

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



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 1^{er} 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.

Vj ku' r ci g' l' p v g p v k q p c m { ' i g h v ' d i e p m

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

Vj ku' r ci g' l' p v g p v k q p c m { ' i g h v ' d i e p m

Introduction

Le sujet du présent rapport est la préservation à long terme de l'information numérique. Chacun sait que cette information est produite journallement 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