

# Communiqué du GIS-DON du 29 août 2008

---

## Pour un disque optique numérique de confiance

La conservation à long terme des données numériques se heurte à trois types de problèmes :

1. Le rapide changement des matériels, que ce soit les média de stockage ou des lecteurs, ainsi que celui des formats d'écriture sur ces médias. L'exemple le plus connu est celui des disquettes souples, qui ont été d'usage très courant, et qu'il devient de plus en plus difficile de lire; mais il existe bien d'autres cas d'obsolescence (format DAT, etc.).
2. Le changement des systèmes d'exploitation des ordinateurs, des logiciels, et des formats de fichiers. L'exemple le plus connu est le changement des formats de fichiers MS-Word, mais il en existe bien d'autres. Certains formats de fichiers semblent en revanche relativement stables, comme Tex/Latex ou Pdf /A.
3. La détérioration des supports physiques eux-mêmes sur lesquels sont écrites les informations numériques. Chacun sait que les disques durs sont susceptibles de « crash » ; il convient de faire des sauvegardes préventives pour ne pas risquer de perdre toute l'information en un instant. Mais peu réalisent que les supports numériques se dégradent constamment, même si on ne les utilise pas. Ce sont des objets physiques fragiles, qu'ils soient magnétiques (bandes, disques durs, etc.) ou optiques (CD-R, DVD-R, etc.) ; de fait, peu sont capables de conserver l'information plus d'une dizaine d'années, parfois moins.

Parmi ces trois problèmes, c'est le dernier qui contient la clé des deux autres : si nous disposions de supports stables pour stocker l'information, cela créerait automatiquement un marché pour des logiciels et des matériels durables permettant leur relecture ; mais, tant que la base du système elle-même est instable, il n'y a aucune raison d'espérer voir émerger par miracle cette durabilité. Pourquoi les lecteurs de disques noirs reviennent-ils à la mode en ce moment ? c'est bien parce que ces disques sont des supports stables de l'information<sup>1</sup> et qu'ils ont été capables de passer plusieurs décennies sans se détériorer.

Paradoxalement, la plupart de nos concitoyens sont conscients des deux premiers problèmes - probablement parce que beaucoup d'entre nous les ont déjà rencontrés dans la vie courante - mais pas du troisième. Dans la mesure où il s'agit d'une dégradation constante et progressive, le phénomène est moins visible, donc plus sournois. Ainsi il ne suffit pas de choisir des formats aussi courants que ceux des CD, dont on peut espérer que le nombre extraordinaire d'exemplaires<sup>2</sup> dans le monde garantira la longévité ; on se heurte néanmoins au fait que les disques optiques enregistrables<sup>3</sup> s'effacent lentement, mais constamment et de façon aléatoire, même si on ne les utilise pas.

Voilà qui va à l'encontre des idées reçues : a priori, les disques optiques numériques enregistrables semblent la solution idéale au problème de la conservation numérique, puisque leur lecture se fait par un procédé qui évite toute usure mécanique (lecture par faisceau laser). Mais ils sont sensibles à des processus de dégradation physique et chimique interne qui rend difficile de garantir leur longévité. Si certains restent encore utilisables au bout de quinze ans, d'autres deviennent illisibles en deux ans, parfois moins, sans qu'il soit aisé de prévoir à l'avance lesquels ; on ne peut plus ensuite les recopier sur un CD neuf.

---

<sup>1</sup> Cet exemple est emprunté au monde de l'analogique, pas du numérique, mais peu importe : il s'agit d'illustrer que la stabilité des supports entraîne un maintien des techniques de lecture.

<sup>2</sup> Des milliards sont fabriqués chaque année !

<sup>3</sup> Nous ne parlons ici que des disques optiques enregistrables, donc utilisables par chacun pour stocker ses informations numériques. Les CD pressés en série, vendus tout enregistrés, ont des durées de vie bien supérieures.

La notion de recopie est en effet essentielle dans le domaine de la préservation numérique. Car, heureusement, l'information numérique possède une propriété miraculeuse : on peut la recopier sans erreur<sup>4</sup>, grâce aux les « codes de correction » qui ont été inventés pour cela par les mathématiciens. Si l'on recopie un CD-R sur un support neuf, ces codes sont mêmes capables de corriger les erreurs qui sont apparues depuis la gravure sur le CD initial ! ...pourvu toutefois qu'elles ne soient pas trop nombreuses. Donc si l'on effectue cette recopie à temps, par exemple tous les trois ans, il n'y a pas de limite dans le temps à la conservation de l'information.

Ainsi, si en théorie l'information numérique peut durer très longtemps, en pratique il faut s'en occuper constamment ; celle dont personne ne s'occupe meurt rapidement, en quelques années. Quel que soit le support matériel utilisé, bande magnétique, disque dur, disque optique, mémoire flash, etc. on ne peut pas enregistrer des documents numériques, puis les laisser au fond d'une armoire ou les cacher dans un coffre-fort pendant 30 ans, espérant les transmettre ainsi aux générations suivantes ; il ne resterait rien d'utilisable.

On entend souvent objecter que le problème est maintenant résolu grâce à l'existence d'Internet, qui permet facilement d'effectuer des copies multiples et distribuées un peu partout dans le monde. Il est vrai que cette délocalisation de l'information est une bonne garantie contre les catastrophes, allant du simple crash de disque dur au vol et à l'incendie. Mais penser que cela assure la pérennité des documents à long terme est illusoire, car il ne faut pas confondre les incidents, accidents ou catastrophes à court terme, avec l'érosion naturelle du temps qui dégrade constamment les supports. Certes, Internet fournit des « tuyaux » pour l'information, mais il ne faut pas confondre les tuyaux et les réservoirs.

En eux-mêmes les réseaux n'apportent donc aucune nouveauté pour le stockage de l'information, ils ne permettent que sa distribution. Chacun comprend d'ailleurs aisément que ce n'est pas une garantie de préservation à long terme que de déplacer le problème à l'autre bout du monde, où l'on ne peut rien voir des procédures de sauvegarde mises en jeu. De plus, pour des raisons de confidentialité, les entreprises ne peuvent pas prendre le risque de délocaliser toute leur information interne, même cryptée, chacun sachant que bien des codes de cryptage peuvent être cassés. Bref, délocaliser l'information en espérant que cela la rende éternelle est un peu croire en la politique de l'autruche.

Une solution serait de créer dans chaque administration, chaque hôpital, chaque entreprise, etc. des services spécialisés uniquement dans la sauvegarde et surtout la constante migration des données, y compris la recopie sur des supports neufs. Leur rôle serait de constamment suivre l'évolution des supports, et de les recopier à temps. Pour les particuliers, ils pourraient s'adresser à des sociétés spécialisées. Pourquoi pas ? Mais, à l'échelon national, un calcul simple montre qu'il faudrait une bonne dizaine de milliers de « curateurs numériques » pour faire ce travail sans fin de fourni, voire plus. On imagine facilement les dépenses correspondantes, en salaires et en matériel !

Puisque c'est techniquement possible, n'est-il pas bien plus simple et raisonnable de fabriquer des disques optiques numériques basés sur des principes physiques sains, de sorte qu'ils vivent une centaine d'années, et que chacun puisse les conserver chez lui ?

Franck Laloë, Président du conseil scientifique

---

<sup>4</sup> Plus précisément, les taux d'erreurs sont tellement faibles qu'ils autorisent des millions (ou plus) de recopies avant d'induire une probabilité appréciable d'erreur.