des normes organisationnelles assez

précises (OAIS par exemple[6]). Elle est facilitée par l'existence de robots, qui gèrent un grand ensemble de bandes magnétiques en cassettes (contenant en tout des centaines de téraoctets) et sont capables de les lire à la demande, de les dérouler régulièrement pour éviter qu'elles collent, de les tester à intervalles réguliers afin d'effectuer à temps les migrations nécessaires. Les coûts qu'elle implique sont avant tout liés aux personnels, locaux, et éventuellement aux robots, mais ne dépendent que relativement peu du prix des supports ; les évaluations habituelles qui ne prennent que ces derniers en compte sont totalement irréalistes.

Cette stratégie n'est pas facile à mettre en œuvre pour un particulier, du fait du suivi régulier et permanent qu'elle suppose ; la tâche n'est pas élémentaire car, comme nous l'avons déjà mentionné, l'existence des codes de correction d'erreur fait qu'un CD peut paraître parfait tant que les codes fonctionnent, puis cesser soudainement d'être utilisable lorsqu'un certain taux d'erreurs est dépassé. Laisser à chacun la tâche d'effectuer les migrations de ses propres données semble donc délicat. On peut alors se placer dans une optique de mutualisation de l'archivage actif privé : est-il envisageable de transposer à grande échelle l'archivage actif des grands organismes, par exemple par la création à une échelle nationale de grands centres spécialisés, afin de permettre la conservation de toute l'information numérique souhaitable, y compris par exemple celle des familles ? Nous reviendrons à l'analyse de cette question au paragraphe 2.4, nous contentant de mentionner pour le moment que le réalisme de cette solution n'a rien d'évident, dès lors que l'on en calcule les conséquences chiffrées.

2.3 ■ *La délégation à un prestataire de service*

Une troisième possibilité découle tout naturellement de ce qui précède : la « stratégie de délégation », qui consiste à confier ses

6. OAIS (*Open Archival Information System*) est un modèle conceptuel destiné à la gestion, l'archivage et à la préservation à long terme de documents numériques ; voir http://fr.wikipedia.org/wiki/OAIS

données à un prestataire de services spécialisé. Moyennant rétribution, celui-ci devient responsable de leur préservation à long terme. Dans beaucoup de cas, grâce à Internet, ce prestataire pourra être très lointain. Pour plus de sécurité, des recopies de sauvegarde dans des continents différents sont même possibles ; les techniques modernes permettent de délocaliser un fichier entre des machines différentes, parfois situées à des milliers de kilomètres. Il est clair que cette façon de faire offre d'excellentes garanties contre la brusque perte de données, comme celle due à un accident naturel. Certains y voient la panacée, d'autant plus que des entreprises commerciales comme Google offrent gratuitement un certain espace de stockage, le financement étant assuré par la publicité. L'idée que la délocalisation de l'information dans le monde entier (ou de sa « virtualisation ») résout tous les problèmes est assurément dans l'air du temps ; elle bénéficie d'une proximité avec des concepts en vogue comme le « *cloud computing* » ou le « *grid computing*». Elle peut certes apporter une solution satisfaisante à des problèmes de sauvegarde, mais il n'est en rien évident qu'il y ait là une voie raisonnable pour l'archivage à long terme.

Des entreprises de toute taille existent qui proposent des services de sauvegarde à distance. Elles fonctionnent souvent selon un principe semblable à celui des compagnies d'assurance : une petite entreprise de stockage numérique s'assure auprès d'une autre compagnie plus puissante, en lui louant de l'espace de stockage où dupliquer ses propres données, et ainsi de suite, ce qui diminue fortement l'impact d'un accident soudain à un échelon quelconque du dispositif.

Cette stratégie n'est toutefois pas non plus exempte de problèmes. Le réseau Internet en lui-même ne stocke évidemment rien, puisqu'il ne peut que mettre en contact un utilisateur avec des réservoirs de préservation lointains. Une information stockée dans un centre en Asie ou en Amérique l'est donc sur une bande ou un disque dur qui, lui non plus, n'a pas une durée de vie infinie ; les mêmes arrêts soudains, les mêmes processus de vieillissement se produisent. Bien évidemment, le prestataire de service devra à son tour résoudre les mêmes problèmes techniques et

choisir entre l'une des deux premières stratégies. Il faut donc le sélectionner avec grand soin pour pouvoir raisonnablement espérer que le travail de conservation sera bien fait avec la continuité requise.

De plus, le système de conservation, qui n'était déjà pas transparent, devient alors plus opaque encore : est-il véritablement plus sûr de faire conserver ses trésors à l'autre bout du monde, sans qu'il y ait la moindre possibilité de voir dans quelles conditions la préservation est effectuée ? En réalité, à l'examen, on réalise que ce système semble plutôt adéquat pour la préservation de l'information à court et moyen terme, quelques années par exemple. Autant il est courant de faire appel à des sociétés de services lointaines pour des services à relativement court terme, autant il n'est pas ordinaire de le faire sur des décennies. Que se passera-t-il si l'entreprise spécialisée dans la conservation à long terme disparaît ? Qui s'occupera de rendre l'information à ses propriétaires ? Ces derniers sauront-ils même qu'ils doivent la réclamer, et à qui ? De plus, les questions de confidentialité peuvent devenir particulièrement cruciales lorsque les données présentent un caractère sensible : chacun sait que le cryptage logiciel des données n'est jamais parfaitement sûr. Enfin, il est évident pour chacun que cette stratégie est particulièrement soumise aux dangers de nature géopolitique.

Dans la même veine, beaucoup parlent depuis quelque temps de la « virtualisation des fichiers » (mot qui, en fait, recouvre des techniques assez variables d'un prestataire à un autre). Cette méthode peut probablement rendre de grands services, mais en termes de préservation à long terme n'apporte aucun avantage très particulier.

2.4 ■ *Le retour à l'analogique*

Quant à l'industrie américaine du cinéma, mondialement connue, elle qui utilise massivement les techniques numériques pour la réalisation des films, elle semble opter pour un étonnant retour en arrière. Un rapport de 70 pages fort bien documenté,

« *The digital dilemma* », écrit en 2007 par le *Science and Technology council of the academy of motion pictures arts and sciences*, en explique les raisons. En particulier, le chapitre 6, *Digital motion picture archiving economics*, donne une argumentation précise concernant les coûts associés à la stratégie de migration, dans une perspective de conservation pour un siècle ou plus.

La conclusion de cette étude est que, pour préserver des films de cinéma à long terme, la meilleur méthode consiste à en faire des copies analogiques (même pour les films tournés en numérique) en grand format et en trois exemplaires monochromes séparés (pour éviter la dégradation des colorants), et à stocker les bobines dans des souterrains profonds où température et humidité sont particulièrement bien contrôlées. Le stockage numérique n'est pas pour autant éliminé, bien évidemment, mais son rôle serait confiné à des usages à court et moyen terme. C'est une belle illustration des problèmes difficiles posés par la conservation de l'information numérique à long terme.

La possibilité technique de graver de façon très résistante des images analogiques sur des substrats optiques existe. Ainsi la firme grenobloise Arnano grave des disques de saphir avec une résolution de 20 nanomètres dont on peut attendre une durée de vie très longue (utilisées en horlogerie de luxe[7], ces inscriptions seraient garanties 2 millions d'années !). Un tel disque peut contenir plusieurs milliers de pages de documents, images, etc. mais, bien sûr, ni fichier sonore ni images animées. Ces techniques sont cependant trop onéreuses pour une utilisation systématique par les particuliers.

7. http://leblogdesmontres.fr/2008/09/15/romain-jerome-titanic-dna-rj21-ultimate-nanocapsule-nano-time/

This page intentionally left blank

Chapitre 3

Les supports d'information

Passons rapidement en revue les supports d'information numérique disponibles à l'heure actuelle.

3.1 ■ *Les disques optiques numériques enregistrables (DONE)*

Les disques optiques sont l'objet principal de ce rapport. La façon dont ils fonctionnent est brièvement rappelée dans l'Appendice 3. Leur production mondiale est considérable : plus de 10 milliards de DONE sont produits chaque année ! La grande commodité qu'ils offrent permet de s'en servir pour stocker des données, pour les envoyer par la poste, pour communiquer aux participants à un congrès le contenu des communications, etc. Pour le moment, les DONE se déclinent en CD-R (ou -RW), DVD-R (ou -RW), et maintenant BD-R (Blu-ray)[8]. Ces supports

8. Les CD correspondent à la première génération de DONE de 12 cm de diamètre, avec une capacité de l'ordre de 600 Mo ; les DVD une capacité d'environ 5 Go ; les BD de 25 Go. Le suffixe – R (*Recordable*) signifie « inscriptible une seule fois », le suffixe – RW (*ReWritable*) « inscriptible et effaçable ensuite pour une nouvelle inscription ».

ont pour eux leur simplicité conceptuelle : ils sont construits tout d'un bloc, sans aucune pièce mécanique mobile. Ils sont donc totalement passifs une fois qu'ils sont enregistrés : ils sont lus de façon purement optique, sans aucune usure mécanique ; leur seul rôle est de renvoyer la lumière qu'ils reçoivent vers un détecteur lumineux. C'est d'ailleurs précisément pour cette raison que beaucoup ont longtemps cru que ces supports seraient éternels, confondant usure lors de la lecture et longévité naturelle. Malheureusement nous savons maintenant que, dans la réalité, ils se dégradent constamment, même quand ils ne sont pas utilisés.

On parle depuis quelques années de super résolution et de disque holographique. Dans le premier cas, l'espoir est d'arriver à mettre sur un seul disque une quantité d'information de l'ordre de 200 Gigaoctets. Hélas, cela apparaît déjà comme faible comparé à la capacité des disques durs magnétiques devenus maintenant courants ; on a un peu l'impression que les DONE ont déjà perdu la course de la capacité. Pour l'holographie, elle permettrait de mettre en œuvre des techniques d'accès en parallèle à l'information, mais il s'agit toujours de programmes de laboratoire qui semblent relativement loin de débouchés industriels (voir cependant § 5.3).

3.2 ■ *Les bandes magnétiques*

Elles sont bien antérieures à la large diffusion dans le public des techniques numériques, puisqu'elles ont été très largement utilisées en enregistrement analogique (leur invention date de 1928). Elles ont commencé à être utilisées pour les données numériques dès 1951. Ce sont donc, historiquement, les premiers supports numériques de stockage qui se sont largement répandus, bien avant les disquettes magnétiques ; elles restent actuellement toujours le pilier de base de la « stratégie active ». Pour des raisons de commodité et de meilleure conservation, les bandes numériques sont souvent contenues dans des cassettes.

En termes de coûts et de capacité de stockage, les bandes fournissent un support particulièrement économique. En revanche,

leur nature physique les rend relativement sensibles à l'usure du temps : au bout de quelques années, la couche d'oxyde dont elles sont recouvertes tend à se décoller et à tomber en poussière ou rester sur la tête de lecture ; le support plastique commence à devenir cassant. Pour être certain qu'une bande reste utilisable, il convient de la dérouler et l'enrouler à nouveau à intervalles réguliers.

Les bandes conduisent à des temps d'accès relativement longs aux données. Elles sont donc utilisées principalement comme support de sauvegarde et de préservation, mais pas comme support de communication (sur Internet par exemple).

3.3 ■ *Les disques durs magnétiques*

Les disques durs ont été introduits en 1956, comme des appareils réservés aux centres de calcul spécialisés. Ils reposent sur un principe physique semblable aux bandes, mais offrent un accès bien plus rapide aux données. Depuis leur apparition, ils n'ont cessé de faire des progrès spectaculaires chaque année, en termes de capacité toujours plus grande et de réduction des prix. L'impact considérable des travaux fondamentaux d'Albert Fert et Peter Grünberg est bien connu ; il leur a valu le prix Nobel en 2007. Actuellement, on peut acheter dans le commerce un disque dur portable d'une capacité de 1 téraoctet pour environ 100 euros, ce qui pouvait sembler inconcevable il y a quelques années seulement.

Les disques durs sont cependant des constructions mécaniques complexes, avec une tête de lecture interne qui se déplace et lit les données à 10 nanomètres environ de la galette magnétique qui tourne très rapidement (autour de 10 000 tours par minute), et sur laquelle la couche active a également une épaisseur de l'ordre de 10 nanomètres (sous une couche de protection de 5 nanomètres environ). Au vu de la précision mécanique extrême requise, on comprend qu'ils soient fragiles et supportent mal les chocs. Ils sont sujets aux pannes brutales (« *crash* »), qui se produisent soudainement, sans prévenir, et entraînent une perte totale des données. On constate dans la pratique qu'un disque dur, une fois arrêté,

a une probabilité non nulle de ne jamais redémarrer, causant ainsi une perte totale de ses données. C'est pourquoi, dans les grands centres, les disques durs restent constamment alimentés, tournant sans cesse pendant toute leur durée de vie. Il y a là un point important à garder à l'esprit : si l'on veut estimer la viabilité d'une stratégie de conservation des données numériques *via* l'utilisation du réseau et de grands centres numériques utilisant des disques durs, il faut prendre en compte la consommation électrique que cela suppose, ainsi que ses effets environnementaux.

Des études récentes[9] montrent la difficulté de prévoir les pannes ainsi que le comportement des disques durs sur quelques années[10]. Si on utilise des disques durs avec un objectif de préservation à long terme, il faut bien sûr prévoir une généreuse redondance de l'enregistrement des données (technique des sites miroirs éloignés), mais aussi garder à l'esprit qu'il s'agit d'une conservation « doublement active » : non seulement les équipes de migration, mais aussi les machines, ne s'arrêtent jamais (contrairement aux bandes magnétiques qui restent dans des armoires).

Des disques durs qui tournent sans arrêt dans le monde entier sont à la base du succès planétaire et du fonctionnement du moteur de recherche Google. Cette entreprise possède dans divers continents un grand nombre (qui semble tenu secret) de hangars remplis de consoles contenant des myriades de disques durs. Ceci permet une grande redondance de l'information, répartie dans le monde entier, et évite les risques dus aux catastrophes comme incendies et inondations. Les pannes qui se produisent sont rapidement réparées, mais les fluctuations qu'elles engendrent expliquent que les réponses du moteur de recherche puissent fluctuer d'un jour sur l'autre ; chacun peut ainsi constater de visu que les données de Google sont sensibles aux pannes de ses disques durs.

---

9. E. Pinheiro, W.D. Weber et L. A. Barroso (Google corp.) « *Failure trends in large disk drive population* », 5e USEMIX conference FAST (2007). B. Schroeder et G. A. Gibson (Carnegie Mellon Univ.), « *Disk failures in the real world* », 5e USEMIX conference FAST (2007).

10. Les techniques RAID (*Redundant Array of Independent Disks*), relativement classiques maintenant, sont efficaces pour se prémunir contre les pannes soudaines des disques durs (*crashes*), mais n'ont pas pour but une préservation à très long terme des données.

3.4 ■ *Les « mémoires flash »*

Depuis quelques années, ces mémoires sont devenues d'usage très courant, du fait de leur grande commodité. Chacun sait qu'elles servent maintenant dans les appareils photo ou enregistreurs vidéo, les « clés USB » pour les données, les baladeurs, et même remplacent les disques durs de certains ordinateurs portables. Avec les mémoires flash, on ne stocke plus l'information sous forme d'orientation magnétique de petits domaines ; on utilise le déplacement de charges électroniques dans de petits transistors avec deux grilles, une de contrôle et une flottante en suspension dans un oxyde. Les mémoires flash sont sujettes à un phénomène d'usure lors des processus d'écriture et de lecture, phénomène qui limite le nombre de cycles à une valeur comprise entre 10 000 et 100 000 ; cependant des progrès sont espérés. En termes de rapidité d'accès, elles sont aussi performantes, voire plus, que les disques durs ; n'ayant aucune pièce mobile, elles n'ont pas l'inconvénient de nécessiter une mécanique de précision et une rotation rapide. C'est actuellement le prix par bit ainsi que le nombre limité de cycles qui sont les raisons principales pour lesquelles elles ne détrônent pas les disques durs dans tous les ordinateurs personnels.

Pour le moment, personne ne considère les mémoires flash comme des supports de stockage à long terme. On peut en effet penser que la migration de charges électriques soit encore plus difficile à éviter que celle de l'orientation des domaines magnétiques, et redouter les conséquences de tous les effets ionisants (rayons cosmiques ou vent solaire par exemple). Mais des études plus précises sur ce sujet restent à faire ; il serait fort intéressant de voir si quelques modifications accessibles comme le changement des tensions ne leur permettraient pas de jouer ce rôle. On peut d'autre part songer à des solutions de « rafraîchissement périodique » de l'information. Dans un contexte de préservation à long terme, la question du nombre de cycles d'écriture devient relativement annexe : si l'on pouvait construire une mémoire flash qui ne permette que 4 ou 5 cycles, mais conserve l'information pour des décennies, un grand pas en avant serait fait.

3.5 ■ *Nouveaux dispositifs*

Il existe bien d'autres supports qui sont en cours de développement. On a beaucoup parlé du « millipède » de IBM, basé sur des techniques du type MEMS, ou encore les mémoires à microscope à force atomique. Il est difficile d'en faire une liste exhaustive, tous les projets n'étant pas publics ; de plus, une idée totalement nouvelle et imprévue peut apparaître à tout moment à partir d'une découverte dans un domaine quelconque de la physique.

Parmi les projets les plus connus, on peut mentionner :

– Les mémoires à changement de phase (PRAM, pour *Phase Change Random Access Memory*), dont les éléments sont faits de matériaux chalcogénures. Ils sont semblables à ceux qui sont utilisés dans des CD-RW et DVD-RW, mais ici c'est un courant (et non un laser) qui est utilisé pour chauffer le chalcogénure, qui se refroidit ensuite lentement ou rapidement selon l'état du bit désiré ; comme l'état vitreux présente une résistivité électrique plus grande que l'état cristallin, la lecture est également électrique et non optique. Une application industrielle de ces PRAM est la fabrication de mémoires résistantes aux radiations, à des fins d'utilisations spatiales.
– Les mémoires RRAM (*Resistive Random Access Memory*) où c'est également la résistance électrique d'un bit élémentaire qui peut être basculée entre deux valeurs ; il s'agit souvent de nanocouches d'oxydes métalliques dont on bascule la résistance par des impulsions de tension, qui y produisent des filaments conducteurs. HP Labs annonce un prototype nommé « memristor ».
– les STT-RAM (*Spin Torque Transfer RAM*), amélioration des mémoires MRAM (magnetic RAM) où c'est la direction de l'aimantation de spin qui stocke l'information dans chaque bit (comme pour un disque dur, mais avec une architecture matricielle). On fait usage de courants électriques dont les électrons sont polarisés (au lieu d'impulsions magnétiques pour les disques durs). Une dizaine de grandes compagnies dans le monde travaille à la mise au point de tels produits.

– les mémoires à nanotubes de carbone (NRAM), présentées comme particulièrement prometteuses en termes de densité, rapidité, insensibilité aux perturbations et même de durabilité. Elles font appel au concept de « *suspended nanotube junctions* », où le nanotube joue à la fois le rôle d'élément structurel géométrique et de conducteur moléculaire. L'action d'un champ électrique produit un basculement entre deux états[11]. Ce projet, développé par une firme privée, est issu de l'Université de Harvard.

Mais on peut aussi s'intéresser à des projets à plus long terme, comme par exemple (cette liste ne prétend pas être exhaustive) :

– Un projet issu de l'université de Berkeley[12] où, cette fois, le nanotube de carbone est utilisé comme un conduit mécanique dans lequel se déplace une minuscule masselotte, faite d'un nanocristal magnétique, dont la position dans le tube permet de stocker l'information avec plus de cent valeurs différentes. Les auteurs annoncent des applications prometteuses pour l'archivage, avec une durée de vie thermodynamique dépassant un milliard d'années ! Ce projet, peut-être un peu futuriste mais très intéressant, est décrit brièvement en Appendice 6.
– La « *Digital Rosetta Stone* ». En fait, ce nom a été choisi par plusieurs projets différents, dont certains n'ont d'ailleurs rien à voir avec la pérennité des supports numériques ; le plus récent (2009) est issu d'une collaboration entre deux universités japonaises (celles de Keio et de Kyoto) et Sharp Corporation[13], qui ont réalisé un empilement de galettes mémoires qui sont ensuite noyées dans un moule de silice. L'accès aux galettes se fait par transmission sans fil, avec un taux de transmission de 150 Mb par seconde ; la capacité est

11. Thomas Rueckes *et al.*, « *Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing* », ou http://www.nantero.com/mission.html

12. G. E. Begtrup *et al.*, « *Nanoscale reversible mass transport for archival memory* », Nanoletters, vol. 9, pp. 1935-38 (2009).

13. http://techon.nikkeibp.co.jp/english/NEWS_EN/20090618/171883/

de quelques To. La durée de vie annoncée est de plusieurs milliers d'années !

Pour revenir aux divers supports disponibles à l'heure actuelle, que conclure de ce rapide panorama ? Pour finir, parmi eux, les seuls qui sont actuellement disponibles et reposent sur un principe de fonctionnement suffisamment simple et passif pour que l'on puisse raisonnablement espérer contrôler leur vieillissement sont les disques optiques numériques (sauf révolution technologique imprévue, pour les mémoires flash par exemple).

Chapitre 4

Une stratégie active généralisée ? Évaluation quantitative

Il est intéressant d'imaginer ce à quoi pourrait mener un mouvement de grande ampleur vers des solutions de stockage et de préservation en stratégie active. Nous l'envisagerons à l'échelle d'un pays comme la France, le jour où le grand public et l'ensemble des acteurs concernés réaliseraient à quel point les supports des DONEs actuels sont en réalité peu fiables, et s'il n'existait à ce moment-là aucune solution alternative crédible de stockage passif.

Nous procéderons en deux temps : l'évaluation de la quantité de données à préserver à long terme, puis les coûts associés. Afin que les résultats soient plus parlants, chaque fois que possible nous prendrons comme unité le foyer français.

4.1 ■ *Le grand public, volume de données*

Par grand public il convient de comprendre non seulement les ménages et les utilisations domestiques et ludiques, mais aussi les

initiatives individuelles de stockage d'information dans le cadre professionnel qui ne sont pas prises en charge par l'entreprise au sein d'un processus structuré (exemple : un ingénieur dans son bureau d'études qui conserve dans le tiroir de son bureau une mémoire de ses projets sur DVD-R).

Le « patrimoine numérique » d'un foyer représente toutes les données qui, pour une raison ou une autre, lui apparaissent dignes d'être préservées sur le moyen et long terme, soit pour des raisons objectives (documents juridiques et financiers, données médicales, certificats...), soit pour des raisons sentimentales, soit pour être léguées aux générations postérieures (photos et vidéos...). Les formats photo et vidéo sont évidemment les plus consommateurs, mais des données issues de documents scannés peuvent également contribuer de façon significative.

Une hypothèse basse ne prend en compte que les photographies et les documents, soit scannés, soit natifs, mais sans inclure la vidéo. Une estimation de l'ordre de 200 photos de 4 Mo chacune[14] conduit à un ordre de grandeur de 1 Go par an. Sur la base d'une hypothèse de 25 ou 50 années d'accumulation, on arrive à la constitution d'un patrimoine cumulé d'informations numériques à préserver de l'ordre de 25 à 50 Go par foyer. Pour la vidéo, on peut estimer que chaque foyer produit annuellement et désire conserver environ une heure de vidéo qui, après compression, représente environ 2 Go, le double en haute définition. L'évolution continue de la qualité des formats vidéo étant compensée en partie par la performance croissante des systèmes de compression, il est raisonnable de compter sur un volume moyen de 4 Go par heure d'enregistrement. On arrive alors à la constitution d'un patrimoine cumulé d'informations numériques à préserver de l'ordre de 100 à 200 Go. Compte tenu de l'évolution des techniques toujours plus consommatrices de données, on peut prendre une marge de précaution et donner une estimation globale

14. Ce chiffre correspond environ à une résolution de 10 Mpixels, actuellement courante.

se situant entre 100 Go et 1 To par foyer[15] – il doit cependant être clair que nous ne recherchons ici que des ordres de grandeur pour évaluer la dimension du problème, à un facteur 10 près, voire plus.

4.2 ■ *Dépense par foyer*

La question est alors de déterminer le coût de la préservation de ce volume d'information, dans le cadre d'une stratégie active de vérifications régulières et de migrations permanentes. Là encore, seule une approximation de l'ordre de grandeur peut être entreprise, les points de comparaison étant rares et dispersés. Il est essentiel de bien prendre en compte le coût humain du dispositif (salaires, locaux et environnement, etc.) : il n'existe pas de stratégie active fiable sans surveillance humaine régulière. Même si des économies d'échelle peuvent probablement être atteintes à terme, il reste prévisible que le facteur travail l'emporte largement sur le coût des équipements et consommables. Ceci se vérifie déjà dans les quelques situations qui ont pu être discutées au sein du groupe PSN, comme à l'INA, au CNES ou à la BNF. Sur ces bases certes parcellaires, le coût annuel de stockage et de préservation de 1 Po dans le cadre d'une stratégie active est ainsi estimé entre 300 K euros et 3 M euros.

Alors que le prix d'achat d'un To « grand public » est aujourd'hui tombé aux environs de 100 euros, avec une durée de vie moyenne de l'ordre de 3 à 5 ans, le prix d'un To faisant partie d'un système SAN (Storage Area Network) avec redondance et infrastructure d'administration se situe plutôt entre 800 euros et 8000 euros, selon les niveaux de performance et de sécurité recherchés. Il n'est pas déraisonnable de converger vers un coût annuel par To, intégrant les facteurs d'échelle mais intégrant

15. Il y a environ 25 millions de foyers en France, ce qui donne l'ordre de grandeur national de 2 500 à 25 000 Po comme évaluation du patrimoine numérique des foyers français, qui se trouve être du même ordre de grandeur que ce que Google met aujourd'hui sur disques durs pour pouvoir travailler sur des copies du Web !

également le plein effet du facteur humain, qui se situe entre 300 et 1 000 euros.

4.3 ■ *Établissements et entreprises*

Peu d'établissements publics en France ont pour le moment mis sur pied une politique systématique de sauvegarde à long terme des données numériques qui soit au même niveau technique que le CNES et la BNF, le CNES ayant plus de recul dans cette mission de conservation. Les estimations faites par le groupe PSN à la suite des auditions de ces deux établissements aboutissent à un coût de l'ordre de 2 000 à 3 000 euros par an et par To.

Avant de donner un chiffre concernant le coût national d'une telle sauvegarde systématique, il faudrait mener une enquête auprès de tous les établissements publics concernant leurs besoins de préservation à long terme des données, sans oublier en particulier les hôpitaux et l'imagerie médicale.

Il faudra ensuite décider dans quelle mesure il serait souhaitable de centraliser le stockage des données, voire d'aller vers un scénario comprenant un centre unique national de préservation des données numériques à très long terme. Un tel scénario permettrait d'une part d'arriver à des économies d'échelle, mais d'autre part présenterait tous les inconvénients bien connus d'une centralisation, en particulier en termes de légèreté de gestion et de réactivité des services. Le choix de la solution optimale est un choix politique qui dépasse les compétences du groupe PSN.

4.4 ■ *Conclusion*

Rapporté à l'échelle d'un pays comme la France qui comprend environ 25 millions de foyers, le coût annuel de préservation de l'information du patrimoine numérique est donc probablement compris entre 2 et 20 milliards d'euros par an (100 à 1 000 euros par foyer et par an), et 2 ou 3 fois moins dans une hypothèse minimale ne couvrant pas les enregistrements vidéo. Ce montant

est bien entendu une estimation macro-économique. Il peut être réparti entre de multiples intervenants, avec le risque néanmoins que la multiplicité des acteurs ne pénalise à la fois le coût global (par réduction des effets d'échelle) ainsi que la fiabilité de la préservation à long terme. Cette approche suppose également une consommation énergétique non négligeable, qui n'a pu être complètement étudiée dans le cadre du présent rapport[16], mais pose des problèmes environnementaux réels.

Le coût global de la stratégie active est donc très supérieur à ce que la grande majorité des acteurs imaginent aujourd'hui. S'il est vrai que les prix des supports numériques chutent constamment (l'exemple des disques durs est particulièrement frappant), il n'y a aucune raison d'attendre une diminution spectaculaire des coûts associés à la stratégie active, puisqu'ils sont principalement liés aux salaires et à la fourniture d'un environnement technique approprié. Cette stratégie s'imposera néanmoins si aucune technologie alternative en stockage passif n'émerge dans les années qui viennent. Le problème est que le temps nécessaire à l'émergence et au développement d'une telle technologie est égal ou supérieur au temps nécessaire pour que la piètre qualité des supports optiques actuels ne devienne cruellement évidente. Il est donc patent qu'un problème sérieux est posé, avec des implications économiques et sociétales notables, et que des décisions rapides seraient hautement souhaitables.

16. Voir par exemple dans *Le Monde* du 26 janvier 2009 « Les centres de données, de plus en plus gourmands en énergie », qui discute en particulier la question en termes de quantité de CO_2 produite et d'impact climatique.

This page intentionally left blank

Chapitre 5

La stratégie passive, les disques optiques numériques

Dans la perspective d'une stratégie de conservation passive, quels supports de l'information peuvent être utilisés ? A priori, nous l'avons vu, ce sont les disques optiques numériques enregistrables (DONE) qui semblent les mieux placés pour jouer le rôle de support pour une stratégie de type « archive et oublie ». Ce sont donc à eux que nous allons limiter la discussion à partir de maintenant.

5.1 ■ *État de l'art, avantages et inconvénients des disques optiques numériques enregistrables*

Les DONE, dans leur forme actuelle, sont loin d'être les objets simples, soumis à une évolution facilement prédictible, que l'on aurait pu espérer. Cette recherche de simplicité et de stabilité ne progresse guère : sous la pression du marché, les solutions retenues par les constructeurs privilégient avant tout la capacité et la

rapidité d'écriture, ainsi qu'un prix très bas ; la longévité n'est guère prise en compte.

Plusieurs constructeurs proposent ce qu'ils appellent des « disques d'archivage », parfois appelés « gold » parce qu'une couche d'or y remplace la couche métallique habituelle réfléchissante (aluminium) ; des produits à deux couches (or et argent) existent également. Si l'oxydation de la couche métallique était la source principale de vieillissement, il y aurait là une solution particulièrement intéressante. Mais, hélas, les études du Laboratoire national d'essais (LNE) ont montré que ce n'est pas le cas : dans la réalité, ces disques ne sont pas nécessairement meilleurs que les disques standard ; ils sont parfois même pires. Plusieurs rapports détaillés et fort instructifs du LNE montrent à quel point la situation est complexe et préoccupante[17].

La construction de disques optiques utilisant des colorants organiques photosensibles est un processus complexe et difficile à maîtriser. Les fabricants ne contrôlent pas la qualité d'ingrédients aussi essentiels que la solution de colorant, qui peut varier d'une livraison à l'autre ; sa composition chimique exacte est d'ailleurs maintenue confidentielle par la firme qui la produit. La fabrication des disques optiques numériques est donc inhomogène et peu reproductible[18], y compris pour un modèle donné d'un fabricant donné ; on constate d'importantes variations de la qualité, en fonction du lot et de la date de fabrication. Parmi la petite fraction des CD-R dont la durée de vie s'est révélée catastrophique (un an ou moins), il se trouve que beaucoup ont été fabriqués en 2003, sans que la raison précise de cette fluctuation soit connue. Une étude récente (2008) du LNE pour la DAF (Direction des Archives de France) montre que les modèles de CD-R qui étaient apparus

17. Voir en Annexe 4 la note synthétique de J. Perdereau (LNE), « Durée de vie des CD-R et DVD-R », octobre 2008.

18. J. M. Lambert et Y. Saunders (LNE) « La conservation des données sur CD-R », étude LNE de juillet 2004 montrant que, sur de 70 CD-R du commerce gravés en 2000, 15 % sont partiellement ou totalement illisibles.

Étude commune INA-LNE-INA « Étude portant sur la fiabilité du stockage de données sur CD-WORM », 2006 ; 8 % des CD-R sont hors normes au bout de 4 ans ; ceux achetés en 2003 étant les moins stables.

J. M. Lambert (LNE), « Étude du vieillissement en conditions climatiques sévères de deux références de CD-R », Fiche R et D du LNE, avril 2007.

comme les meilleurs dans l'étude précédente (2006) sont, deux ans plus tard, parmi les moins résistants[19]. Une autre étude du LNE[20] a montré que, dans les archives départementales numérisées sur CD-R, beaucoup de supports sont dégradés voire parfois illisibles, entraînant la perte de l'information qu'ils contenaient. Certes, le LNE et le NIST (National Institute of Standards) ont également observé que certains disques sont relativement bons, et semblent capables de conserver l'information pour 10 ou 20 ans, peut-être plus. On est alors dans un cas favorable où les DONE ont une meilleure longévité qu'un support magnétique ; mais il semble pour le moment très difficile de dégager les règles qui permettent de garantir à l'avance qu'une fabrication de disques sera bonne.

Il faut dire que la course vers une capacité de stockage toujours plus grande n'arrange pas les choses. La raison physique en est simple : les processus de dégradation d'un objet physique, par exemple par diffusion de molécules d'eau entre couches, sont d'autant plus rapides que l'échelle est plus petite (le temps caractéristique de diffusion varie comme le carré de la distance). Un objet contenant des inscriptions très denses et minuscules est nécessairement plus sensible à de tels processus qu'un objet de dimensions plus grandes. De plus, toujours dans l'optique d'augmenter la capacité, les disques Blu-ray stockent l'information très près de leur surface, protégés par une couche extrêmement mince (et d'ailleurs assez difficile à fabriquer). Par leur conception même, ces nouveaux supports devraient vieillir encore plus vite que les CD et DVD habituels.

Pour finir, l'impression qui se dégage est que la confusion entre « stockage » et « archivage à long terme », signalée dans

---

19. J. M. Lambert et J. Perdereau (LNE), « Qualité des CD-R disponibles sur le marché pour l'archivage des données numériques », rapport LNE de juillet 2008 ; « Qualité des DVD + R et DVD-R disponibles sur le marché pour l'archivage des données numériques », rapport LNE d'octobre 2008 soulignant la faible qualité des premiers DVD-R, qui s'est heureusement améliorée depuis.

20. Y. Saunders (LNE) « Qualité des disques optiques numériques conservés dans différents services d'archives », rapport final pour la Direction des Archives de France (décembre 2006) indiquant que « pour un disque sur trois, le risque de perdre des données est bien réel, et très élevé pour un sur cinq ».

Cette étude ainsi que quelques autres sont disponibles sur le site de la DAF :

http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage/

l'introduction de ce rapport, a bloqué l'émergence d'un bon DONE d'archivage. Nombreux sont ceux qui n'ont pas encore réalisé la différence entre les deux fonctions, et la nécessité de réaliser des DONE spécifique pour chacune d'entre elles. Ce n'est probablement pas en se contentant d'améliorer un peu la fabrication des disques de stockage actuels que l'on fera apparaître le disque d'archivage dont la société a besoin !

5.2 ■ *Processus physicochimiques mis en jeu*

On peut légitimement s'étonner d'un écart aussi grand entre les espoirs mis en jeu initialement dans le lancement des disques optiques numériques, supposés durer « éternellement », et ce que donnent les mesures de vieillissement artificiel, comme celles du LNE. C'est qu'il y a loin entre l'idée d'un processus d'enregistrement et de lecture presque idéal, réalisé par un faisceau lumineux sans aucun contact mécanique, ou encore le concept de « gravure » qui évoque une inscription ayant la solidité des monuments égyptiens, et la réalisation des DONE. En pratique, le processus d'inscription des données est complexe et met en jeu des mécanismes différents et mal connus (cf. Appendice 3).

Cette complexité se répercute immédiatement sur la difficulté d'étude des phénomènes de vieillissement. S'il s'agissait d'une réaction chimique unique et bien identifiée, on pourrait lui attribuer une énergie d'activation E_a, et prédire que la vitesse d'évolution correspondante varie comme l'exponentielle du rapport $(-E_a/k_BT)$, où T est la température absolue et k_B la constante de Boltzmann. On pourrait alors mesurer cette vitesse d'évolution à plusieurs températures différentes et en extraire E_a, ce qui permettrait ensuite de prédire la vitesse d'évolution à toute température. Cette méthode classique rendrait ainsi possible, à partir de mesures de vieillissement accéléré effectuées à hautes températures, d'en déduire par extrapolation la vitesse de vieillissement à température ordinaire, généralement beaucoup plus faible. C'est en gros le principe de la norme ISO dont il a été question au paragraphe 1.3, à ceci près qu'elle prend en compte deux paramètres,

température et humidité, et donc deux extrapolations exponentielles, au lieu d'une seule (elle ignore cependant bien d'autres mécanismes de vieillissement, comme ceux dus à la pollution des contaminants chimiques, l'illumination, etc.).

Quand on réalise que les extrapolations en question portent entre des temps de mesure de l'ordre du mois et des durées de vie calculées qui sont parfois des dizaines d'années, on se rend déjà compte combien la méthode est sujette à caution : de toutes petites erreurs de mesures se traduiront par de grandes barres d'erreur sur le résultat final. Mais il y a plus grave encore : si deux processus élémentaires de vieillissement agissent en parallèle, l'un de poids important associé avec une constante de temps longue, l'autre de faible poids mais de constante de temps courte, les mesures de vieillissement accéléré ne détecteront que le premier (à de petites erreurs près), puisqu'il est alors dominant. Mais, aux temps longs, la situation s'inverse, et c'est le second qui domine ; l'extrapolation vers les temps longs devient alors totalement fausse puisqu'elle ne prend en compte que celui qui est devenu négligeable – elle donc est bien trop optimiste.

Évoquer cette possibilité n'est pas académique, mais naturel au vu des résultats expérimentaux. En effet, si le processus de vieillissement des DONE était unique, les courbes mesurées en vieillissement artificiel seraient des fonctions du temps monotones simples, toutes semblables, à une homothétie près. Or il suffit de jeter un coup d'œil sur les figures du paragraphe 2 de l'Appendice 5 pour voir que c'est loin d'être le cas : une grande variété de comportements se manifeste. On a au contraire l'impression que pour certains DONE intervient un premier processus, pour d'autres un second, la cassure dans la courbe temporelle étant la signature claire d'un phénomène multiprocessus. Nombreuses sont les possibilités : blanchiment du colorant (par exemple à la lumière), diffusion des molécules d'eau dans la structure, oxydation ou décollement des couches métalliques, évolution en volume du polycarbonate, etc. Certains processus peuvent effectivement donner lieu à un comportement latent, comme on l'observe dans certains cas bien connus des chimistes, qui parlent alors de « processus d'initiation » ou de « phase d'induction »,

suivie ensuite d'une « phase rapide ». Par exemple, on peut imaginer que les défauts dans la matrice de polycarbonate diffusent thermiquement, se regroupent jusqu'à former des vides, qui à leur tour diffusent et induisent le vieillissement observé. De plus, les expériences de vieillissement artificiel mettent souvent en évidence l'apparition de défauts par taches localisées, un signe supplémentaire du caractère inhomogène du processus.

Pour finir, la seule chose vraiment certaine qui résulte des mesures de vieillissement artificiel est leur résultat brut : la résistance des DONE à des températures et humidités extrêmes. Il est évidemment logique d'utiliser ces données comme un indicateur de longévité, puisque personne ne peut pour le moment proposer de meilleur critère. On peut espérer que les DONE qui résistent le mieux en vieillissement artificiel sont également ceux qui retiendront le plus longtemps des données lisibles en conditions de stockage réel, mais, comme nous l'avons vu, même ceci n'est pas absolument certain. Ces mesures sont certes précieuses, mais il convient de conserver la plus grande prudence concernant les extrapolations et les prédictions de durées de vie en termes d'années, qui sont probablement fausses dans un grand facteur.

5.3 ■ *Variantes des disques optiques numériques*

Un espoir pourrait être offert par les disques holographiques. La motivation de leur développement semble être principalement l'augmentation de la capacité, ainsi que la possibilité d'une lecture parallèle de plusieurs faisceaux d'information. Mais, de plus, si on arrive à écrire l'information au cœur de l'objet et non plus tout près de sa surface, on peut raisonnablement espérer augmenter considérablement leur résistance à toutes les attaques physico-chimiques venant de l'extérieur. Il faut toutefois que le milieu lui-même d'enregistrement soit stable et ne se dégrade pas spontanément, par exemple à la lumière. Une compagnie américaine commercialise un produit de ce type (*In Phase Technology*). Les disques utilisent une couche de photopolymère sensible à la lumière, de sorte qu'ils doivent être placés dans une cartouche

d'un type adapté. Ils ne peuvent donc être lus que par un matériel spécifique vendu par la même entreprise ; l'ordre de grandeur du prix des lecteurs et des disques serait environ 100 fois celui du matériel utilisé par le grand public.

Mentionnons la commercialisation des disques UDO (*Ultra Density Optical disc*) par une firme britannique, Plasmon UK, qui a récemment déposé son bilan. Utilisant un média qui leur est propre mettant en œuvre un changement de phase (semblable aux CD-RW), cette firme garantissait une longévité de 50 ans. Hélas la lecture du support est impossible avec tout lecteur autre que celui vendu par Plasmon UK : on est alors prisonnier d'un seul fournisseur et, en cas de faillite, rien ne garantit plus la disponibilité de l'information à long terme.

Un nouveau projet, dénommé « Millenniata », est récemment apparu, avec la collaboration d'un certain nombre de membres de l'université de l'Utah[21]. Ses promoteurs annoncent une mise sur le marché avant la fin de l'année 2009 d'un DONE qui devrait durer mille ans (d'où le nom, combinaison de « millenium » et de « data »). Peu de renseignements techniques sont disponibles. D'après les quelques informations qu'a pu recueillir le groupe PSN, il s'agirait d'un DONE avec une couche sensible assez classique déposée sur un substrat de polycarbonate – donc pour finir assez proche par certains aspects des DVD-RW actuels. On peut s'interroger sur la longévité d'un disque en polycarbonate pendant un millénaire… Mais, que ce projet apporte ou non une solution technique réelle, il illustre bien un certain bouillonnement d'idées actuel autour du sujet ; de plus en plus d'acteurs commencent à percevoir l'importance et l'urgence du problème.

Une exception à ce tour d'horizon un peu sombre : le Century Disc, une invention française du milieu des années 80 lancée par la société Digipress, société ensuite revendue à diverses occasions. C'est un Century Disc que la NASA a mis dans une navette spatiale pour envoyer des informations sur notre civilisation terrestre d'autres civilisations lointaines ! Ces disques, en verre trempé, utilisent un procédé de lithogravure qui les rend très stables dans

21. Voir http://www.millenniata.com/index.html

le temps ; ils passent haut la main tous les tests de vieillissement du type de ceux que fait le LNE. Leur principal inconvénient est que leur fabrication n'est pas automatisée, mais manuelle. Le prix d'un Century Disc CD ou DVD est actuellement de l'ordre de 100 euros (cf. Appendice 2).

5.4 ■ *Quelques pistes vers un disque optique numérique enregistrable de bonne longévité*

Idéalement, les DONE stables du futur devraient être lisibles, voire même enregistrables, sur les mêmes appareils que les DONE ordinaires. La très large diffusion d'un standard de lecture est en effet une garantie concernant la disponibilité à long terme des lecteurs ; si des disques numériques sont lisibles par des appareils très courants, existant en très grand nombre dans le monde, la stabilité à long terme du système est assurée.

Cependant, demander une compatibilité totale sans aucun aménagement des appareils actuels est peut-être trop exiger. Le point essentiel est plutôt que la lecture du DONE du futur ne demande qu'une adaptation raisonnable des lecteurs disponibles. Après tout, les lecteurs de disques optiques numériques actuels sont déjà capables de lire un très grand nombre de formats et de s'adapter aux caractéristiques du support (réglage automatique de la puissance du laser par exemple). C'est nécessaire pour lire des disques allant du CD au BD, incluant la musique et de nombreux « codec » différents pour la vidéo. Un élargissement des capacités de lecture vers celle de disques de grande longévité n'est donc pas exclu ; mais une condition indispensable pour assurer la pérennité du système est que cela n'entraîne pas de changement radical du prix de l'appareil.

Dans l'Appendice 2, nous mentionnons des projets fort intéressants qui pourraient mener à un DONE avec une durée de vie satisfaisante, voire excellente ; nous nous limitons à quelques projets français, sans prétendre à l'exhaustivité. Pour le moment, ils résultent plus d'initiatives à relativement petite échelle que de grands projets structurés, que ce soit dans le secteur public

ou privé ; c'est pourquoi il est difficile d'évaluer à ce stade leur chance d'aboutir rapidement à un produit accessible au grand public. On ne peut cependant que regretter que des voies qui semblent prometteuses restent relativement peu explorées. Nous nous privons ainsi peut-être de progrès technologiques et de solutions innovantes au problème de l'archivage à long terme.

On peut également penser aux disques holographiques où l'information est écrite en volume. De cette façon, on s'affranchit de tous les problèmes liés aux diffusions et pollutions en surface. Comme de plus la technique holographique utilise un plus grand volume du disque et permet ainsi une capacité théorique très grande (3 à 4 téraoctets), on comprend l'intérêt qu'elle suscite. Mais la compatibilité avec les lecteurs existants est alors totalement perdue.

L'avenir des disques optiques numériques est incertain à long terme. Ils sont actuellement produits par milliards, mais sont menacés par les progrès spectaculaires des techniques concurrentes, en particulier du type mémoire flash. Ils ont largement perdu la course à la capacité face aux disques durs ; les prix par Go restent comparables, mais les taux de transfert sont moins élevés. Pour finir, au lieu d'une concurrence perdue d'avance avec les autres supports, leur principal espoir de survie semble être d'exploiter leurs spécificités en termes de simplicité du dispositif de stockage et de contrôle des processus. On peut espérer la mise sur le marché de produits dont on puisse garantir la longévité de façon certaine, un siècle ou plus ; même si la capacité reste limitée et si le prix est relativement élevé, le besoin d'archivage est tel qu'il y a là un important marché potentiel à exploiter. L'avenir dépendra donc en bonne partie de la capacité des fabricants à effectuer une reconversion par rapport aux tendances actuelles, et de réaliser une avancée significative dans le domaine de la longévité.

This page intentionally left blank

Conclusion et recommandations

Dans tous les pays, le besoin sociétal est le même, réel et pressant. L'Unesco a rédigé une « charte sur la conservation du patrimoine numérique » qui est reproduite en Appendice 1. Hélas, aucune réponse claire ne peut être apportée pour le moment, faute d'une bonne solution acceptable. Celle de la « préservation active » semble pour le moment la plus sûre, et c'est celle des quelques grands établissements spécialisés. Elle n'est cependant probablement pas transposable à l'échelle de tous les besoins, y compris familiaux et personnels.

Les disques optiques numériques enregistrables, par leur simplicité de principe, ne semblent pas loin de remplir le cahier des charges, mais il faudrait améliorer leur construction actuelle ; des études sont nécessaires pour cela. Pour le moment, ces études ne sont entreprises ni par le secteur privé (à notre connaissance) ni par les laboratoires publics, faute d'un programme permettant ces recherches. Elles seraient pourtant d'une grande utilité pour la société, mais la difficulté est qu'elles ne sont pas de la recherche, ni fondamentale, ni industrielle, puisque leur objectif n'est pas aujourd'hui celui des constructeurs. Espérons que ce

rapport les incitera à tenir compte davantage de l'impératif de l'archivage à long terme, et des possibilités de nouveaux marchés que pourraient créer des produits donnant des garanties solides sur ce plan.

L'industrie japonaise semble avoir pris conscience du problème. En témoigne la création toute récente (août 2008) d'un centre indépendant de tests des disques optiques numériques enregistrables (DONE), le « *Archive Disc Test Center* », dans le cadre du projet Bifröstec. Depuis début 2009 ce centre dispose d'un laboratoire et de 5 enceintes de vieillissement accéléré (température et humidité) ainsi que de testeurs donnant accès aux paramètres des disques optiques enregistrés. Les fournisseurs pourront, moyennant finances, soumettre des lots de 100 DONE à ce centre indépendant, qui mesurera leurs propriétés sous vieillissement accéléré. Si les résultats sont satisfaisants, le centre attribuera un label, concrétisé par un logo « *Optical Disc Archive Time Test* » qui indiquera aux acheteurs que les disques correspondants ont bien résisté aux tests. Il est clair que cette initiative japonaise est très intéressante, et qu'il serait de l'intérêt des pays européens de s'y associer dès que possible. Le problème intéresse de nombreux pays[22]. La France, en particulier avec l'expérience acquise par les travaux du LNE, est bien placée pour jouer un rôle dans ce domaine.

---

22. À titre d'illustration, nous reproduisons un extrait du site de l'université de Montréal (http://blogues.ebsi.umontreal.ca/stra_num/) :

« *L'information numérique est devenue essentielle à presque tout ce que les Canadiens font dans leur vie quotidienne* […] *Une partie de notre patrimoine numérique a déjà disparu et une grande partie est en péril.* »

Voici quelques phrases extraites du site de la Bibliothèque et Archives du Canada : (http://www.collectionscanada.gc.ca/scin/012033-1005-f.html) :

« *Au cours des prochaines années, tous les Canadiens devraient vraisemblablement être touchés par les problèmes de conservation de l'information. La fragilité des données numériques et la rapide obsolescence de leurs supports menacent les précieuses collections personnelles et familiales d'œuvres musicales, de DVD et de photographies.*

Pourtant, peu de Canadiens mesurent l'ampleur du problème. Aussi est-il important de sensibiliser les créateurs, les représentants des gouvernements, les organes de financement et le grand public. Il faut notamment attirer leur attention sur les risques de perte d'information numérique et ses conséquences, de même que préciser les stratégies et les mesures à prendre pour éviter une telle perte. Nous aurons besoin de leaders dans ce domaine. Il faudra également fournir des exemples percutants de perte ou de risques de perte d'information. »

Recommandations

Un message d'alerte au grand public, aux établissements et à l'administration, doit être lancé. Le problème n'est pas spécifique à la France, mais mondial. Une action au niveau européen, ou dans le cadre de l'Unesco, semblerait souhaitable, afin de faire prendre conscience de l'urgence d'une politique concertée dans ce domaine. À l'échelon national et européen, nous proposons quatre recommandations :

1. **Débloquer les études sur le sujet.** Engager rapidement une étude réellement scientifique des phénomènes de vieillissement des supports, notamment des supports optiques, visant à dégager des recommandations fiables en matière de standardisation de formats de supports d'archivage longue durée. Cette étude serait de préférence à mener en liaison avec les autres travaux similaires dans le monde, en particulier au Japon et aux États-Unis.
   Les laboratoires et équipes qui font des propositions de recherche dans ce domaine doivent pouvoir trouver un soutien financier ; il faut impérativement éviter de voir leurs projets rejetés par les agences de moyens du fait qu'ils n'entrent dans aucune catégorie prévue, ce qui conduit à leur abandon. Lancer rapidement un appel à projets ambitieux visant à remplacer la technologie d'enregistrement optique actuelle (CDR et DVDR), basée pour le moment sur des processus physicochimiques complexes et mal contrôlés, par des technologies plus robustes et prévisibles.
2. **Éviter la perte des compétences dans le privé et le public.** Recenser les compétences publiques et privées dans le domaine et prendre les mesures urgentes nécessaires à la préservation des compétences clés, avant qu'elles aient complètement disparu de l'Europe. Des actions conservatoires seront très certainement nécessaires avant la fin de l'année 2010, compte tenu des menaces pesant sur les dernières équipes de R & D industrielle qui sont compétentes dans le domaine.

3. **Favoriser l'innovation et l'apparition d'une offre industrielle de qualité**. Sur la base des résultats des études et projets identifiés ci-dessus, favoriser l'émergence d'une nouvelle offre technologique et industrielle. Soutenir vigoureusement les quelques entreprises qui ont déjà effectué des avancées vers la réalisation de disques optiques numériques enregistrables de très bonne longévité. Ce soutien peut prendre la forme de contrats ou de commandes de dispositifs réellement innovants par les organismes publics concernés par l'archivage à long terme.
4. **Élaborer une véritable politique d'archivage numérique.** S'assurer au sein de chaque ministère que les données numériques importantes (documents officiels, examens médicaux conservés dans les hôpitaux, etc.) sont bien l'objet du suivi indispensable à leur survie. Évaluer l'intérêt d'une mutualisation des moyens, dans la perspective d'une stratégie active à l'échelon national, ou de la création d'un centre de conservation des données numériques à long terme équipé de robots permettant le suivi nécessaire à grande échelle (mettre dans la balance les économies d'échelle avec les inconvénients habituels d'une centralisation).

Ces initiatives doivent être soutenues par des financements publics, notamment par la France, l'Allemagne (aux niveaux fédéral et Länder) et les Pays-Bas, premiers pays concernés par la localisation des compétences clés, ainsi que par l'Union européenne.

LIFE-EXPECTANCY FOR DIGITAL DATA

How long can digital data be stored?

A commissioned report by the "PSN" Academic Working Party on Life Expectancy for Digital Supports. PSN-WP has Members common to both the Academy of Sciences and the National Academy of Technologies of France (NATF).

Members of the PSN-WP :

Erich SPITZ, Chairman,
Academy of Sciences and National Academy of Technologies of France (NATF)

Jean-Charles HOURCADE,
National Academy of Technologies of France (NATF)

Franck LALOË,
LKB/École Nationale Supérieure,
Chairman of the Advisory Scientific Council of GIS-DON

This page intentionally left blank

Abstract

Why should we pay any attention to long-term preservation of digital data, at a time when digital storage capacities have never been so abundant or so cheap ? The simple answer is that we must never mistake two distinct notions, **data storage** and backup on one hand and **archiving** on the other. The spectacular progress of hard-disk devices, as well as the drop in their market prices, make it an easy affair to store information data. They are generally duplicated several times to counter the risk of unforeseeable system failures, such as a hard disk "crash". However, archiving in this way over several decades or even centuries raises a totally different issue than backup actions. The reason is that the material used for the digital support usually has a life expectancy of roughly 5 to 10 years. As and when a hard disk is approaching its life-end and thereby presenting a risk of total data loss, it becomes necessary to recopy the data onto a new support. Bearing in mind that the ageing of support materials is difficult to foresee, only a constant follow-up of data quality enables you to guarantee long range archiving, but, alas, with a high associated organisation cost.

Recordable optical digital disks (RODDs) have lesser capacities but are sometimes believed to offer better life expectancies. This is implicit in the way we commonly say that we "engrave" data on a recordable disk. The very notion of "engraving" alludes to the age-old inscriptions that were engraved in stone and marble and discovered later on seemingly timeless antiques. Unfortunately, the resulting feeling of security is misleading : no recording support material or process currently on the market can guarantee a high degree of conservation of the data recorded beyond 5 to 10 years, approximately.

Our societies today generate ever-increasing amounts of information, while, at the same time, the expected operational life of the supports available has never been so short. While this question is addressed satisfactorily by major specialised institutions, it is almost totally ignored by the public at large as well as by the majority of public and private institutions, and business sector companies. There is a real danger that huge amounts of so-called personal data (medical, scientific, technical, administrative, etc. information) could simply disappear in years to come.

The Academic PSN-Working Party was set up in Autumn 2008 by the Academy of Sciences and the National Academy of Technologies of France (NATF), after both Academies had taken note of this worrisome situation and decided to issue a status report on the subject.

This report has an intentionally limited scope : indeed, this is a necessity given the numerous digressions possible. The report thus focuses on that part of the information world that conserves a long-term intrinsic value, whether it be in the form of personal documents, family souvenirs, medical data, etc., or public information (*e.g.,* scientific data acquired from single, non-repeatable experiments, etc.). The report discusses several possible strategies : "archive and forget", sometimes called the passive strategy, certainly the most natural method ; the active strategy relying on perpetual migration ; delegated contract archiving to a service provider ; the possible return to analogue recording.

The various storage support materials are reviewed (recordable optical digital disks (RODDs), magnetic tapes, hard-disk drive mechanisms, flash memories, etc.). with a brief discussion of their respective assets and weak points. The fourth part of the report assesses possible generalisation of an active strategy to meet the total requirements of Society, whether in terms of personal documents, those issued by public authorities, or those possessed by public or private enterprise. The final section of the report takes a closer look at recordable optical disks, for which a series of alarming measurements have been carried out recently. These studies underscore the 'non fundamental' characteristics of the problems to be solved : if the processes leading to ageing of recordable optical disks are not under control, the reasons lie more in the priorities adopted to ensure access to the market place rather than to essential reasons, *viz,* data preservation issues. The report suggests certain approaches that could lead to RODDs with notably better life expectancies.

A set of four recommendations is issued by the PSN-Working Party at the end of the report.

This page intentionally left blank

Conclusion and recommendations

On a world-wide sale, the societal need is almost identical, real and urgent. UNESCO adopted and proclaimed a "Charter on the Preservation of the Digital Heritage", reproduced below, in Appendix 1. Alas, no clear solution to the archiving problem is currently available, for lack of a decisive and thereby widely acceptable method. At this point in time, the approach known as "active preservation" seems apparently to be the safest ; it is effectively used by several major specialist institutions. However, it probably cannot be extended to meet all requirements of Society, inclusive of family and personal data archiving.

Recordable optical disks (RODDs), given their intrinsic simplicity, apparently are close to meeting the specification requirements, but nonetheless require an improvement in their present construction process, *i.e.,* further studies are needed to achieve this aim. Such studies are not carried out in the private sector (to the best of our knowledge), nor by public laboratories, for lack of official research programmes that would permit such research. The latter would, however, prove very useful to Society. The difficulty is that such studies are considered as being neither

basic nor industrial research, since their object does not coincide with the aims of manufacturers. The authors of this report entertain the hope that the industrial sector will be encouraged to develop new media that are adequate for long term data preservation. There are possible new market segments that could lead to creation of successful new products.

The Japanese industrial sector seems to have become aware of the reality of the problem. Witness the quite recent (August 2008) establishment of an independent test centre for digital optical disk evaluation, the "Archive Disc Test Center", in the framework of the Bifröstec Project. Since early 2009, this research establishment offers laboratory facilities (including 5 devices for accelerated ageing with controllable temperature and humidity) as well as testing equipment to monitor the quality of data of RODDs. The disk suppliers, for a contract price, can provide batches of 100 RODDs to this independent centre, which then will carry out property variations under accelerated ageing processes. If the results prove satisfactory, the centre issues a logo "Optical Disc Archive Time Test". This serves to inform future purchasers of disks that they have indeed satisfactorily complied with the testing requisites. Clearly, this Japanese initiative is very attractive and it would be in the interests of European countries to become partners as soon as possible. The problem indeed is of interest to numerous countries. In particular France, with its experience at the LNE national testing standards laboratory, could advantageously play a role in this area.

Recommendations

A warning should be issued to the public at large and to our national administrations and establishments. The problem is not restricted to France ; it is a world-wide issue. Action at a European level or within the scope of UNESCO seems advisable, if only to ensure that our politicians realise that it is now urgent to adopt a concerted policy in these matters. At both national and European levels, the authors make 4 recommendations :

1. **Launch and implement research in this area**. Rapidly undertake a truly scientific study of support ageing phenomena, notably optical support materials. The purpose is to issue reliable recommendations for standards applicable to long-range archival support materials and processes. Such research should preferably be conducted jointly with other similar projects round the world, in particular in Japan and in the USA.

   Laboratories and research teams that are in a position to make proposals for research work in this area should be able to obtain appropriate financial funding ; it is primordial that their proposals are not rejected by the national funding agencies, only on the grounds that they do not fit into any of the agency's pre-set categories. It would be relevant, in this light, to launch an ambitious programme aimed at the replacement of present processes for optical recording (CD-Rs and DVD-Rs) – that use complex physicochemical processes that are not fully mastered – by more rugged and foreseeable technologies.
2. **Avoid loss of relevant skills and know-how in both public and private sectors.** Identify and list all public and private skills relevant to the field and take urgent steps to safeguard key skills and associate know-how, before they disappear completely from the European scene. Conservatory action will certainly be needed before end 2010, given the financial threats facing the last few industrial European R & D teams that have skills in this field.
3. **Encourage innovation and the advent of a high quality industrial offer.** Based on the results of the research outlined above, efforts should be deployed to favour the emergence of a new industrial and technological offer. Vigorous support should be forthcoming for those few companies that have already made progress towards mass fabrication of guaranteed long-life expectancy digital optical disks. Support here could come as contracts or orders placed by public authorities with vested interests in implementing long-term archival policies, for truly innovative devices in this field.

4. **Implement a real digital archival policy.** One should ascertain, within each Government department or ministry, that all important digital data (official documents, medical records held by hospitals, etc.) are subject to a follow-through inspection process, which is the only way to ensure long-range survival. Assessment should be made of the possibilities for sharing preservation structures, with the prospect of implementing an active national scale strategy. One possibility is the creation of a specific centre for very long-term preservation of digital data, which centre would be equipped with robots appropriately programmed to ensure large scale inspection and follow-through data conservation processes (balancing economy of scale and the usual draw-backs of all centralised processes).

Such initiatives should be supported by public financing, especially in the cases of France, Germany (at both federal and Länder levels) and the Netherlands, that together represent the three countries most concerned by localisation of key skills and know-how mentioned above, and of course by the European Union.

Appendices

Nous incluons à la suite de ce rapport quelques appendices, soit afin de préciser certains points trop détaillés pour le texte principal (Appendice 2, ou Appendice 5 sur les travaux du LNE), soit pour illustrer le propos (Appendices 3 et 4 visualisant le principe de fonctionnement d'un DONE ou montrant des photos de disques endommagés par leur vieillissement).

This page intentionally left blank

Appendice 1

Charte de l'Unesco sur la conservation du patrimoine numérique (15 octobre 2003)

La Conférence générale,

Considérant que la disparition du patrimoine, quelle qu'en soit la forme, constitue un appauvrissement du patrimoine de toutes les nations,

Rappelant que l'Acte constitutif de l'Unesco stipule que l'Organisation aide au maintien, à l'avancement et à la diffusion du savoir en veillant à la conservation et protection du patrimoine universel de livres, d'œuvres d'art et d'autres monuments d'intérêt historique et scientifique, que son programme « Information pour tous » offre une plate-forme aux débats sur les politiques à adopter et les mesures à prendre en ce qui concerne l'information et la sauvegarde du savoir enregistré et que son programme « Mémoire du monde » vise à assurer la conservation du patrimoine documentaire mondial et son accessibilité universelle,

Reconnaissant que ces ressources en matière d'information et d'expression créatrice sont de plus en plus produites, diffusées, obtenues et conservées sous forme numérique, créant ainsi un nouvel héritage – le patrimoine numérique,

Consciente que l'accès à ce patrimoine offrira de plus larges possibilités de création, de communication et de partage des connaissances entre tous les peuples,

Comprenant que ce patrimoine numérique risque de disparaître et que sa conservation dans l'intérêt des générations présentes et futures est une question urgente qui intéresse le monde entier,

Proclame les principes ci-après et **adopte** la présente Charte.

Le patrimoine en tant que patrimoine commun

Article premier – Champ d'application

Le patrimoine numérique se compose de ressources uniques dans les domaines de la connaissance et de l'expression humaine, qu'elles soient d'ordre culturel, éducatif, scientifique et administratif ou qu'elles contiennent des informations techniques, juridiques, médicales ou d'autres sortes, créées numériquement ou converties sous forme numérique à partir de ressources analogiques existantes. Lorsque des ressources sont « d'origine numérique », c'est qu'elles existent uniquement sous leur forme numérique initiale.

Les documents numériques comprennent, parmi un large éventail de formats électroniques qui ne cesse de se diversifier, des textes, des bases de données, des images fixes et animées, des documents sonores et graphiques, des logiciels et des pages Web. Ils sont souvent éphémères, et leur conservation nécessite des mesures volontaires d'entretien et de gestion dès leur création.

Beaucoup de ces documents ont une valeur et une importance durables et constituent par conséquent un patrimoine qui doit être protégé et conservé pour les générations présentes et futures. Ce patrimoine, qui ne cesse de grandir, peut exister dans n'importe quelle langue, n'importe quelle partie du monde et n'importe quel domaine de la connaissance ou de l'expression humaine.

Article 2 – Accès au patrimoine numérique

La conservation du patrimoine numérique a pour but de faire en sorte qu'il demeure accessible au public. Il s'ensuit que l'accès aux documents du patrimoine numérique, en particulier ceux qui sont dans le domaine public, ne doit pas faire l'objet de restrictions excessives. En même temps, les informations sensibles et personnelles doivent être protégées contre toute forme d'intrusion.

Les États membres pourraient vouloir coopérer avec les organisations et institutions compétentes pour favoriser l'instauration d'un environnement juridique et pratique qui maximise l'accessibilité du patrimoine numérique. Il convient de réaffirmer le principe d'un juste équilibre entre les droits légitimes des créateurs et des autres titulaires de droits et les intérêts du public touchant l'accès aux documents du patrimoine numérique et d'en faciliter la réalisation, conformément aux normes et accords internationaux.

Protection contre la perte de patrimoine

Article 3 – La menace de perte

Le patrimoine numérique mondial risque d'être perdu pour la postérité. Les facteurs qui peuvent contribuer à sa perte sont l'obsolescence rapide du matériel et des logiciels qui servent à le créer, les incertitudes concernant les financements, la responsabilité et les méthodes de la maintenance et de la conservation et l'absence de législation favorable à sa préservation.

L'évolution des attitudes n'a pas suivi celle des technologies. L'évolution numérique a été trop rapide et trop coûteuse pour que les pouvoirs publics et les institutions élaborent en temps voulu et en connaissance de cause des stratégies de conservation. La menace qui plane sur le potentiel économique, social, intellectuel et culturel du patrimoine, pierre angulaire de l'avenir, n'a pas été pleinement saisie.

Article 4 – Nécessité d'agir

Si rien n'est fait contre les menaces actuelles, la perte du patrimoine numérique sera rapide et inéluctable. Il est dans l'intérêt des États membres d'encourager des mesures juridiques, économiques et techniques visant à sauvegarder ce patrimoine. Une campagne d'information et de sensibilisation s'impose d'urgence pour alerter les décideurs et le grand public en leur faisant prendre conscience aussi bien du potentiel des supports numériques que des problèmes pratiques de conservation.

Article 5 – Pérennité de l'information numérique

La pérennité du patrimoine numérique est fondamentale. Pour le conserver, il faudra prendre des mesures pendant toute la durée de vie de l'information, du moment où elle est créée à celui où l'on y a accès. La conservation à long terme du patrimoine numérique commence avec la conception de procédures et de systèmes fiables qui produisent des objets numériques authentiques et stables.

Mesures requises

Article 6 – Élaborer des stratégies et des politiques

Des stratégies et des politiques doivent être élaborées pour protéger le patrimoine numérique en tenant compte du degré d'urgence, de la situation locale, des moyens disponibles et des prévisions d'avenir. Ce sera plus facile si les créateurs, titulaires du droit d'auteur, et les détenteurs de droits voisins et autres parties

prenantes travaillent en coopération à la définition de normes communes compatibles et qu'ils partagent les ressources.

Article 7 – Sélectionner ce qu'il convient de conserver

Comme pour tout type de patrimoine documentaire, les principes de sélection peuvent varier d'un pays à l'autre, même si les principaux critères appliqués pour décider des documents à conserver doivent être leur importance ou leur valeur culturelle, scientifique, de preuve ou autre, sur la durée. Il est évident que les documents « d'origine numérique » doivent avoir la priorité. Les choix opérés et tout réexamen ultérieur doivent pouvoir être justifiés et reposer sur des principes, politiques, procédures et normes bien définis.

Article 8 – Protéger le patrimoine numérique

Les États membres ont besoin de cadres juridiques et institutionnels appropriés pour assurer la protection de leur patrimoine numérique.

Élément clé de la politique nationale de conservation, la législation en matière d'archives et de dépôt légal ou volontaire dans des bibliothèques, archives, musées et autres dépôts publics doit être étendue au patrimoine numérique.

L'accès aux documents du patrimoine numérique en dépôt légal doit être assuré, dans les limites de restrictions raisonnables, sans que cela nuise à leur exploitation normale.

Les cadres juridiques et techniques protégeant l'authenticité sont indispensables pour éviter la manipulation ou l'altération volontaire du patrimoine numérique. Ils exigent que le contenu, la fonctionnalité des fichiers et la documentation soient conservés dans la mesure nécessaire pour garantir l'authenticité des documents.

Article 9 – Préserver le patrimoine culturel

Le patrimoine numérique n'a, par essence, aucune limite temporelle, géographique, culturelle ou formelle. Il est propre à une culture, mais virtuellement accessible à tout un chacun dans le monde. Les minorités peuvent s'adresser aux majorités, les particuliers à un auditoire mondial.

Le patrimoine numérique de tous les pays, régions et communautés doit être conservé et rendu accessible pour donner au fil du temps une image équilibrée et équitable de tous les peuples, nations, cultures et langues.

Responsabilités

Article 10 – Rôles et responsabilités

Les États membres peuvent juger bon de confier à un ou plusieurs organismes la responsabilité de coordonner la conservation du patrimoine numérique, en mettant à leur disposition les ressources nécessaires. Le partage des tâches et des responsabilités peut se faire en fonction de l'expertise et des rôles existants.

Des mesures doivent être prises pour :

(a) engager les concepteurs de matériel et de logiciels, les créateurs, éditeurs, producteurs et distributeurs de documents numériques ainsi que les autres partenaires du secteur privé à coopérer avec les bibliothèques nationales, archives, musées et autres organisations chargées du patrimoine public en vue de conserver le patrimoine numérique ;

(b) développer la formation et la recherche et veiller au partage des expériences et des connaissances entre les institutions et associations professionnelles concernées ;

(c) encourager les universités et autres établissements de recherche, tant publics que privés, à assurer la conservation des données issues de la recherche.

Article 11 – Partenariats et coopération

La conservation du patrimoine numérique exige des efforts soutenus de la part des gouvernements, des créateurs, des éditeurs, des industries du secteur et des institutions chargées du patrimoine.

Vu la fracture numérique actuelle, il est nécessaire de renforcer la coopération et la solidarité internationales pour permettre à tous les pays d'assurer la création, la diffusion et la conservation de leur patrimoine numérique ainsi que la possibilité d'y accéder en permanence.

Les industries, les éditeurs et les médias sont vivement encouragés à promouvoir et partager les connaissances et les compétences techniques.

Favoriser les programmes d'enseignement et de formation, les accords en matière de partage des ressources et la diffusion des résultats de la recherche et des meilleures pratiques démocratisera l'accès aux techniques de conservation numérique.

Article 12 – Le rôle de l'Unesco

Il incombe à l'UNESCO, en vertu de son mandat et de ses fonctions :

(a) de prendre en considération les principes énoncés dans la présente Charte dans le déroulement de ses programmes et d'en promouvoir l'application au sein du système des Nations unies et par les organisations intergouvernementales et non gouvernementales internationales qui s'occupent de la conservation du patrimoine numérique ;

(b) de servir de point de référence et d'offrir aux États membres, aux organisations intergouvernementales et non gouvernementales internationales, à la société civile et au secteur privé une enceinte où ils puissent unir leurs efforts pour élaborer des objectifs, des politiques et des projets en faveur de la conservation du patrimoine numérique ;

(c) de favoriser la coopération, la sensibilisation et le renforcement des capacités et de proposer des principes directeurs éthiques, juridiques et techniques normalisés, pour étayer la conservation du patrimoine numérique ;

(d) de déterminer, à partir de l'expérience tirée durant les six prochaines années de l'application de la présente Charte et des principes directeurs, s'il est nécessaire d'adopter d'autres instruments normatifs en vue de la mise en valeur et de la conservation du patrimoine numérique.

Appendice 2

Quelques projets français

– Un projet respectant bien la compatibilité des lecteurs a été monté récemment par le CEA-LETI (Grenoble), en s'appuyant sur des compétences de plusieurs laboratoires comme le Service Chimie de Rennes, le Département de Photochimie Générale de Mulhouse, et le CIRIMAT de Toulouse. Il s'agit d'utiliser des matériaux permettant au maximum de remplacer les matériaux organiques, donc fragiles, par des matériaux inorganiques. Lorsque le disque comprend à la fois les deux types de matériaux, on parle de disque « hybride ». Le projet régional « Phaistos » a déjà permis au LETI d'obtenir un certain nombre de résultats concernant un substrat hybride à pistes structurées par nano-impression photochimique de sol-gels hybrides organique/inorganique, ainsi que la réalisation d'un empilement inorganique avec couche enregistrable de type chalcogénure (à changement de phase cristalline). L'équipe en charge du projet au LETI a envisagé de continuer ses travaux en collaboration avec le GIS-DON (groupe d'intérêt scientifique sur les disques optiques numériques) et les laboratoires

cités précédemment, avec trois objectifs : substrat structuré de pistes (polycarbonate remplacé par des matériaux inorganiques ou hybrides) ; développement d'empilements enregistrables complètement inorganiques basés sur des matériaux du type chalcogénures thermiquement stabilisés, ou des matériaux thermo-oxydables de type oxydes de fer spinelles ; élaboration d'une couche de protection améliorée. Malheureusement, ce projet intéressant n'a pas trouvé de financement en 2008, et a du être abandonné.

– Le Century Disc offre des perspectives prometteuses, même s'il doit encore évoluer pour remplir le rôle d'un support d'archivage passif selon le principe « archive et oublie » accessible financièrement aux particuliers. Il s'agit de disques en verre trempé dans lesquels l'information est gravée par un procédé de lithogravure, donc directement dans le verre, pour assurer une excellente stabilité dans le temps. Sous sa forme actuelle, il inclut une couche métallique qui le rend totalement compatible avec les lecteurs ordinaires. On peut également penser à une version encore plus stable, et sans couche métallique ; il faudrait alors adapter l'optique et l'électronique des lecteurs pour traiter des signaux avec des contrastes plus faibles, mais cela ne semble pas infaisable.

Deux voies semblent se présenter pour le futur du Century Disc en tant qu'outil d'archivage pour les particuliers :

a) la plus ambitieuse, celle de la mise au point de graveurs individuels, comme ceux de CD et DVD actuels, permettant à chaque personne de graver chez elle des Century Discs. C'est probablement un programme de plusieurs années au moins.
b) celle de la mise au point de machine de recopie vers un Century Disc, manœuvrée par des professionnels. On peut alors imaginer un modèle commercial semblable à celui des photographes de quartier d'autrefois : chacun pourrait apporter ses données sur des DVD-R ordinaires, et commander la recopie sur des disques optiques numériques de

très bonne longévité, qu'il reprendrait sur place quelques jours plus tard.

La voie a) serait évidemment idéale, mais elle semble encore hors de portée pour au moins quelques années, sauf découverte inattendue. En revanche, de bons espoirs existent de rendre accessible la voie b), et d'arriver à des machines qui permettent, pour quelques dizaines d'euros, la recopie des données d'un DVD vers un support très stable. Pour le moment, le prix des Century Disc reste encore environ 10 fois trop élevé pour être accessible aux particuliers.

This page intentionally left blank

Appendice 3

Schéma du processus d'enregistrement d'un disque optique numérique

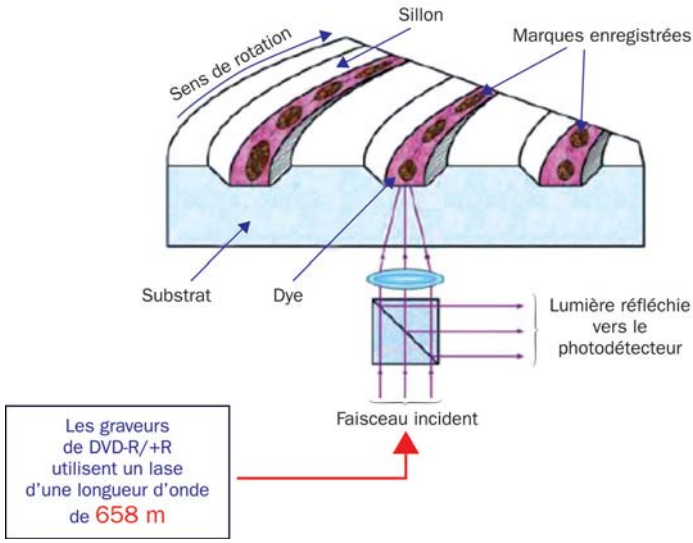
(Images fournies par J. J. Wanègue)

Dans un CD-R, on trouve en plans superposés dans le disque :

– un substrat en polycarbonate (épaisseur 1,2 mm), dans lequel est imprimé un sillon en spirale ; il forme la piste que va suivre le faisceau laser pendant la rotation du disque ;
– une couche très mince de colorant constitué de molécules photosensibles, qui remplit le sillon. C'est la « couche sensible » ;
– un film métallique réfléchissant, généralement en aluminium (ou parfois en or) ;
– enfin tout au dessus une couche d'apprêt permettant d'écrire ou d'imprimer sur sa surface.

Le laser vient du dessous, traversant toute l'épaisseur du polycarbonate pour atteindre la couche sensible. Grâce à un asservissement mécanique et opto-électronique, son point de focalisation suit le sillon. Lors d'une opération de gravure, le laser émet de façon intermittente, laissant des marques de longueurs variables

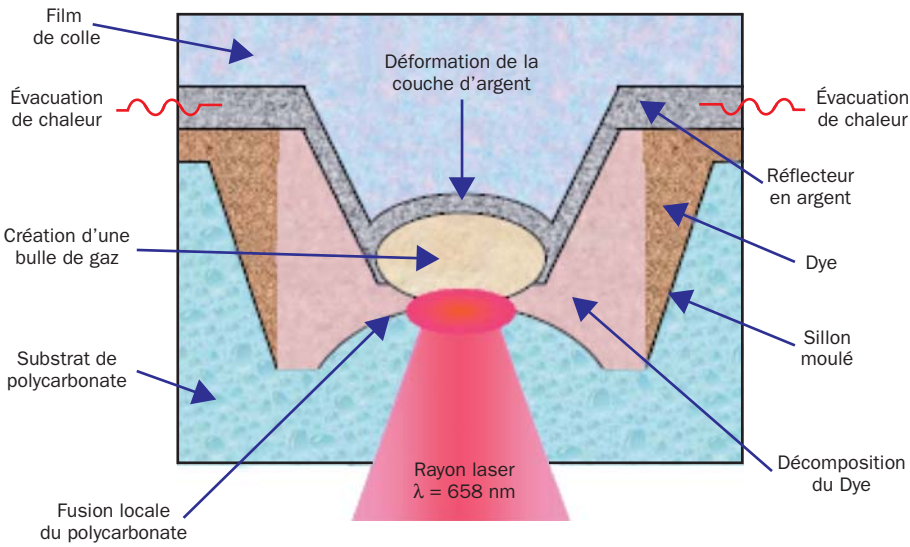
qui codent l'information binaire. Les marques proviennent de ce que le colorant absorbe la puissance lumineuse émise par le laser, ce qui déclenche toute une série de phénomènes : blanchiment du colorant par décomposition chimique (qui le rend plus transparent et change son indice de réfraction) ; destruction brutale d'une partie du colorant sous l'effet de l'apport d'énergie, qui résulte dans la formation d'une bulle autour du point d'impact ; par contrecoup, à la fois la matrice du polycarbonate et la couche métallique sont déformées. C'est une petite explosion locale, aux effets relativement mal contrôlés ; le poids relatif des divers processus de formation des marques dépend de la vitesse d'écriture, mais aussi d'autres facteurs.



Structure générale d'un DVD enregistrable : DVD-R ou DVD+R.

Lors de la lecture, le laser émet une puissance beaucoup plus faible que lors de l'écriture, de sorte que la gravure devient impossible. Le but du dispositif est alors de détecter les modifications des propriétés optiques du faisceau lumineux renvoyé par le disque sous l'effet des marques. Les changements de la transparence et de l'indice de réfraction du colorant, ceux du substrat, ainsi que la déformation de la couche métallique, se conjuguent pour créer toute une série d'effets qui modifient ce faisceau lumineux de

retour : absorption, diffraction optique, interférences optiques, ou encore diffusion lumineuse par les irrégularités. La juxtaposition de ces processus, dont l'importance relative varie selon la fabrication et la façon dont l'enregistrement a été fait, explique pourquoi il semble très difficile de prédire la longévité des CD-R. Des systèmes élaborés de traitement du signal sont nécessaires pour en extraire l'information utile, notamment à partir de ses variations dans le temps (dérivation) lors des transitions entre marques et absences de marques. Le sillon contient également d'autres informations utiles, en particulier par des oscillations latérales qui sont détectées et utilisées pour une synchronisation temporelle.



Représentation schématique de la formation des marques (pits) lors de l'enregistrement d'un DVD-R (ou d'un CD-R).

Le CD-RW n'utilise pas de colorant, mais un matériau sensible du type chalcogénure (composé contenant un ion négatif chalcogène, O, S, Se ou Te). L'effet du laser est d'induire un changement de phase qui se répercute sur une modification des propriétés optiques : à l'état vierge, le matériau est cristallin et présente une bonne réflectivité mais, sous l'effet du laser, passe vers un état amorphe où cette réflectivité devient plus faible. Jusqu'à maintenant, les mesures de longévité des CD-RW ont conduit à des

résultats moins bons qu'avec les CD-R, mais il n'est pas impossible qu'une autre conception des CD-RW permette de faire mieux, car ils s'affranchissent des problèmes liés à la stabilité chimique des colorants.

Le DVD-R est semblable au CD-R dans son principe, mais sa couche sensible à colorant se trouve en sandwich entre deux galettes de polycarbonate d'une épaisseur de 0,6 mm. Il faut utiliser une couche de colle pour assembler ces deux galettes, ce qui peut constituer un point faible en termes de longévité. L'utilisation de marques plus petites et d'un pas plus resserré du sillon que dans le CD permet d'obtenir une capacité de stockage 10 fois supérieure environ.

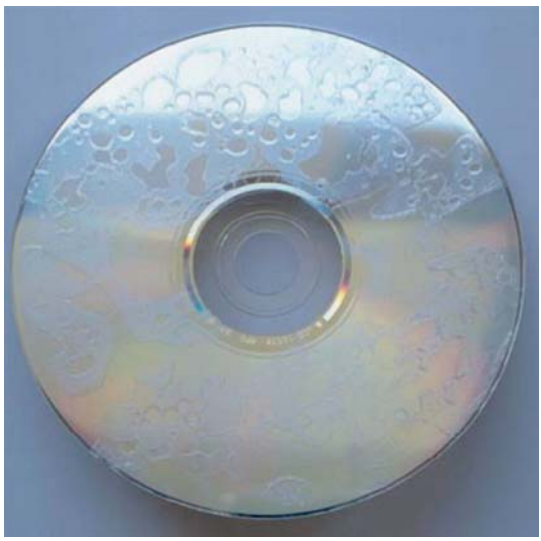
Le HD-DVD enregistrable était un peu une transposition du DVD-R, avec une couche sensible protégée au milieu de l'épaisseur du disque, mais il a été abandonné. Tout le marché du disque dit BD appartient maintenant au Blu-ray qui, lui aussi, utilise un laser de longueur d'onde plus courte que le DVD, ce qui permet d'augmenter la capacité de stockage (25 Go environ en simple couche). Comme pour le CD, la couche sensible du Blu-ray est située près d'une surface, mais cette fois le laser l'illumine directement sans passer à travers toute l'épaisseur du polycarbonate, ne traversant qu'une fine couche optique de 0,1 mm. Ce dispositif est évidemment fragile, mais il est recouvert d'une couche de protection avec « hard coat » qui, selon les fabricants, résiste extrêmement bien aux rayures et aux salissures.

De façon très générale, la prédiction de la longévité à long terme d'objets n'est possible que s'il s'agit de dispositifs relativement simples, dont le principe et le fonctionnement sont bien compris. De plus, pour un principe d'enregistrement donné, chaque fois que l'on augmente la capacité, on est conduit à réduire l'échelle des bits d'informations. Or, plus un objet est petit, plus il est sensible au vieillissement. Par exemple, pour un vieillissement dû à un processus de diffusion moléculaire, la constante de temps de ces derniers varie comme l'inverse du carré des dimensions. Il existe donc une certaine exclusion mutuelle entre les notions de capacité et de longévité, au moins tant qu'on s'en tient à un principe de fonctionnement donné.

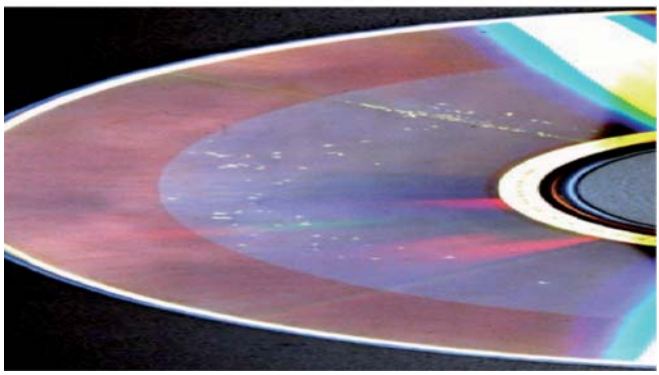
Appendice 4

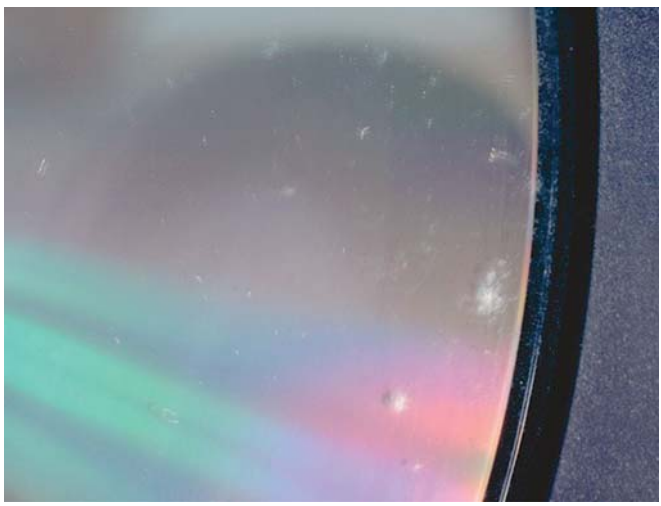
Quelques images illustrant le vieillissement de disques optiques enregistrables

(Documents fournis par J. M. Fontaine, de l'Institut Jean le Rond d'Alembert)



Cloques





Moisissures

Exposition à la lumière

This page intentionally left blank

Appendice 5

Une synthèse du LNE faite à l'occasion de l'audition au groupe PSN (octobre 2008)

DURÉE DE VIE DES CD-R ET DVD-R
(Jacques PERDEREAU – LNE)

Le CD audio est apparu dans les années 1980. Ces CD ont montré leur bonne tenue dans le temps. Les quelques problèmes rencontrés étaient souvent liés à des défauts de fabrication concernant le vernis et l'étiquette imprimée par sérigraphie.

Cette fiabilité reconnue du CD non enregistrable a servi de justification à l'utilisation du disque optique enregistrable pour l'archivage des données. Le CD-R a été utilisé pour stocker les données à partir de 1991 et le DVD-R à partir de 1998. Vers la fin des années 90 le CD-R a commencé à être massivement utilisé pour l'archivage et le DVD-R à partir de 2004-2005.

Actuellement la production annuelle de CD-R et DVD-R est de l'ordre de 15 milliards de disques. Le disque optique est aujourd'hui le type de support enregistrable le plus utilisé.

1. *Le CD-R*

En 2000-2003, le principal problème de l'archivage sur CD-R était le problème de l'adaptation du graveur vis-à-vis du disque (vitesse d'écriture et stratégie d'écriture optimisée ou non). Seule la sensibilité à la lumière de la couche enregistrable (*dye*) en cyanine est citée dans quelques articles et en particulier par TDK qui vantait son nouveau *dye* « *Advanced cyanine* » plus stable que la cyanine ordinaire. La phtalocyanine, autre *dye* utilisé pour les CD-R, est réputée, quant à elle, très stable.

Les normes ISO 18921 (CD - 2000) 18927 (CD enregistrable - 2002), proposent des « méthodes d'estimation de l'espérance de vie » basées sur l'étude de la durée de vie des disques dans des conditions de températures et d'humidités élevées et sur l'utilisation du model d'Eyring. En se servant de ces normes, les fabricants annoncent des durées de vie qui vont de 50 à 300 ans.

En fonction de toutes ces données, le CD-R est le support d'archivage unanimement recommandé. Il est en effet réputé nettement plus fiable que les supports magnétiques (disques, bandes).

2004 : Étude LNE sur l'état de conservation de 60 références de CD-R

Le LNE a entrepris en septembre 2003 de vérifier l'état de conservation d'une collection de CD-R qui avaient été gravés entre juillet 2000 et décembre 2002. Les caractéristiques de l'échantillon sont les suivantes :

- 69 CD-R (60 références) gravés entre 2000 et 2002 ;
- Dye : Cyanine 44 %, Phtalocyanine 45 %, AZO 11 % ;
- Conservation : sans lumière, 20 à 23 °C, hygrométrie faible (40 à 50 %).

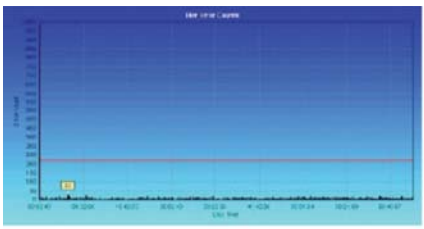
Les analyses de qualité initiale (taux d'erreurs, paramètres optiques) ont été réalisées juste après la gravure. En septembre 2003, après un contrôle de qualité des 69 CD-R, le résultat fut surprenant.

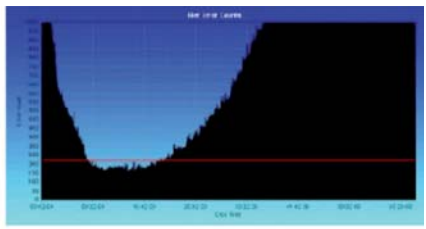
En effet, certains disques, dont la gravure initiale était pourtant de bonne qualité, se sont révélés fortement dégradés, avec des taux d'erreurs dépassant largement les limites admissibles pour garantir la lisibilité. Le dépassement de ces limites a été atteint dans certains cas pour une durée de conservation inférieure à 1 an.

Les graphes suivants montrent les taux d'erreurs (en fonction de leur localisation sur le disque) de 2 références de CD-R pour l'analyse initiale et celle de septembre 2003.

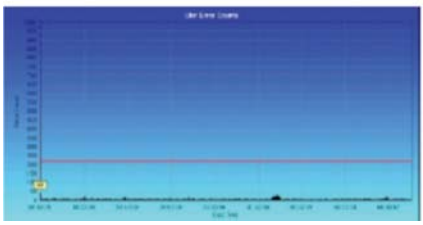
Mesure initiale **Mesure en septembre 2003**

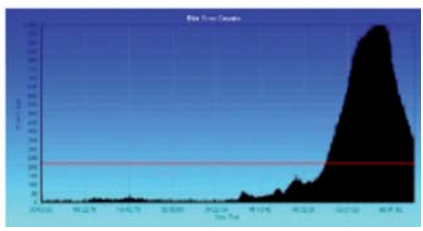
Disque avec une forte dégradation sur sa totalité (réf. 25-D)





Disque avec une dégradation sur une large zone (réf. 26-D)





Les résultats principaux de cette étude sur les CD-R âgés de 3 ans au plus sont les suivants :

- 15 % des CD-R sont partiellement ou totalement illisibles ;
- certains après moins d'un an de conservation ;
- les dyes concernés : en majorité cyanine et peu de phtalocyanine.

Généralement les disques ne sont pas dégradés uniformément sur toute leur surface, et les zones les plus dégradées peuvent être situées aussi bien près du centre, au milieu ou près du bord extérieur disque.

2005 – Étude de la collection de l'INA

Cette collection comprend 160 000 CD-R qui ont été gravés de 1994 à 2003. L'échantillon pour l'évaluation est constitué de 1500 CD-R également réparti sur les 10 années.

Les principaux enseignements que l'on peut tirer de cette étude sont :

- 8 % des CD-R sont en dehors des limites de lisibilité données par l'Orange Book ;
- les CD-R les plus anciens (1994) font partie des CD-R qui présentent le meilleur état de conservation ;
- les CD-R de 2003 sont ceux qui présentent le plus de défauts. Nous avons pu mettre en évidence l'évolution d'un type de CD-R sur un intervalle de seulement 6 mois.

2005/2006 – Étude d'échantillons de CD-R de 23 services départementaux des archives

Cette étude menée pour la Direction des Archives de France avait pour objectif de faire un sondage sur l'état de conservation des stocks de CD-R de 23 services d'archives. L'étude est réalisée à partir d'un échantillon de 254 disques prélévés.

Caractéristiques de l'échantillon :

- gravure entre 1999 à 2005 ;
- 7 références de CD-R ;
- conditions de gravure et de stockage variables.

Les résultats sont inquiétants :

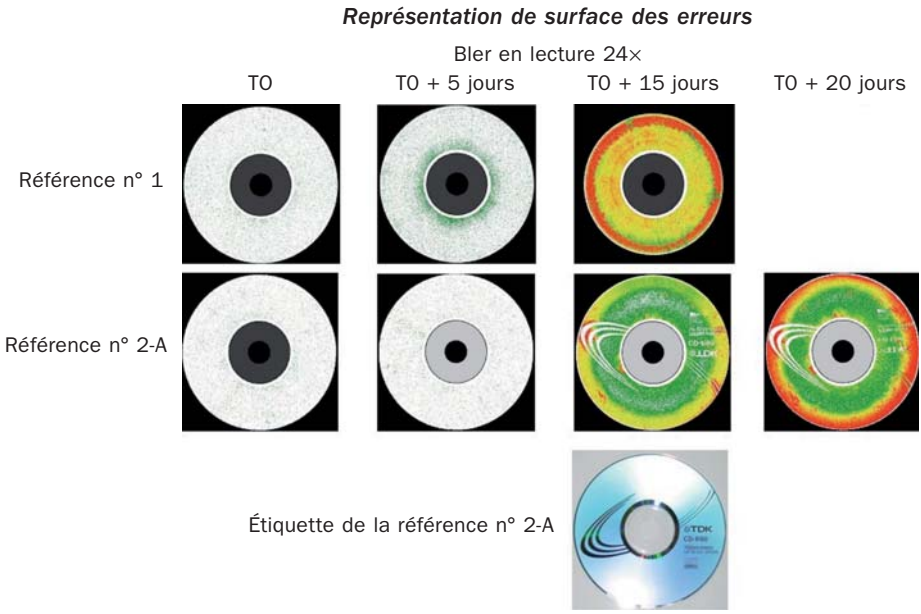
- 20 % des CD-R au-delà des limites dont 13 % difficilement récupérable sans perte de données ;
- certains services ont une majorité de disques au-delà des limites.

Conclusion sur l'analyse des collections

La dégradation, dans le temps, des propriétés de certains CD-R a été mise en évidence dans ces 3 études. De plus, des analyses de CD-R d'autres collections ont confirmé ces résultats.

Il existe une très grande disparité de comportement entre les références de CD-R, puisque les durées de vie observées peuvent aller de 1 an à plus de 10 ans. Ceci va à l'encontre des affirmations des fabricants.

Il n'existe pas d'explication sur les causes de ces dégradations données par les fabricants ou dans des publications. Nous avons observé fréquemment que la dégradation n'est pas homogène du centre vers le bord du disque et que les types de dégradations sont différents d'une référence de disque à l'autre (exemple entre Verbatim et TDK ci-après). Il y a à l'évidence plusieurs phénomènes qui peuvent concourir à la dégradation des CD-R.



Évolution des taux d'erreurs de 2 disques en fonction du temps d'exposition à 80 °C et 85 % HR.

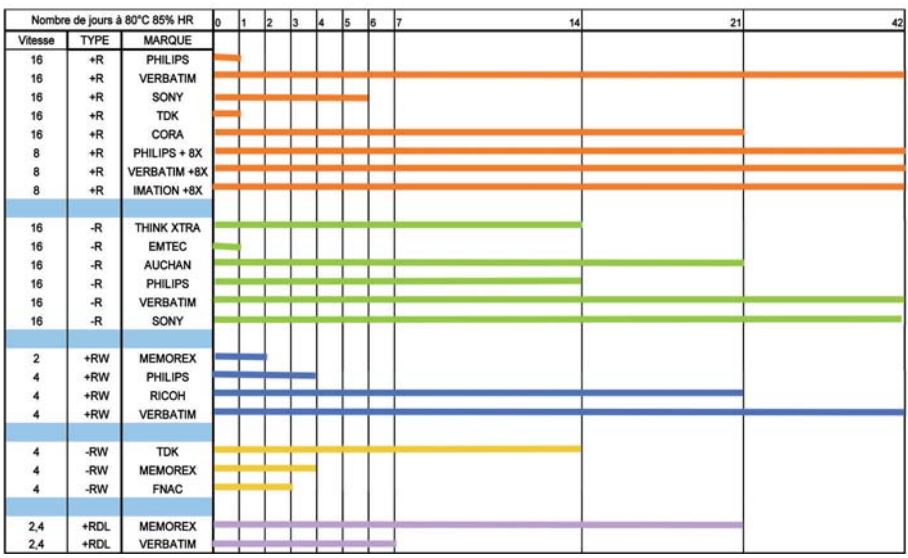
2. *Le DVDR utilisé comme support d'archivage*

Le DVD-R est moins utilisé par les institutions que le CD-R en raison de ses défauts de jeunesse aussi bien au niveau du collage des 2 disques que de la difficulté de mise au point du *dye* en fonction des différentes vitesses d'écriture de 1X à 16X.

En terme de durée de vie, les fabricants annoncent plutôt une duré de 50 ans que de 100 ans. Nous avons très peu de données expérimentales si ce n'est la durée de vie de 6 mois pour un *dye* qui était utilisé pour les DVD-R en 2004/2005.

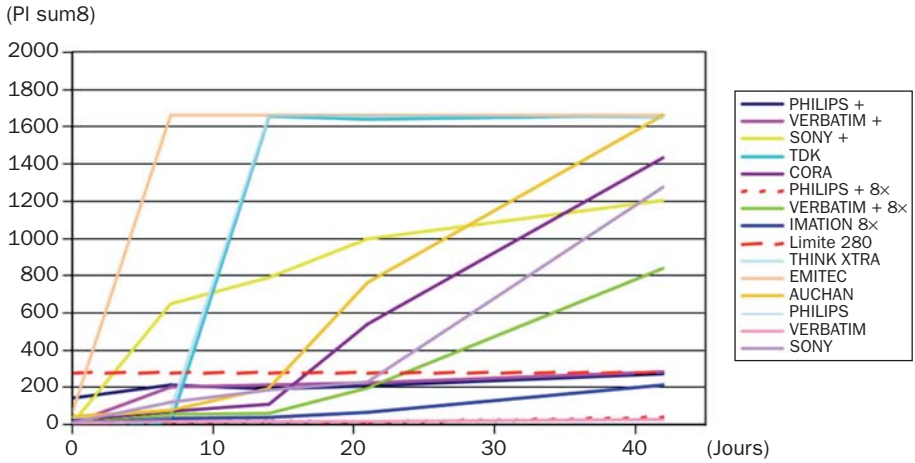
Les seules expérimentations de durée de vie réalisées aussi bien par le NIST et que par le LNE sont faites dans des conditions de températures et d'humidités élevées (60 à 85 °C et 70 à 85 % HR).

Comme pour le CD-R, les résultats montrent une très grande disparité de comportements puisque les durées de vie peuvent varier dans un rapport de 1 à 40 comme le montre les résultats de l'étude sur les DVD faite pour l'INC en 2006. Le graphique ci-après représente la durée de vie (en jours) des différents DVDR dans des conditions de température de 80 °C et d'humidité relative (HR) 85 %.



Durée de vie (en jours) de différents DVD-R à 80 °C et 85 % HR.

De plus la variation des taux d'erreurs en fonction du temps d'exposition à température et humidité élevées n'est pas forcément régulière et identique d'un disque à l'autre, comme le montre le graphe ci-après.



DVD-R : évolution du taux d'erreur Plsum8 en fonction du temps.
Conditions d'exposition (80 °C – 85 % HR).

3. *Conclusions du LNE sur l'utilisation des CD-R et DVD-R pour l'archivage*

Les CD-R et les DVDR ont des durées de vie très différentes selon les marques et même pour les différents modèles d'une même marque. Il a été constaté que certains modèles de disque étaient fortement dégradés au bout d'un 1 an et que d'autres n'avaient pas évolué au bout de 10 ans. Il en est de même pour les autres supports d'enregistrement des données comme les disques durs.

Les CD-R et les DVDR peuvent être utilisés pour l'archivage des données mais il faut choisir les meilleurs et mettre en place un suivi dans le temps des archives.

Il n'existe pas de méthodes fiables pour déterminer la durée de vie. En effet, les causes des dégradations n'étant pas identifiées, il

est illusoire d'utiliser les normes d'estimation de durée de vie par extrapolation pour choisir les disques les plus stables.

Le GISDON se propose d'une part d'étudier les causes des dégradations et d'établir des procédures de qualification des disques pour un label « Archivage » et d'autre part de concevoir un support de type Blu-ray dédié à l'archivage. Depuis 3 ans les partenaires du GISDON n'ont pas pu avancer dans ces recherches en l'absence de financement. Par manque de procédures de qualification étayées, nous nous contentons de sélectionner les disques qui résistent le mieux dans des conditions de température de 60 ou 80 °C et avec 85 % d'humidité relative.

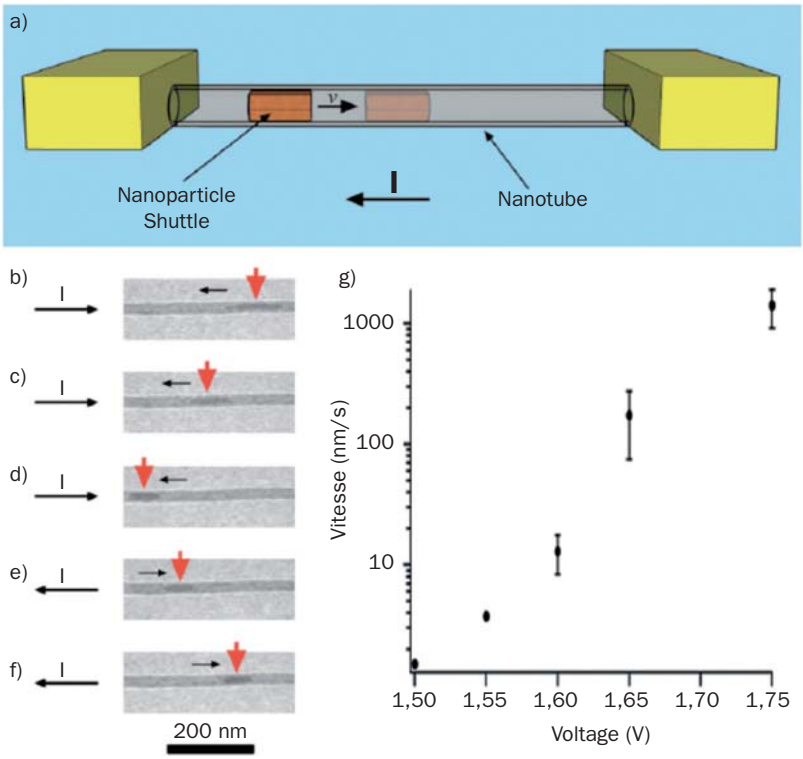
Il convient de remarquer que l'on ne doit pas se fier uniquement à la marque même si certaines marques se révèlent en moyenne meilleures que d'autres. On peut, par exemple, citer le cas de VERBATIM qui a produit des CD-R de bonne stabilité pour les dyes de 8X à 24X, un CD-R se dégradant en 1 an (*dye* 32X), à nouveau un CD-R stable pour 40X et pour finir un moyennement stable pour le 52X. Des exemples similaires peuvent exister pour les marques MPO, RITEK, KODAK, TDK, FUJI, SONY…

On peut se demander si les fabricants maîtrisent bien les phénomènes de vieillissement des disques, car dans les deux dernières études réalisées pour la Direction des Archives de France, les disques spécifiques « Archivage » se sont révélés parfois être les moins stables.

Appendice 6

Mémoires à nanotubes de carbone

Ce projet est basé sur l'utilisation d'un appareil électromécanique, dans lequel une masselotte faite d'une nanoparticule de fer se déplace dans un nanotube de carbone creux et multiparois, et peut être poussée dans une direction ou une autre par l'action d'un courant électrique (électromigration). La lecture s'effectue par des petites impulsions électriques qui ne perturbent pas la position de la masselotte. Il se trouve que la résistance électrique du système dépend de cette position, par un mécanisme qui n'est pas clairement élucidé pour le moment.



Mémoire à nanotubes de carbone.

Système utilisant une nanomasselotte comme mémoire. La figure a) est une vue d'artiste représentant la nanomasselote encapsulée dans un nanotube de carbone, à l'intérieur duquel elle se déplace comme le ferait une navette. Lorsqu'un courant de quelques centaines de microampères passe dans le tube, la nano masselotte se déplace dans le sens opposé du courant. Les figures b) à f) sont des images TEM de la masselotte se déplaçant d'environ 200 nm (les contacts sont hors du champ sur la droite et sur la gauche). Dans les figures b) à d), le courant va vers la droite et la masselotte vers la gauche ; dans les figures e) à f), le courant est inversé, et la masselotte revient à sa place originale. Elle constitue ainsi une sorte de navette qui peut être déplacée sur toute la longueur du nanotube d'une façon réversible et reproductible, et servir à enregistrer des données numériques. La figure g) montre l'effet d'un ajustement du courant sur la vitesse de la particule.

Appendice 7

Quelques idées reçues

On entend souvent dire que la conservation à long terme des données est une question de formats de fichiers, et qu'elle implique de conserver des lecteurs et des ordinateurs de toutes les générations. On cite couramment à l'appui de ce point de vue l'évolution des formats MS Word et les problèmes d'incompatibilité associés. Mais il y aussi tous les logiciels de CAO, de composition graphique, d'architecture avec représentation 3 D, ou plus prosaïquement de comptabilité. Un exemple particulièrement critique est fourni par les jeux : si l'on se donnait comme but de conserver éternellement les jeux vidéo, il faudrait conserver non seulement les logiciels, mais aussi les systèmes d'exploitation ainsi que les consoles elles-mêmes.

Mais il est irréaliste de vouloir conserver toute l'information sous tous les formats possibles, alors que de nouveaux formats apparaissent constamment ! Ce qui est nécessaire est que chacun sache, s'il veut conserver à long terme une information sélectionnée, quel logiciel et quel format utiliser, et sur quel support l'inscrire. Un format bien construit comme Tex/Latex restera probablement lisible dans 100 ans, ne subissant aucune pression

commerciale pour lui imposer des évolutions constantes ; de plus il est utilisé par une grande communauté scientifique internationale. Un autre exemple de format dont l'espérance de vie est grande est PDF/A.

De toute façon, si dans quelques siècles on sait que tel ou tel support ou de type de fichier contient une information importante, on saura toujours arriver à la lire ! Champollion a fait plus difficile. Pensons également aux platines pour disques vinyl qui réapparaissent actuellement dans le commerce, du fait que ces disques ont une grande durée de vie.

On entend parfois dire que la préservation à long terme du numérique est avant tout une affaire de métadonnées : il n'est utile de préserver une information que si on sait à quoi elle correspond. Les métadonnées sont une condition nécessaire à une bonne préservation, mais cela ne veut pas dire qu'elles sont suffisantes ! À quoi serviraient des métadonnées parfaites sur une information qui s'est effacée ? De plus, ajouter a posteriori des métadonnées est souvent possible, et ne présente pas plus de difficulté technique que la préservation des données elles-mêmes. Les métadonnées ne sont donc pas le cœur du problème.

On entend également souvent affirmer que la question centrale est une bonne adéquation du « couple graveur-CD », comme si le choix d'un graveur adéquat permettait de s'assurer d'une bonne longévité. En réalité, les processus de gravure et de vieillissement des DONE sont indépendants. Il tombe sous le sens que, pour une bonne longévité, il est préférable de partir d'une bonne gravure, avec peu d'erreurs ; mais ce n'est pas le choix d'un bon graveur qui ralentira les processus de vieillissement ultérieurs. Là non plus n'est pas le cœur du problème.

Une autre idée courante depuis quelques années est que le réseau Internet permet la délocalisation et l'abstraction des données, sur des ordinateurs répartis dans le monde entier, et que cela résout le problème de leur préservation à long terme. Nous l'avons déjà vu au paragraphe 2.3 : prendre Internet pour un réservoir de stockage serait aussi naïf que confondre, en plomberie, les tuyaux et les réservoirs. Au bout du compte, le réservoir doit exister quelque part dans le monde, ou les risques de fuite

sont les mêmes. Ce n'est pas en éloignant le problème à l'autre bout du monde, ou en le diluant entre de très nombreux ordinateurs, qu'on construit un système plus robuste à l'échelon de quelques décennies. Cette technique semble mieux adaptée pour la conservation à moyen terme, quelques années par exemple, et pour les données qui ne sont pas trop sensibles.

This page intentionally left blank

Liste des auditions du groupe PSN

(Les comptes rendus de ces auditions sont disponibles)

25 septembre 2008, Monsieur F. Le Carvennec, expert indépendant

2 octobre 2008, Messieurs D. Uhde et H. Richter (Thomson Allemagne)

16 octobre 2008, Messieurs J. M. Rodes et F. Berahou, INA

24 octobre 2008, Monsieur Jacques Perdereau, LNE

5 novembre 2008, Messieurs J.-M. Fontaine (LAM, Institut Jean Le Rond d'Alembert) et J. J. Wanègue, expert indépendant

14 novembre : Monsieur P. André, Plasmon France (Caen)

17 novembre 2008 : Messieurs J. Raby et A. Fargeix, LETI (Grenoble)

17 décembre 2008 : Monsieur L. de Poix, entreprise MPO

19 décembre 2008 : Messieurs A. Beaufort, T. Ledoux et L. Duplouy (BNF)

21 janvier 2009 : Madame F. Banat-Berger et Monsieur M. Jacobson (Direction des Archives de France)

28 janvier 2009 : Madame M. Campana et Monsieur J. M. Besse (ministère de l'Industrie)

9 mars 2009 : Monsieur S. Deleonibus, CEA-LETI (Grenoble)

9 mars 2009 : Monsieur T. Ihashi (Bifrostec, Japon)

9 mars 2009 : Messieurs J. P. Gleyzes, M. Buégué, et Madame D. Boucon (CNES)

11 mars 2009 : Madame G. Pinson, Messieurs J. B. Henniart et C. Girard, Secrétariat d'État au Développement de la Région Capitale

2 avril 2009 : Monsieur L. Ranno, Département nanosciences, Institut Néel, UJF, Grenoble

20 mai 2009 : Messieurs F. Daumas, directeur du CINES (Montpellier) ; M. Auffret et O. Rouchon, département archivage et diffusion du CINES

25 juin 2009 : Monsieur Christian Amatore, membre de l'Académie des sciences

25 juin 2009 : Monsieur Gilles Lachaud, directeur de l'Institut de Mathématiques de Luminy

29 juin 2009 : Messieurs T. Ihashi (Bifrostec) et A. Inoué (Mitsubishi Chemicals)

22 octobre 2009 : Madame E. Dion et Monsieur D. Reizine (APHP, radiologie)