

mentée sans erreurs, devient-elle parfaitement sûre ? Pas encore. Il reste les attaques indirectes, qui sont comparables à la découverte de la combinaison numérique protégeant un coffre-fort grâce à l'écoute au stéthoscope des bruits de son mécanisme de verrouillage... C'est ce que l'on nomme les attaques par canaux auxiliaires ou canaux cachés.

Ce type d'attaque n'est malheureusement pas l'apanage des films d'espionnage : même un cryptosystème parfaitement réalisé peut être sujet à des fuites d'informations parasites qui, bien analysées et exploitées, suffisent parfois pour le casser. Il arrive ainsi qu'un processeur émette des rayonnements exploitables, qu'il consomme de l'énergie suivant un profil traduisant les tâches de transmissions qu'il est en train d'accomplir, qu'un algorithme de chiffrement ait des temps d'exécution plus ou moins longs selon la valeur de la clef, etc.

De telles fuites ont déjà été utilisées pour casser certaines implémentations du cryptage RSA. Comment ? Les opérations d'élévation à la puissance nécessaires dans cet algorithme peuvent être réalisées soit

par multiplication, soit par élévation au carré ; or ces deux opérations induisent des profils de courant électrique différents, lesquels, mesurés et rapportés à leur cause, ont *in fine* permis de déterminer la clef !

Attention aux fuites !

Les failles de ce type sont les plus difficiles à déceler. Les modèles de cryptosystèmes, qui ignorent toute la couche technique sous-jacente au fonctionnement des ordinateurs et des réseaux, sont inutiles pour les repérer. Une fois qu'elles sont connues cependant, il arrive que des contre-mesures simples suffisent à les faire disparaître, par exemple un brouillage supplémentaire de données, une égalisation des signaux ou de la consommation électrique, le blindage des puces pour les rendre opaques à l'espionnage électromagnétique, etc. Autant de défenses qui tiendront jusqu'à ce qu'un autre attaquant découvre encore une fuite insoupçonnée d'information...

La conception et la mise en œuvre d'un cryptosystème impliquent ainsi une science

subtile. Cependant, toute la subtilité du monde ne peut rien contre l'utilisateur d'un coffre-fort qui laisse traîner les clefs ou contre la maladresse d'un utilisateur de communications chiffrées qui emploie le même mot de passe trivial (son prénom) pour se connecter au site de son groupe de fans de football, puis à sa banque et ensuite à sa messagerie d'entreprise...

De plus, les révélations du *Guardian* et du *New-York Times* suggèrent que le secret le mieux gardé de l'Internet aujourd'hui serait que les agences de renseignement anglo-saxonnes ont secrètement imposé aux « industries de l'Internet américaines et étrangères » d'insérer des « vulnérabilités dans leurs systèmes de cryptage ». Ces portes dérobées leur assureraient des entrées dans les protocoles de communication sécurisée largement utilisés en ligne, tels HTTPS (le protocole de transfert hypertexte sécurisé), la voix sur IP (téléphonie par Internet) ou encore SSL. Un coffre-fort doté d'une trappe indécidable, mais connue des espions, protégera bien mal nos secrets, c'est clair... ■

LES { Partagez les savoirs } RENDEZ-VOUS DU MUSÉUM

Entrée gratuite

CONFÉRENCES Cycle : Le Grand Herbar

Lundi 4 novembre — 18h : De la plante à l'image, l'icône liée à l'herbier de Paris
A. Gérin, conservateur en chef des Bibliothèques ; D. Lamy, ingénieur de recherche au CNRS, membre associé du Centre Koyré, dép. Systématique et Évolution, Muséum

Lundi 18 novembre — 18h : La Palynothèque de l'Herbier : historique, rôles et recherches actuelles
T. Deroin, maître de conférences, botaniste ; P.-H. Gouyon, professeur, dép. Systématique et Évolution, Muséum

Lundi 25 novembre — 18h : Les algues : du terrain aux collections en passant par la molécule
L. Le Gall, maître de conférences, phycologue, dép. Systématique et Évolution, Muséum

COURS PUBLICS Cycle : À la recherche de nos origines

G. Levi, directeur de recherches au CNRS, biologiste du développement, dép. Régulations, Développement, Diversité moléculaire, Muséum

Jeudi 28 novembre — 18h : La vie de l'embryon : de la fécondation à la naissance

Grand Amphithéâtre du Muséum — 57 rue Cuvier

MÉTIERS DU MUSÉUM

Dimanche 24 novembre — 15h : Astrophysicien
M. Gounelle, professeur, dép. Histoire de la Terre, Muséum

Auditorium de la Grande Galerie de l'Évolution — 36 rue Geoffroy St-Hilaire

Au Jardin des Plantes
Détails sur jardindesplantes.net, dans "l'agenda du jardin"

La quête d'un support numérique durable

Les disques durs, les CD, les DVD, les mémoires flash... se dégradent en quelques années. Cette situation est largement sous-estimée, mais diverses stratégies et de nouveaux dispositifs sont mis en place.

Franck Laloë et Erich Spitz

Les greniers des vieilles maisons abritent souvent des trésors. Quand on fouille dans des tiroirs ou dans des coffres, il n'est pas rare de retrouver de vieilles photos, des films huit millimètres ou des lettres manuscrites. Aujourd'hui, la plupart des documents, aussi bien les images que les vidéos et les textes, sont enregistrés sur des supports numériques, tels les disques durs des ordinateurs, les CD et les DVD.

À quoi ressembleront les chasses au trésor dans les greniers d'ici 50 ans ? Comment les supports numériques résistent-ils au passage du temps ? Les constructeurs ont annoncé que les supports numériques avaient une longue durée de vie. Les rares études réalisées montrent, au contraire, que ces supports se détériorent en quelques années. Dès lors, se pose la question de la pérennité des données numériques. Comment les conserver ? Ce problème est méconnu du grand public. Pourtant, il touche les documents personnels ayant

une valeur sentimentale (les photos et les lettres), mais aussi les informations plus sensibles, telles les données administratives, financières, médicales ou scientifiques.

Aujourd'hui, le numérique est le véhicule dominant de l'information. Il a supplanté les supports analogiques parce qu'il présente de nombreux avantages : une grande densité d'information permettant un gain de place considérable, la facilité des recherches dans l'information stockée, un accès rapide, une circulation quasi instantanée et son universalité – la même clef USB peut contenir des textes, du son, des vidéos, des images et des tableaux de données. La question de la pérennité des données a été reléguée au second plan. Pour l'industrie, la priorité était de développer des dispositifs rapides et contenant des quantités d'informations toujours plus importantes.

La question de la pérennité reste essentielle. Nous aborderons les différents aspects



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

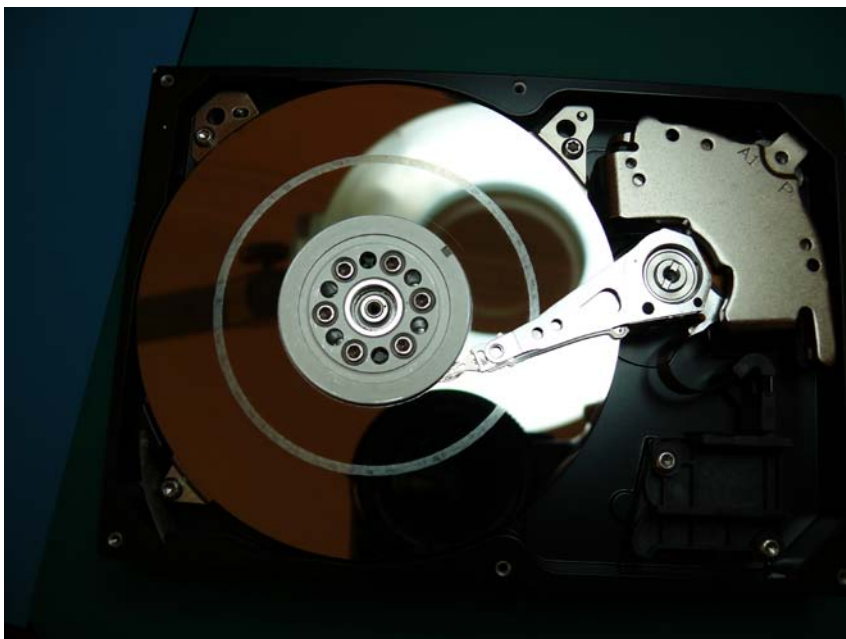
endommagé, sur un CD de données un peu ancien par exemple, la copie éliminera les erreurs apparues entre-temps – tant que le taux d'erreurs apparues n'est pas trop grand. Le fichier sera restauré dans son état initial !

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.



1. CE DISQUE DUR A ÉTÉ ENDOMMAGÉ par la tête de lecture qui a touché le matériau ferromagnétique à la surface du disque et tracé le sillon circulaire. Les données enregistrées sur le disque et situées dans le sillon sont définitivement perdues.

Qui possède encore un lecteur de disquette souple ou un lecteur de disque Zip ou Jaz ? De nombreux supports sont devenus obsolètes parce qu'ils ont été remplacés par des modèles plus performants ou parce qu'ils ont été des échecs commerciaux.

Ces deux aspects de la pérennité de l'information peuvent être réglés avec un peu d'organisation et de bon sens. Transposer au numérique le dépôt de métadonnées qui était déjà essentiel pour l'analogique n'est pas hors de portée ; de plus, on peut parfaitement choisir des systèmes d'exploitation, des logiciels et des formats de fichiers ouverts – dont le code est accessible et que chacun peut utiliser – ou très répandus, qui offrent des perspectives raisonnables de pérennité (voir l'article de Roberto Di Cosmo dans ce numéro).

Ces deux premiers problèmes sont importants, mais ils ne doivent pas masquer le troisième, qui est bien plus grave. On peut prendre toutes les précautions que l'on voudra en créant des métadonnées parfaites et en utilisant des formats robustes : si les données elles-mêmes disparaissent, cela ne sert évidemment à rien ! Ce problème a longtemps été sous-estimé et négligé. Hélas, nous allons voir que cet aspect de la pérennité des données pose des difficultés techniques et économiques qui sont loin d'être résolues.

La variété des dispositifs de stockage disponibles dans le commerce et leur évolution rapide peuvent donner l'impression qu'il existe de nombreuses méthodes différentes pour stocker et sauvegarder de l'information numérique. Certains supports sont-ils adéquats pour de l'archivage de longue durée ? En réalité, ces dispositifs fonctionnent sur un petit nombre de principes physiques et se rangent dans quelques catégories simples qui subissent toutes des détériorations en quelques années.

Tous les supports s'altèrent

La première catégorie fait appel à l'enregistrement magnétique : le système d'enregistrement est conçu pour orienter de façon contrôlée des petits domaines dans une couche magnétique déposée sur une surface plane. La lecture s'effectue ensuite par une tête électromagnétique sensible à cette orientation. C'est ainsi que fonctionnent les disques durs, les bandes magnétiques (ou LTO, pour *Linear Tape Open*) et même les disquettes de formats divers, dont la plupart ont maintenant disparu à cause du succès des mémoires flash. Les disques durs présentent de grandes capacités de stockage pour un coût faible, et avec un accès rapide aux données. Ils sont cependant fragiles et

soumis à des pannes soudaines et imprévisibles (ou *crash*) pouvant entraîner une perte totale des données (par exemple, si la tête de lecture touche la couche magnétique, voir la figure 1). Pour ne pas perdre ses données, une possibilité est la sauvegarde sur plusieurs disques en parallèle (tels les systèmes redondants RAID, pour *Redundant Array of Independent Disks*). La longévité des disques durs est difficile à évaluer, car elle dépend de nombreux facteurs indépendants. On s'accorde à penser néanmoins que, hors *crash*, une durée de vie de cinq ans est un bon ordre de grandeur. Pour optimiser cette longévité, il faut laisser tourner les disques durs en permanence, ce qui minimise les incidents. Les grands centres de données appliquent cette méthode coûteuse en électricité. On estime que les centres de données contribuent ainsi à plus de deux pour cent de la production de dioxyde de carbone d'origine humaine.

Il y a quelques années, les bandes magnétiques étaient considérées comme obsolètes, car moins performantes que les disques durs. Mais, précisément du fait de la nécessité de l'archivage des données, les bandes magnétiques ont effectué un retour spectaculaire sur la scène du numérique : elles conservent les données de façon passive, sans consommation électrique. Cependant, leur durée de vie est aussi limitée, de l'ordre de cinq ans, temps au bout duquel il devient indispensable d'effectuer une « migration » – c'est-à-dire de recopier les données sur une bande neuve, si on ne veut pas les perdre. Il existe d'ailleurs des robots dont la fonction est de gérer un grand nombre de bandes, d'y suivre le nombre d'erreurs et d'effectuer une migration vers des bandes neuves dès que c'est nécessaire.

La seconde catégorie de supports numériques fait appel à un enregistrement électrique et non magnétique : dans les mémoires flash, de très nombreux petits transistors stockent une charge électrique qui définit la valeur du bit enregistré. Les mémoires flash sont devenues d'usage universel, dans les appareils photo, les caméscopes, les ordinateurs, les téléphones, etc. Elles autorisent des accès aux données encore plus rapides que les disques durs, mais le prix par mégaoctet de données est plus élevé. Quant à la longévité des données stockées sur des mémoires flash, on dispose de peu d'informations à ce sujet. On sait cependant que les rayons cosmiques et

déchargent progressivement les transistors et effacent donc les données. On considère souvent que la durée de vie est de l'ordre de cinq ans.

La troisième catégorie est constituée par les disques optiques numériques (DON), tels que les CD, les DVD et les disques *Blu-ray*. Les temps d'accès aux données enregistrées sont bien plus longs que ceux des disques durs et des mémoires flash. De plus, leur capacité est nettement inférieure à celle des disques durs. On fabrique pourtant chaque année des milliards de disques optiques numériques, encore que leur nombre soit en constante diminution chaque année; beaucoup pensent qu'ils sont en voie d'extinction progressive. En termes de longévité toutefois, ce sont les meilleurs supports disponibles actuellement: les disques «pressés» fabriqués en grande série ont une durée de vie qui serait comprise entre 20 et 30 ans. Cependant, ce n'est pas le cas des disques enregistrables, qui reposent sur des principes physiques plus délicats (inscription dans un colorant composé de molécules photosensibles ou, pour les disques réenregistrables, un changement de phase qui rend la surface plus ou moins réfléchissante). Les études du Laboratoire national de métrologie et d'essais en France et du Laboratoire lutherie acoustique musique de l'Université Paris 6 ont montré que certains disques enregistrés perdent leurs données en un an seulement! D'autres, heureusement, ont des durées de vie dix fois plus longues.

Les disques optiques offrent des perspectives très intéressantes en termes de longévité. Il y a une vingtaine d'années une société française, *Digipress*, a commercialisé un CD ayant une excellente durée de vie, puisque les données étaient directement gravées dans du verre. Hélas, l'invention devait être trop en avance sur son temps, à une époque où beaucoup croyaient encore à la longévité des supports numériques. Ce produit n'a pas trouvé son marché. Mais, depuis deux ans, une société allemande, *Syplex*, a repris l'idée et propose des DVD gravés dans le verre dont la longévité est remarquable, comme l'ont montré des études du Laboratoire national de métrologie et d'essais (voir l'encadré page 82). L'avantage d'une telle solution est que les données sont stockées dans le format DVD, très répandu dans le monde, et qu'elle ne demande

aucun lecteur plus spécifique qu'un lecteur standard. Quelques autres produits sont proposés sur le marché et annoncés avec une grande longévité, mais à notre connaissance, aucun n'a subi des tests aussi soigneux que ces disques. Il faut espérer que les esprits aient maintenant suffisamment évolué pour que ce nouveau produit se répande rapidement, en particulier par des commandes institutionnelles, mais rien ne le garantit.

Graver dans le verre

Quelles stratégies peuvent être adoptées pour préserver l'information numérique à long terme avec les dispositifs existants? Deux approches sont possibles: la stratégie dite passive, qui consiste à écrire l'information sur un support pérenne et à le stocker dans un endroit sûr; la stratégie dite active, qui consiste à écrire l'information sur un support numérique standard et à en suivre constamment l'évolution, afin d'effectuer une migration vers un support neuf lorsque nécessaire. La stratégie passive est celle que l'humanité a pratiquée depuis toujours, mais elle utilisait des supports analogiques – que l'on songe à la bonne conservation des parchemins anciens, voire aux écritures cunéiformes sur des pierres! Elle semble donc la plus naturelle, mais dans le cas du numérique, elle demande un investissement initial, par exemple pour la confection de DVD gravés dans le verre.

La stratégie active ne demande pas un tel investissement. Dans la plupart des cas,

LES AUTEURS

Franck LALÔÉ est directeur de recherche au Laboratoire Kastler Brossel de l'École normale supérieure. Il est président du Groupement d'intérêt scientifique SPADON. Il a été rapporteur du groupe Pérennité des supports numériques.

Erich SPITZ est membre de l'Académie des sciences et de l'Académie des technologies. Il a présidé le groupe Pérennité des supports numériques.

BIBLIOGRAPHIE

J. Zhang *et al.*, *5D data storage by ultrafast laser nanostructuring in glass*, compte rendu de conférence CLEO: Science and Innovations, 2013.

J. Perdureau, *Tenue dans le temps des disques enregistrables DVDR et DVD en verre (Syplex) à température et hygrométrie élevées*, rapport du GIS DON, LNE, 2012.

J.-C. Hourcade *et al.*, *Longévité de l'information numérique; les données que nous voulons garder vont-elles s'effacer?*, rapport à l'Académie des Sciences et à l'Académie des Technologies, EDP Sciences, 2010.

L'altération, un processus mal compris

Les différents supports numériques ont des durées de vie relativement courtes. De nombreux facteurs entrent en jeu et il est difficile de dire précisément quels phénomènes sont en cause et lesquels sont les plus importants. Pour les disques durs, les pannes mécaniques sont une difficulté majeure. Mais d'autres problèmes peuvent surgir: certains éléments du disque stockant l'information peuvent devenir défectueux et illisibles.

Les rayons cosmiques peuvent décharger les transistors des mémoires flash et modifier l'infor-

mation. Les supports optiques, les CD et les DVD, correspondent à un empilement de couches de protection en plastique, de couches contenant les données et d'une pellicule métallique qui sert de réflecteur pour le laser de lecture des données. La pellicule plastique est sensible à la surexposition aux ultraviolets de la lumière du Soleil et à la chaleur. Les tests de longévité de ces supports ont aussi montré qu'ils sont sensibles aux réactions d'oxydation par l'eau et par l'air. Les supports optiques enregistrables sont encore moins stables.

Les facteurs externes ne sont pas les seuls à influencer la durée de vie des supports. La qualité des matériaux utilisés et les précautions prises lors de la fabrication sont aussi importantes. Le Laboratoire national de métrologie et d'essais a montré que sept pour cent des DVD enregistrables du commerce présentaient des microtaches. Elles seraient dues à des impuretés présentes dans la couche réfléchissante et accéléreraient le processus d'oxydation, réduisant la durée de vie du support à un an.

l'archivage ne se fait pas localement, mais dans des centres de données spécialisés, ou « fermes », contenant de nombreuses armoires remplies de disques durs. Ces centres sont gérés par du personnel spécialisé, dont une des missions est de procéder aux migrations de supports quand cela est nécessaire. Cette solution fonctionne de façon satisfaisante, et elle est bien dans la ligne de l'engouement actuel pour le *cloud*. Elle pose néanmoins des difficultés. La première est que, sauf s'il dispose d'un centre informatique évolué, le propriétaire des données perd le contrôle de l'archivage; il doit faire confiance à un prestataire de service, sachant que toute erreur de sa part peut engendrer une perte de données irréversible. Il faut aussi prévoir à l'avance le cas où ce prestataire viendrait à cesser ses activités, et les procédures de restitution des données. Le second problème est celui de la confidentialité: il est évident que des données sensibles sont mieux protégées si elles sont à l'abri dans un coffre-fort sur place que si elles sont transportées à des milliers de kilomètres.

La comparaison des coûts des deux stratégies est bien sûr déterminante. Celui de la stratégie passive est facile à évaluer: le coût de transfert des données sur des supports spéciaux et de très grande longévité peut être élevé; par exemple, un DVD de cinq gigaoctets de données gravées dans le verre coûte actuellement à peu près 150 euros, soit plus de 100 fois le prix d'un DVD enregistrable standard! Il est clair que la stratégie passive implique une dépense initiale importante. Ensuite, les coûts sont réduits: celui lié à l'immobilisation d'un espace d'archivage, par exemple dans une cave ou un coffre-fort.

La méthode passive, plus économique

Pour la stratégie active, il en va tout autrement. L'erreur classique est de ne prendre en compte que les coûts des supports nécessaires pour les migrations, coûts qui sont négligeables. Les coûts réels sont en fait induits par la nécessité du suivi technique (salaires des informaticiens,

locaux avec environnement technique) et la consommation électrique des « fermes » de disques durs. Ces coûts croissent donc linéairement avec le temps, contrairement à ceux de la stratégie passive, qui restent quasi constants; chacun comprend donc qu'il existe un certain délai à partir duquel c'est la stratégie passive qui devient la moins onéreuse.

Des comparaisons entre les coûts des deux méthodes ont été effectuées. Ainsi, pour l'archivage de 500 gigaoctets, la stratégie active sur bande magnétique est la moins chère pendant six à sept ans, puis c'est la stratégie passive sur DVD gravés en verre; pour l'archivage de un téraoctet, le point de croisement est repoussé à 10 ans environ. Il est intéressant de noter que cette étude montre que la stratégie active sur disques durs RAID, pourtant la plus répandue aujourd'hui, n'est jamais la moins onéreuse, que ce soit à court ou à long terme.

À partir de ce simple constat, les organismes et entreprises qui ont besoin de conserver d'importants volumes de données pendant plus de dix ans devraient opter

Évaluation de la tenue dans le temps du DVD en verre Sylllex

Des études menées par le Laboratoire national de métrologie et d'essais (LNE) sur le vieillissement naturel des disques optiques numériques enregistrables CD-R et DVD-R ont montré une grande disparité de comportement des disques dans le temps, avec des durées de vie comprises entre un an et plus de 20 ans selon les modèles. C'est en contradiction avec les durées de vie de 50 à 300 ans annoncées par les fabricants.

Les membres du Groupement d'intérêt scientifique SPADON ont souhaité étudier des solutions différentes des DVD-R classiques, tel le DVD *GlassMasterDisc* de la Société Sylllex. Il n'utilise ni polycarbonate ni de couche d'enregistrement organique, points faibles potentiels des DVD-R classiques. L'information est directement gravée dans du verre.

L'estimation de la durée de vie de ce disque aurait pu se faire en utilisant la méthode normalisée par vieillissement accéléré en température et humidité (ISO/CEI 16963). Cependant, d'une part, cette norme n'a pas été validée expérimentalement, d'autre part, elle ne semble pas applicable aux DVD, au vu des résultats de plusieurs études antérieures, probablement parce que les facteurs d'altération physico-chimiques sont multiples. Il n'est donc pas possible d'estimer la durée de vie réelle des DVD à partir d'un test de vieillissement accéléré. Pour étudier l'efficacité du *GlassMasterDisc* pour l'archivage, le GIS-SPADON et le LNE ont

choisi d'évaluer la tenue dans le temps de ce support exposé à des températures et des taux d'humidité élevés et de comparer les résultats aux DVD enregistrables classiques. Les six modèles choisis comme éléments de comparaison sont réputés comme ayant une bonne tenue dans le temps (cas des DVD-R utilisant une couche enregistrable inorganique) ou ont montré dans les études précédentes du LNE une tenue au vieillissement accéléré (80 °C et 85 pour cent d'humidité) supérieure à celle des autres modèles.

L'étude a consisté à placer les DVD dans une enceinte climatique (voir la photographie ci-contre de l'enceinte CLIMATS) dans des conditions dites très sévères de température et d'humidité, pour une durée maximale de 1 000 heures, et cela par tranches de 250 heures. Le DVD *GlassMasterDisc* étant en verre, la température a été choisie égale à 90 °C, soit 10 °C de plus que pour les précédentes études de DVD, tout en maintenant l'humidité relative à 85 pour cent. La fin de vie des DVD correspond à un taux de données numériques erronées dépassant les limites normalisées ou à l'illisibilité du disque.

Il ressort de cette étude, qu'en vieillissement accéléré, le DVD *GlassMasterDisc* a une durée de vie nettement supérieure à celle des DVD-R existants sur le marché. Après 1 000 heures d'exposition dans ces conditions délétères, le DVD *Glass-*



MasterDisc ne présente pas d'altération significative de sa qualité, alors que le meilleur des autres modèles de DVD-R a une durée de vie maximale de 250 heures.

Le DVD *GlassMasterDisc* présente donc le meilleur potentiel, parmi les DVD enregistrables existants sur le marché, pour réaliser un archivage de très longue durée des données numériques.

Jacques Perdereau,
Ingénieur au LNE

Actuellement
en kiosque

POUR LA SCIENCE

- Le bestiaire des ARN
- La mémoire des cellules
- La matière noire du génome
- Le cerveau modelé par l'expérience

Vers une médecine
personnalisée

Le magazine thématique de l'actualité scientifique

N° 81 Octobre-Décembre 2013

L'hérédité sans gènes

Comment l'environnement rebat les cartes

DOSSIER HORS-SÉRIE

www.pourlascience.fr

le site de référence de l'actualité scientifique internationale



Disponible sur



n°81 - 120 pages - prix de vente : 6,95€

pour la solution passive, la moins chère de toutes. Il est étonnant de constater qu'il n'en est rien, et de très loin, car presque personne ne s'intéresse encore aux supports numériques pérennes. La raison en est peut-être que beaucoup de décideurs, et même de particuliers, n'ont pas pris la mesure du problème et préfèrent conserver leurs illusions plutôt que d'envisager des catastrophes futures potentielles ; et cela nécessite d'engager des politiques à moyen et long termes. La stratégie active reste donc à peu près la seule utilisée dans le monde, à quelques exceptions près, comme à l'Institut national de l'audiovisuel, en France, pour les émissions radiophoniques.

Cependant, évitons un alarmisme excessif : les données les plus importantes, celles du robot *Curiosity* sur Mars ou celles du CERN par exemple, ne seront pas perdues ; les sommes nécessaires à leur conservation seront très probablement dépensées, car elles restent une faible fraction du coût de ces grands programmes de recherche. De façon plus générale, toutes les données qui touchent à des domaines d'activité où l'argent abonde ont toutes les chances de survivre : les banques, les notaires et d'autres professions gérant des sommes importantes sont et resteront capables de mettre en œuvre la stratégie active dans de bonnes conditions.

Assurer la pérennité des données est difficile dans d'autres secteurs. Pour les hôpitaux, les données médicales doivent être légalement conservées au minimum 20 ans, une gageure dans un contexte de restrictions budgétaires. Et dans bien d'autres cas, on peut être encore plus inquiet. Il est d'ores et déjà connu que les données issues de la numérisation de documents analogiques ont parfois été perdues (dans certaines archives départementales par exemple), et l'on est loin d'avoir une vue générale sur la question. Quant aux données personnelles et familiales, tout porte à croire qu'elles n'auront pas la durée de vie des photos ou des films huit millimètres que nous ont transmis nos parents.

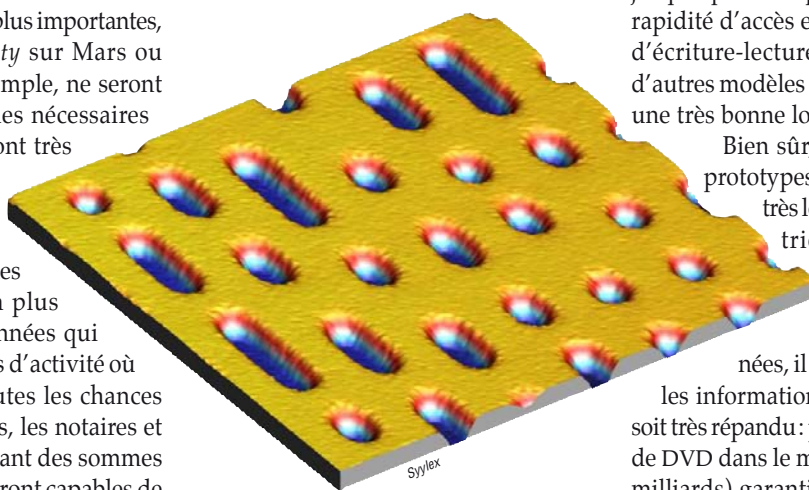
Peut-on raisonnablement espérer voir émerger des supports numériques peu onéreux, permettant le stockage de grandes quantités d'informations et garantissant leur conservation pendant des siècles

sans migration ? Il semble que les grands laboratoires industriels se sont longtemps concentrés sur la fiabilité, le volume de données stockées et la rapidité d'accès, mais une prise de conscience progressive et l'émergence d'un marché centré sur la longévité pourraient changer la donne. De plus, des structures publiques commencent à s'intéresser au sujet. Le Groupe d'intérêt scientifique SPADON (*Supports pérennes d'archivage pour les données numériques*) regroupe plusieurs laboratoires universitaires et du CNRS. Il suit et coordonne les progrès techniques dans ce domaine, et suscite de nouvelles recherches. De même, le groupe Pérennité des informa-

respondent aux trois dimensions d'espace habituelles, puisque l'inscription se fait en volume, plus deux dimensions pour la taille et l'orientation des nanostructures. La capacité de stockage extrapolée est de plus de 300 téraoctets par disque de silice, et la durée de vie serait pratiquement infinie ! On a aussi beaucoup parlé récemment des possibilités offertes par les skyrmions, des configurations quantiques particulières dans les matériaux magnétiques, ou même de l'enregistrement de données dans l'ADN (voir l'article de Christophe Dessimoz dans ce numéro) ! Enfin, de façon moins futuriste, rien n'interdit d'imaginer que les mémoires flash, jusqu'à présent optimisées en termes de rapidité d'accès et de nombre de cycles d'écriture-lecture, puissent engendrer d'autres modèles de dispositifs assurant une très bonne longévité.

Bien sûr, il ne s'agit là que de prototypes de laboratoire, encore très loin de toute offre industrielle. De plus, il faut garder à l'esprit que, pour un archivage à très long terme des données, il est essentiel de stocker les informations dans un format qui soit très répandu : par exemple, le nombre de DVD dans le monde (des dizaines de milliards) garantit probablement qu'ils resteront lisibles dans un siècle ou plus. Un support d'information numérique pérenne doit donc atteindre un certain niveau de diffusion pour être un bon candidat à l'archivage.

Aujourd'hui, le DVD gravé dans le verre est la meilleure piste. Il résiste bien à l'altération, mais sa capacité est limitée et son prix relativement élevé. C'est probablement pourquoi il est très peu utilisé. L'enjeu de la pérennité des données numériques est important, et il est crucial de trouver des solutions. Certes, cela nécessite une bonne organisation avec l'usage de métadonnées et de normes sur les formats pour les conserver sur le long terme. Mais l'aspect le plus important et le plus difficile à régler est celui de l'altération des supports numériques. La prise de conscience du problème va encourager la recherche et le développement de dispositifs à longue durée de vie. Il est difficile de dire si une solution pourrait apparaître bientôt et d'où elle viendrait. De l'optique ? De la physique du solide ? ■

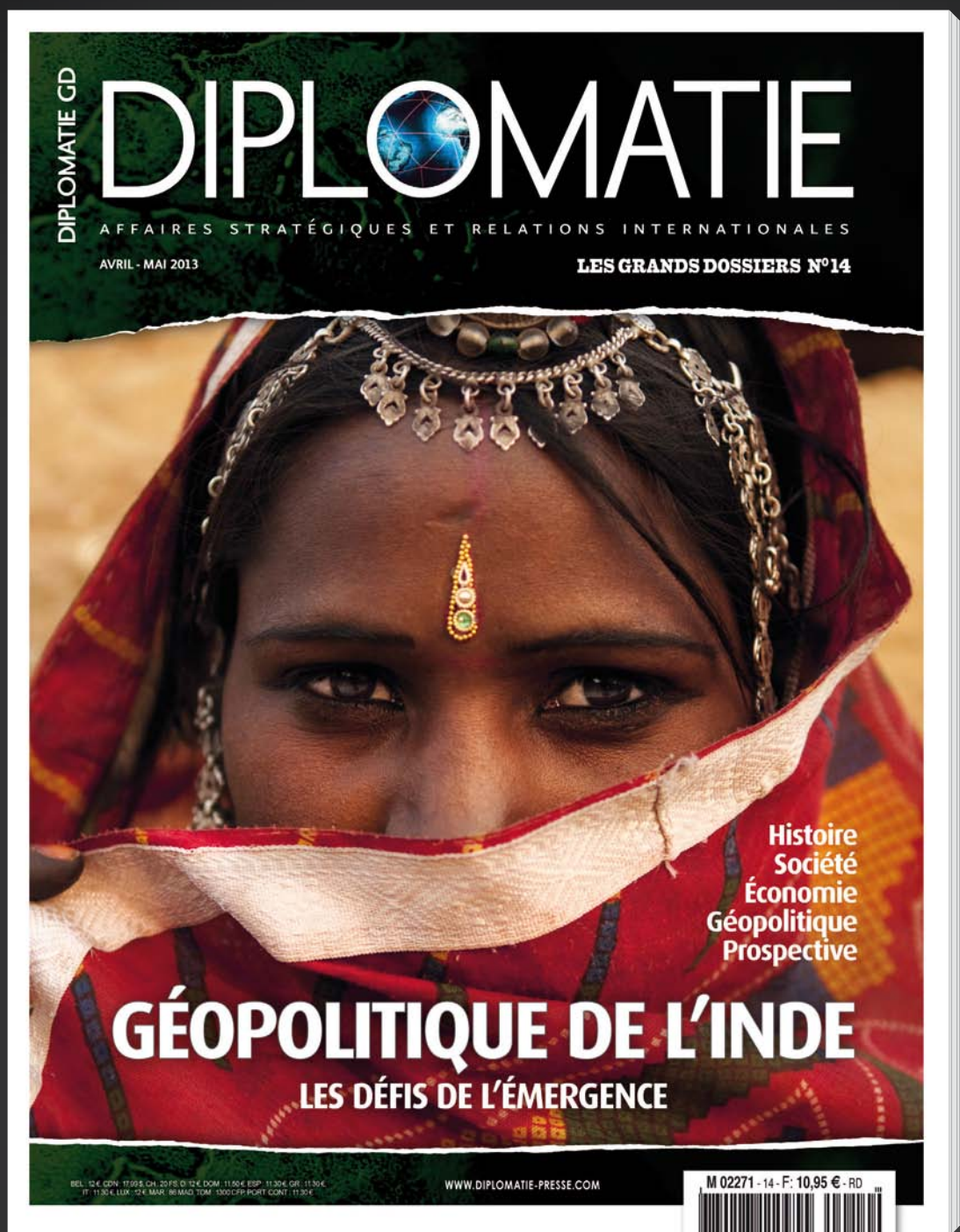


2. LA SURFACE D'UN DVD EN VERRE vue au microscope à force atomique. L'information est gravée directement dans le verre. Ce support offre la meilleure durée de vie et la solution la plus économique à long terme.

tions numériques, auquel participent, entre autres, la Bibliothèque nationale de France et le Centre informatique national de l'enseignement supérieur (CINES), effectue un suivi depuis plusieurs années.

Ces efforts commencent à porter leurs fruits. Une voie prometteuse semble être fournie par des dispositifs optiques, avec l'écriture par des lasers à impulsions (nanosecondes ou femtosecondes) à l'intérieur même d'une structure vitreuse. Récemment, dans le cadre du projet européen *Femtoprint*, un groupe de l'Université de Southampton a annoncé avoir mis au point une technique d'enregistrement « à cinq dimensions » en créant, avec un laser ultrarapide, des nanostructures dans le verre – les cinq dimensions cor-

En vente chez votre marchand de journaux



100 pages • Format 23x30 cm • 10,95 € (codification Prestalis : 02271)

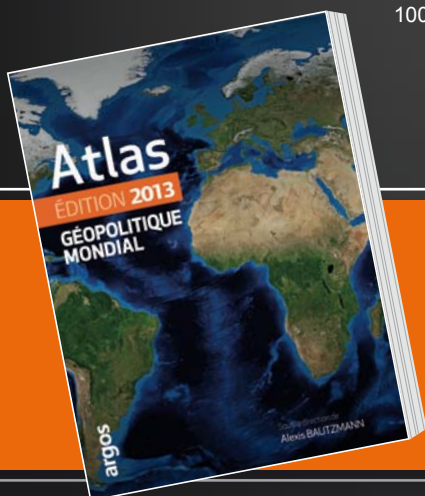
Commandez DIPLOMATIE sur Internet

WWW.GEOSTRATEGIQUE.COM

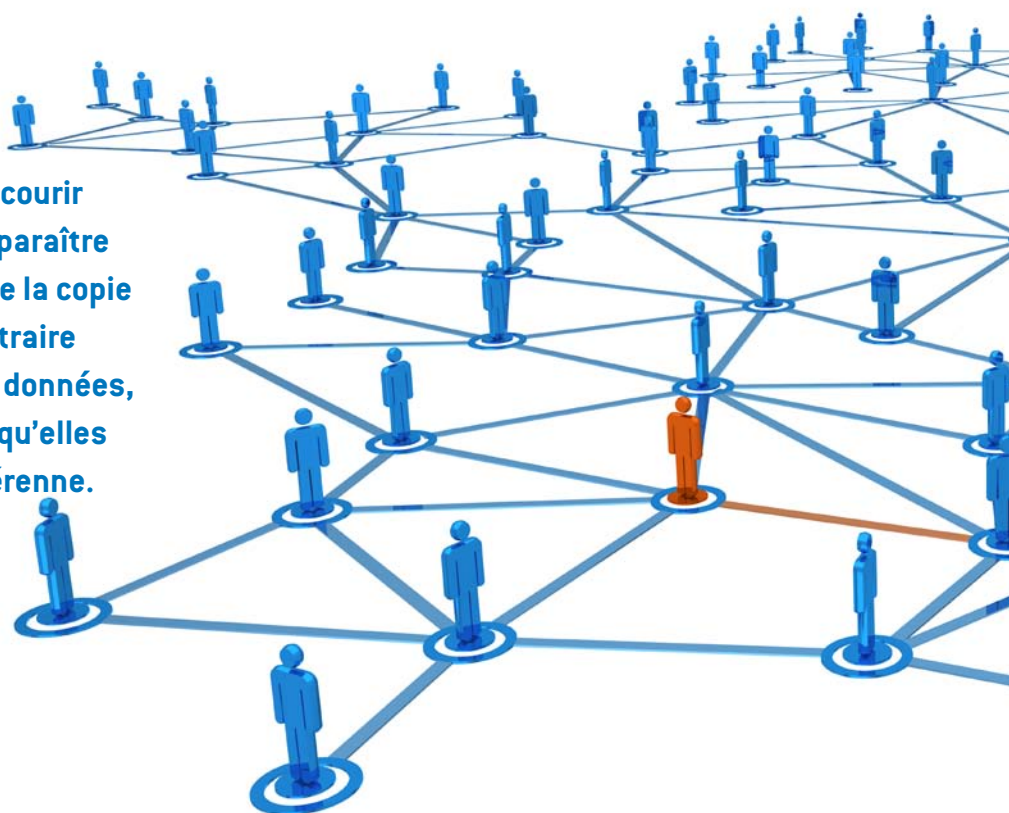
Et ne manquez pas en librairie

ATLAS GÉOPOLITIQUE MONDIAL 2013

270 cartes et graphiques pour décrypter toute l'actualité internationale
(Éditions ARGOS/AREION - Diffusion PUF - 23x30 cm - 212 pages - 22,50 €)



Centraliser les données, c'est courir le risque de les voir toutes disparaître d'un seul coup. L'avènement de la copie numérique facile invite au contraire à partager le plus possible les données, afin d'augmenter les chances qu'elles soient conservées de façon pérenne.



Roberto Di Cosmo

Organiser le partage pour préserver les données

Qu'il s'agisse d'histoire, de philosophie, d'avancées scientifiques, d'œuvres littéraires ou artistiques, ou tout simplement d'histoires individuelles, l'humanité a toujours cherché à préserver ses savoirs et à les transmettre aux générations futures. Pour ce faire, elle a d'abord employé les récits oraux et les mythes. Puis l'invention de l'écriture a entraîné celle des livres, et tout a changé le jour où ces « recueils » ont pu être multipliés grâce à l'imprimerie, avant d'être partagés en masse dans les bibliothèques. Aujourd'hui, la copie numérique et Internet permettent une diffusion des connaissances plus efficace que jamais. Analysons les modalités, les risques et les promesses de la transmission numérique du savoir, une innovation, qui, une fois de plus, change notre vie culturelle.

Dupliquer des informations a longtemps été difficile. Pensons, par exemple, au coût prohibitif de la reproduction d'un livre par un copiste avant l'invention de l'imprimerie. La seule stratégie possible consistait à réunir les précieux recueils en un endroit à la fois sûr et accessible, où l'on pouvait les protéger tout en les rendant disponibles à la consultation à l'aide d'index centralisés. La grande bibliothèque d'Alexandrie, où, pendant l'époque hellénistique (entre -336 et -30), on tenta de rassembler « tout le savoir du monde », illustre que cette stratégie est ancienne. La destruction de cette bibliothèque, dans les premiers siècles de notre ère, montre les risques, bien connus des informaticiens, que courent les systèmes centralisés, puisqu'il suffit d'en détruire le centre pour tout faire disparaître...

Grâce aux avancées techniques du dernier demi-siècle, nous vivons aujourd'hui une nouvelle rupture dans l'histoire culturelle de l'humanité.

La copie parfaite...

Pour la première fois, nous disposons d'instruments pour répliquer parfaitement un nombre toujours plus grand de biens immatériels. Les moyens informatiques permettent en effet de numériser – c'est-à-dire d'encoder sous forme d'une suite de chiffres binaires – du texte, du son, des images et de la vidéo, puis de répliquer à volonté ces informations sans perte de qualité et pour un coût négligeable. Tout cela est devenu banal, et le domaine des biens duplicables à l'identique ne cesse



de s'élargir, comme l'illustre l'arrivée des scanners et des imprimantes 3D.

Dès lors, on est tenté de penser que le problème ancien de la préservation des données est résolu... Après tout, quand il est possible de dupliquer, sans perte de qualité, et à coût quasi nul, autant de fois qu'on le souhaite, comment pourrait-on encore perdre de l'information ?

... n'existe pas !

Mais à y regarder de plus près, on s'aperçoit vite qu'il n'en est rien. Non seulement les techniques actuelles de reproduction n'ont pas résolu le problème, mais elles l'ont en fait compliqué. En premier lieu, les mêmes outils qui rendent facile la duplication de l'information ont aussi pour conséquence d'engendrer dans la pratique une quantité gigantesque d'information : pour l'année 2011, on a calculé que 1,8 zettaoctet de données ont été produites (voir la figure 2), et huit zettaoctets sont annoncés pour 2015 ! Précisons que la plus grande partie de cette information n'est pas créée directement par l'homme, mais par nos machines, qui gardent non seulement

les traces numériques de nos opérations informatiques, telles la transmission ou la conversion de fichiers, mais aussi les données et les métadonnées (données sur les données) produites par nos échanges sur les réseaux sociaux...

Décider de ce qui doit être préservé dans cette gigantesque masse et de ce qui peut tomber dans l'oubli n'est pas aisé. Ainsi, la Bibliothèque du Congrès des États-Unis archive désormais les messages envoyés par les citoyens américains sur le réseau *Twitter*, alors que ces commentaires et autres micro-informations devaient à l'origine n'être qu'éphémères...

Même si l'on disposait d'une méthode pour décider exactement de ce qui doit être préservé, la conservation à long terme des objets numériques serait bien plus problématique qu'il n'y paraît. À la différence des livres et des tableaux, qui livrent immédiatement leur contenu, l'accès aux objets numériques suppose un environnement technique complexe. Ce dernier comporte, d'une part, des supports matériels de stockage de l'information susceptibles de se dégrader et, d'autre part, des logiciels complexes, utilisant des formats définis et fonctionnant seulement sur certains processeurs qui, eux, sont susceptibles de devenir rapidement obsolètes. À moyen et long termes, les risques de pertes de données numériques sont donc considérables.

Illustrons cette réalité peu connue par un exemple : si l'on retrouvait aujourd'hui un document électronique rédigé il y a seulement... 30 ans, sans doute aurait-il la forme d'une disquette magnétique de 5 pouces 1/4 contenant un fichier *WordStar*. À supposer que cette grosse disquette ne soit pas trop dégradée, ce qui est fort peu probable (voir l'article de Franck Laloë et Erich Spitz dans ce numéro), nous nous heurterions à une autre difficulté : les lecteurs de tels supports ont presque tous disparu. De même, le logiciel *WordStar* n'est plus employé (car il n'est plus disponible), de sorte que le format dans lequel il enregistrait les données n'est plus reconnu par les traitements de texte actuels (voir la figure 4).

L'ESSENTIEL

- La réplication numérique de données et leur diffusion par Internet semblent faciliter la conservation de la mémoire humaine.
- En réalité, elle la menace, car les moyens techniques de cette réplication se périment et rendent les données inaccessibles.
- Les offres d'hébergement gratuit des grands opérateurs d'Internet, pour utiles qu'elles soient, ne constituent pas non plus une solution sans risques.
- Il reste à refaire sur le réseau ce que l'humanité a toujours fait : répartir sa mémoire entre tous !

© Shutterstock.com/festa

1. LE RÉSEAU HUMAIN est depuis toujours le dépositaire de l'expérience de l'humanité. Pour en faire aussi le dépositaire de sa mémoire numérique, encore faut-il faciliter les échanges de données entre pairs...

Le cas du projet *Domesday* de la BBC (société britannique de radiotélévision) fournit un exemple de ces difficultés techniques. En 1986, soit quelque 900 ans après une entreprise similaire (voir l'encadré page 90) ordonnée par Guillaume le Conquérant (1027-1087), la BBC a lancé un recensement et un sondage sur l'ensemble du Royaume-Uni. Les données, recueillies par plus d'un million de volontaires dans les 23 000 cellules territoriales définies pour l'occasion, constituaient une sorte d'immense instantané de la société britannique; elles furent mises à la disposition du public au moyen d'un vidéodisque de type *LaserDisc*, un support des données dont on considérait à l'époque qu'il serait durable. Or, dès 1998, il fut supplanté par les DVD, de sorte qu'à partir du début des années 2000, on ne trouvait déjà pratiquement plus de lecteurs de *LaserDisc*, ni d'ordinateurs où pouvait tourner le logiciel nécessaire pour présenter le contenu du disque. Résultat: plusieurs équipes durent fournir un important effort de recherche afin d'émuler (d'imiter) le fonctionnement des lecteurs de *LaserDisc* dans un environnement informatique actuel.

Ce cas montre que pour préserver des logiciels, il importe de maintenir les connaissances nécessaires pour les faire exécuter par une machine, et donc de disposer de leur code source et de leur environnement d'exécution. Dans le cas du *LaserDisc*, quelques années avaient suffi pour que tout cela disparaisse! C'est bien à cause de telles situations qu'aujourd'hui, on reconnaît de plus en plus que l'usage de formats de données ouverts (qui peuvent être librement utilisés et modifiés, et dont les spécifications

sont publiques et sans restriction d'accès) et de logiciels libres (dont le code source est disponible pour tous) est indispensable à la pérennité des informations.

Mais une fois le problème de la lecture des *LaserDisc* du projet *Domesday* résolu, on s'aperçut que d'épineux problèmes légaux restaient à démêler. Selon le droit d'auteur en vigueur, la réalisation d'une copie exigeait que l'on recueille d'abord les autorisations des millions d'individus ayant contribué au projet, ainsi que celles des auteurs des logiciels utilisés... À une époque où reproduire l'information est devenu banal, l'imposant arsenal juridique des droits d'auteurs nous interdit nombre de répliques, et ce sur des durées de plus en plus longues... Si le droit protégeant la propriété intellectuelle sert aujourd'hui essentiellement à préserver les modèles économiques ayant précédé l'ère d'Internet, il a souvent pour effet de bloquer le développement des modèles nécessaires à l'économie numérique.

Copier est-il hors la loi ?

Si nous voulons préserver les données, nous sommes donc confrontés à un dilemme: nous devons d'un côté les recopier sur de nouveaux supports (avant leur inévitable dégradation) et sous de nouveaux formats (avant leur inévitable obsolescence); d'un autre côté, nous devons respecter des lois qui nous interdisent de réaliser ces copies... Notre civilisation est donc confrontée à la question suivante: qui doit se charger du renouvellement continu des données et dans quelles conditions? Plusieurs



2. EN 2011, 1,8 ZETTAOCTET de données a été produit, et diverses estimations en annoncent huit pour 2015. Un zettaoctet représente 1 000 milliards de gigaoctets (c'est-à-dire 10^{21} octets). Ainsi, en quelques jours, l'humanité accumule aujourd'hui plus d'information qu'elle n'en a produit en deux millions d'années avant l'invention d'Internet.



3. LES SUPPORTS DE STOCKAGE se suivent et ne se ressemblent pas, tant en termes de capacité que de format d'enregistrement. Ci-dessus, quelques-uns des formats datant des 30 dernières années: à gauche, une disquette 5 pouces 1/4, au milieu une disquette 3 pouces 1/2 de haute densité, puis un CD-ROM et, à droite, une clef USB actuelle.



4. WORDSTAR est un logiciel de traitement de texte introduit en 1978, qui fonctionnait sous DOS, le système d'exploitation pour ordinateurs personnels de Microsoft. Il occupa une position dominante au début des années 1980, mais fut supplanté par *Word* dans les années 1990, puis abandonné et oublié, au point de devenir inutilisable aujourd'hui.

réponses sont possibles. Chacune a ses avantages et ses inconvénients.

Une première possibilité serait de sous-traiter la préservation des données à quelques grands opérateurs chargés de les recopier régulièrement sur les nouveaux supports avant que les anciens ne se dégradent et, éventuellement, de les convertir dans les nouveaux formats. À cet égard, les offres d'hébergement gratuit ne manquent pas, qu'il s'agisse d'abriter nos photographies, vidéos, documents bureautiques ou encore notre courrier. Le plus souvent, elles sont le fait de grands fournisseurs de services sur Internet, tels *Google*, *Amazon*, *Facebook*, *Apple*, *Microsoft*, etc. Toutefois, les récentes révélations sur le programme de surveillance PRISM de l'Agence américaine de sécurité soulignent un phénomène général : nous payons ces services en fournissant involontairement des informations très détaillées sur nos activités, préférences, relations personnelles, achats, voyages, etc.

Ces données sont employées à notre insu, ce qui se traduit non seulement par une surveillance occulte, mais aussi de plus en plus souvent par des violations à grande échelle de la vie privée, par exemple par des vendeurs de toutes sortes de produits.

D'autres grands acteurs ont lancé un gigantesque effort de préservation des données disponibles sur Internet ou sur certains réseaux sociaux : fondée en 1996, l'*Internet Archive* fait régulièrement des copies des sites disponibles sur la Toile et conserve une vaste collection d'images, de vidéos, de livres et de logiciels. S'agissant des œuvres littéraires, musicales ou cinématographiques, de grandes sociétés se réservent le droit exclusif de les archiver, et interdisent au public de les copier ou de les échanger ; en outre, elles mettent parfois en place des mesures de protection contre la copie qui entravent la conservation de ces données à long terme.

Aussi, tous ces acteurs, aussi puissants soient-ils, finissent-ils par constituer des points de fragilité comparables à la grande bibliothèque d'Alexandrie des temps hellénistiques. Malgré les investissements faits pour créer des centrales de serveurs plus sûres et mieux organisées, la concentration de grandes masses de données sous le contrôle d'une seule organisation est toujours source de fragilité. Et au-delà du problème technique de la sauvegarde des données contre les pertes accidentelles, il



5. L'AGENCE AMÉRICAINE DE SÉCURITÉ (NSA ou *National Security Agency*) exploite la concentration des données personnelles chez les grands fournisseurs de services sur Internet pour surveiller la population. Dans le projet PRISM, révélé par le lanceur d'alerte Edward Snowden, elle se fonde sur la loi américaine régulant les activités d'espionnage pour se fournir en données à décrypter chez les gros opérateurs, tels *Google*, *Amazon*, *Facebook*, etc.

■ L'AUTEUR



Roberto DI COSMO, professeur d'informatique à l'Université Paris-Diderot, est membre du Laboratoire Preuves, programmes et systèmes du CNRS.

s'agit aussi de trier les informations qui méritent d'être préservées et celles qui peuvent être effacées. Or le fait qu'une seule organisation, ayant sa logique et ses intérêts particuliers, en décide, introduit manifestement un énorme risque de destruction. Aurions-nous aujourd'hui des éditions complètes initiales des *Fleurs du mal* de Charles Baudelaire (1821-1867) ou de *Madame Bovary* de Gustave Flaubert (1821-1880), si, il y a plus d'un siècle, alors que ces chefs-d'œuvre étaient censurés, leur conservation avait été confiée à une organisation étatique centralisée ?

L'échange entre pairs

Une solution alternative de conservation des données de l'humanité à long terme provient du très grand nombre des ordinateurs personnels et de la croissance exponentielle de la capacité de stockage des mémoires disponibles pour le grand public. Le déploiement d'un système d'échange pair à pair (de l'anglais *peer to peer*) rend envisageable la transformation du parc mondial d'ordinateurs personnels en une gigantesque base de données distribuées, puis de laisser chacun y répliquer et y partager ses données.

Sous réserve que les ordinateurs personnels continuent à exister (!), cette solution aurait l'avantage de faire l'impasse sur une autorité centrale fixant seule ce qui mérite d'être conservé. Cela permettrait à

En 1085, Guillaume le conquérant ordonna un recensement général en Angleterre, qui devint ensuite la base indiscutable pour établir les taxes royales. Les Anglais nommèrent *Domesday Book* (le livre du jugement dernier) le volume où les résultats étaient consignés (*ci-dessous à gauche*). En 1986, la BBC décida de stocker les données du nouveau *Domesday* – un nouveau recensement et une enquête générale qu'elle avait organisés – sur des vidéodisques de type *LaserDisc*, support que l'on pensait alors pérenne (*au centre*). Toutefois, il se périma si vite que, 25 ans plus tard, il était déjà devenu impossible d'en lire le contenu ! Un long et coûteux effort technique fut alors nécessaire pour reproduire sur un ordinateur moderne le fonctionnement d'un lecteur de *LaserDisc*, ce qui conduisit au *Domesday reloaded* (*Domesday refait*), une nouvelle version consultable (*à droite*) qui, par prudence, a été versée aux Archives nationales britanniques...



l'humanité de décider spontanément des informations qu'elle souhaite conserver... Les années fastes du téléchargement de musique sur les anciens systèmes pair à pair nous ont donné un aperçu de ce qu'un tel réseau peut accomplir : à côté de tubes et autres succès commerciaux plus ou moins faciles, on y retrouvait souvent des morceaux de musique rares et, sinon, indisponibles. Il suffit en effet qu'existe un petit groupe de passionnés par tel ou tel type de données pour que celles-ci soient préservées quelque part dans le réseau social. Cela constitue un atout en faveur de la conservation de données à long terme, étant donné qu'il est difficile d'identifier quelle information pertinente aujourd'hui le sera demain ou, inversement, quelles données apparemment sans intérêt aujourd'hui pourraient se révéler précieuses dans l'avenir.

Comment résoudre les conflits d'intérêt ?

Encore faut-il que l'accès aux données dans la durée soit garanti par l'usage de formats ouverts et de logiciels libres. Dès lors, il serait utile de mettre en place une infrastructure de recherche répartie, comme les données, partout dans le monde, pour que les informations y soient stockées et indexées.

BIBLIOGRAPHIE

J. Rothenberg, *Avoiding technological quicksand : Finding a viable technical foundation for digital preservation*, rapport au *Council on Library and Information Resources*, janvier 1999 (www.clir.org/pubs/reports/rothenberg).

P. et S. Aigrain, *Sharing : Culture and the Economy in the Internet Age*, Amsterdam University Press, 2012.

Th. Priol, *Vers le tout-en-réseau ?*, Dossier *Pour la Science*, n° 66, janvier-mars 2010.

Sur le parc de PC prévisible à l'échelle mondiale : www.gartner.com/id=1602818

Le principal obstacle n'est pas technique, mais économique. Nombre d'acteurs des industries de la musique, du cinéma et du livre militent pour qu'on restreigne les échanges entre pairs et pénalise ceux qui y participent ; fondés sur la vente de copies individuelles, leurs modèles économiques s'écroulent si l'on donne la possibilité de réaliser à volonté un nombre illimité de copies numériques de toute œuvre. C'est pourquoi ils souhaitent l'application de lois contraignant les internautes à se détourner du modèle pair à pair pour passer à des échanges cryptés, lesquels ne peuvent que restreindre l'accès public aux données.

Nous ne pouvons détailler ici les propositions qui ont été faites pour concilier ces intérêts contradictoires, mais il nous paraît essentiel de souligner que le libre partage sans entrave est un instrument essentiel pour la préservation de nos données pour les futures générations. Il doit être encouragé et accompagné. De la même façon que les collections privées de livres ont préservé des titres dont les éditeurs ont disparu, les très nombreuses collections de copies numériques individuelles de données nouvelles ou anciennes seront pour nos descendants un moyen de conserver la culture actuelle, pourvu qu'elles soient enregistrées dans des formats ouverts, avec des logiciels libres. ■

Les mémoires du FUTUR

Comment stocker et traiter le déluge d'informations produites au quotidien ? Les mémoires numériques actuelles se rapprochant de leurs limites, les scientifiques développent de nouveaux dispositifs.

Luca Parniola

En 2012, l'industrie a fabriqué plus de puces électroniques de mémoire que l'agriculture n'a produit de grains de riz ! Une telle production est indispensable pour stocker et traiter les énormes quantités d'informations créées tous les jours. Dans le monde, $2,72 \times 10^{21}$ octets de données ont ainsi été produits en 2012. Les dispositifs actuels sont-ils capables de soutenir la croissance rapide du volume de données ? Cela implique des capacités de stockage importantes et des dispositifs qui permettent d'accéder rapidement aux données. Or les techniques utilisées aujourd'hui semblent proches de leurs limites. Aussi les laboratoires développent-ils une nouvelle génération de mémoires numériques, plus performantes et moins coûteuses.

La nécessité de conserver des informations a débuté très tôt dans l'histoire de l'humanité. Les premiers supports étaient analogiques – parmi les nombreux dispositifs, on trouve les tablettes d'argile, les parchemins ou les livres. Ils avaient pour principal inconvénient d'être encombrants. L'émergence du

L'ESSENTIEL

- Les appareils numériques utilisent des mémoires rapides d'accès pour interagir avec le microprocesseur et des mémoires à grande capacité de stockage, mais plus lentes. Chaque type a ses avantages et ses limites.
- Les mémoires universelles, rapides et capables de stocker de grands volumes de données, n'existent pas.
- Les scientifiques conçoivent de nouveaux dispositifs, fondés sur des effets physiques différents, qui seront rapides d'accès, fiables, économes en énergie...

numérique a bouleversé cet état de fait. Toute donnée – musique, texte, image – est convertie en une séquence de bits (des « 0 » et des « 1 ») qui permet, entre autres avantages, un gain de place important, un traitement automatisé, la consultation à distance, etc.

Le premier exemple de mémoire numérique est celui des cartes perforées utilisées sur les métiers à tisser Jacquard à partir du début du XIX^e siècle. Ces cartes présentaient une suite de trous (un trou correspondant, par exemple, à 0 et son absence à 1). Elles contrôlaient le mouvement des crochets du métier à tisser et ainsi automatisaient la fabrication des textiles. Ces cartes perforées se limitaient à indiquer une succession d'instructions.

En 1834, le mathématicien anglais Charles Babbage a eu l'idée d'utiliser les cartes perforées avec sa machine à calculer. Les cartes contenaient les instructions et les données pour les opérations à effectuer. Cette innovation a conduit à l'invention des ordinateurs et au développement des supports de mémoire numériques. Ceux-



© Vasya Kobelew/shutterstock

ci ont connu un essor important au cours des dernières décennies. Ils représentent aujourd'hui l'essentiel des supports de mémoire. Durant ces 20 dernières années, les performances de ces dispositifs ont aussi encouragé la production massive de données (voir la figure 1). La mise en mémoire étant devenue simple et peu chère, il n'y avait plus besoin de limiter le volume de données à stocker.

Les mémoires universelles : utopiques ?

Si l'accroissement du volume d'informations à stocker se poursuit au même rythme, les mémoires d'aujourd'hui seront-elles à la hauteur, en capacité de stockage comme en rapidité d'accès ? Il est probable que non. Les ingénieurs rêvent de mémoires dites universelles qui seraient notamment rapides d'accès, capables de stocker de grands volumes de données, fiables et économes en énergie. Malheureusement, un tel dispositif relève plus de l'idéal vers lequel tendre que d'une réalité concrète.

Les systèmes numériques d'aujourd'hui associent toujours plusieurs dispositifs : des mémoires dites de travail (ou « mémoires vives »), qui conservent de petits volumes de données mais qui ont des temps d'accès très courts, et des mémoires de stockage, qui peuvent enregistrer d'énormes quantités d'information, mais au prix d'un temps d'accès plus long. En associant ces deux types de mémoires, il est possible de construire un système qui peut exécuter des opérations en peu de temps et conserver un grand volume de données.

De nombreux concepts sont à l'étude et certains pourraient être opérationnels dans les années à venir. Pour comprendre l'intérêt de ces dispositifs, il faut commencer par s'intéresser aux mémoires utilisées aujourd'hui, avec leurs qualités et leurs limites. Je présenterai ensuite quelques solutions techniques qui pourraient succéder aux dispositifs actuels d'ici cinq à dix ans.

Comme je l'ai évoqué, nous ne savons pas fabriquer de mémoires universelles. La nature ne semble pas faire mieux dans ce domaine. Il existe, par exemple, différents modèles théoriques de la mémoire du cer-

veau humain. Dans tous les cas, la mémoire est subdivisée en sous-systèmes spécialisés en différentes tâches. On observe ainsi de nombreuses analogies entre la mémoire biologique et les mémoires artificielles. Dans un modèle classique des années 1960, on distingue la mémoire sensorielle qui retient une grande quantité d'informations provenant des différents sens pendant un temps très bref et sans faire d'interprétation. Certaines informations sont ensuite transmises à la mémoire à court terme, dont la capacité est limitée et où les informations sont disponibles pendant plusieurs dizaines de secondes. Cette mémoire permet, entre autres, à un lecteur de ne pas oublier le début de la phrase pour pouvoir en comprendre le sens.

Enfin, une mémoire à long terme permet de conserver des informations longtemps et en quantités importantes. L'évolution a probablement conduit à cette structure qui permet par exemple de traiter rapidement une information perçue, comme quand nous sommes confrontés à un prédateur ou à un autre danger.

Depuis les années 1960, ce modèle de la mémoire humaine a subi des modifications et s'est enrichi de sous-catégories (voir l'article de Franck Chaillan dans ce numéro). Cependant, la conclusion est que le cerveau n'est pas une mémoire universelle et que la combinaison de mémoires spécialisées semble présenter des avantages.

L'industrie suit un schéma similaire, puisqu'elle associe différents types de mémoires. Si, à plusieurs reprises, de nouveaux dispositifs ont été présentés comme universels, des études approfondies les ont toujours mis en défaut. Quoi qu'il en soit, les fabricants se concentrent sur l'amélioration de leurs caractéristiques : la durée de vie, la fiabilité, etc. Une propriété importante est le temps d'accès, la durée nécessaire pour la lecture ou l'écriture d'une information sur une mémoire ; elle est comprise entre la nanoseconde et la milliseconde, suivant les dispositifs.

Le processeur d'un ordinateur, qui exécute des calculs, a besoin d'obtenir très rapidement les instructions et les données nécessaires à l'opération. Il doit donc être associé à des mémoires de travail à temps d'accès très court, pour ne pas ralentir les performances du processeur. De telles mémoires coûtent cher, et leur densité d'information stockée est faible : elles ne peuvent pas emmagasiner de grands