

L'ESSENTIEL

- La pérennité des données numériques est un problème peu étudié par l'industrie et les organismes publics.
- L'altération des supports est la difficulté principale pour l'archivage. Les dispositifs les plus courants ont des durées de vie courtes.
- La méthode d'archivage durable la moins onéreuse aujourd'hui consiste à graver des DVD en verre.



© R. MACNAY PHOTOGRAPHY, LLC/shutterstock.com

de la pérennité de l'information numérique et soulignerons que le problème le plus délicat est celui de l'altération des dispositifs de stockage. Des solutions existent, mais elles sont encore peu répandues.

Notons d'abord que sauvegarde et archivage à long terme des données ne sont pas synonymes. La première a pour but de prévenir les pertes soudaines de données (accidents divers, vols, etc.), la seconde, la disparition progressive des données utiles à long terme. En ce qui concerne la sauvegarde, comment garantir que la duplication d'un fichier est fidèle à l'original? Cette difficulté trouve une solution efficace dans les codes correcteurs, des algorithmes qui identifient les erreurs de copie.

Dupliquer sans erreurs

Cet aspect des méthodes numériques passe souvent inaperçu des utilisateurs : la possibilité de copier un document numérique avec la garantie que la copie ne contient pas d'erreurs. En effet, quand vous enregistrez un texte, aucune altération accidentelle ne viendra s'y glisser et changer une lettre en une autre pendant la sauvegarde. Cette fiabilité ne se retrouve pas dans les supports analogiques. Il suffit de penser aux textes anciens qui nous sont parvenus par le travail séculaire des moines copistes qui, malgré leur soin extrême, ont introduit des erreurs ; il est humainement impossible de faire des copies identiques. C'est d'ailleurs valable pour toute technique analogique : quelle que soit sa qualité, chaque recopie introduit un peu de « bruit » qui s'accumule de façon exponentielle de copie en copie. Au bout d'un certain nombre de sauvegardes, l'information initiale devient inexploitable.

Derrière cette merveille de la copie sans erreurs se cache le travail de mathématiciens, tels les Américains Irving Reed et Gustave Solomon, qui, dans les années 1960, ont mis au point des codes de correction d'erreurs numériques permettant des millions de copies sans erreurs d'un fichier. L'algorithme est complexe, mais son principe est simple : il faut « suréchantillonner » l'information, c'est-à-dire donner plus d'éléments qu'il n'est nécessaire pour restituer l'information. La redondance permet de vérifier qu'il n'y a pas eu d'erreurs dans la copie. Les algorithmes sont si efficaces que si nous recopions un fichier légèrement

endommagé, sur un CD de données un peu ancien par exemple, la copie éliminera les erreurs apparues entre-temps – tant que le taux d'erreurs apparues n'est pas trop grand. Le fichier sera restauré dans son état initial !

Le numérique semble avoir les mêmes vertus magiques que le Phénix renaissant de ses cendres. Et pourtant, il a un talon d'Achille, et pas des moindres : la préservation à long terme des données. Certes, il est de peu d'intérêt de conserver la plupart d'entre elles, dont la pertinence est éphémère ; mais il est essentiel d'en préserver la petite fraction qui garde un intérêt pendant des années, des décennies, parfois des siècles. L'intérêt de certaines données croît avec leur ancienneté, comme les données des satellites d'observation de la surface terrestre : en période de réchauffement climatique, il devient important de comparer les observations actuelles aux plus anciennes possibles, afin de discerner des évolutions à long terme. Les données des grands instruments scientifiques (telles celles du CERN, du satellite *Planck*, du radiotélescope SKA ou *Square Kilometer Array*) sont irremplaçables, rien ne garantissant qu'il sera possible un jour de répéter ces expériences très coûteuses. À plus petite échelle, les données du cadastre, de la comptabilité des entreprises, les données médicales, etc., doivent être conservées pendant longtemps. Enfin, à titre individuel, chacun d'entre nous souhaite conserver les souvenirs de sa vie personnelle et familiale.

La conservation de données numériques n'est donc pas une question anodine, puisqu'elle touche de nombreux domaines. Elle est confrontée à trois obstacles : documenter les fichiers pour savoir de quoi il s'agit, conserver le matériel et les logiciels nécessaires à la lecture du fichier et lutter contre le vieillissement des supports physiques.

Les deux premiers points sont ceux auxquels chacun songe avant tout. Toutes les formations de documentaliste enseignent l'importance des métadonnées. Celles-ci regroupent les informations qui accompagnent un document : sa nature, sa date, ses auteurs, des mots clefs associés, etc. Ces éléments facilitent une exploitation future.

Outre le problème des métadonnées, chacun a une anecdote à raconter de disquettes contenant des fichiers importants, mais illisibles faute de matériel adéquat.