

ADN,
QUARTZ,
VERRE...

QUEL SUPPORT DE STOCKAGE POUR DEMAIN ?

La capacité des data centers ne suit pas le rythme général de production des données à l'échelon mondial. Il est donc temps de repenser la manière de stocker nos contenus numériques.

Si rien n'est fait, la planète risque de crouler sous les data centers, ces bâtiments qui renferment l'infrastructure informatique (serveurs, routeurs, disques durs, etc.) nécessaire à la conservation et à la distribution des données numériques. La start-up Pearcode, basée au sein de la technopole Sophia-Antipolis, dans les Alpes-Maritimes, en dénombre plus de huit millions aujourd'hui, contre un peu moins de 500 000 en 2012. Et ce nombre pourrait encore exploser à l'horizon 2040. Le biologiste cellulaire François Képès, membre de l'Académie des technologies et coauteur du rapport *Archiver les mégadonnées au-delà de 2040 : la piste de l'ADN⁽¹⁾*, assure qu'un millième des terres émergées pourrait être dévolu au stockage des données à cette date. ■■■

LE GLOBAL MUSIC VAULT

entend conserver toute la musique du monde dans de petites plaques de verre. Son DG, Luc Jenkinson, en présente une au compositeur Alfons Karabuda.

**En 2040,
un millième des
terres émergées
pourrait être
dévolu au stockage
des données**

A photograph of two men standing in a European-style city street. The man on the left is older, with grey hair and a beard, wearing a black turtleneck and a dark jacket. The man on the right is younger, with dark hair and a beard, wearing a blue jacket over a black shirt. He is holding a transparent, disc-shaped object (a 5D holographic disc) between his fingers. The background shows historic buildings and a cobblestone street.

8 millions

C'est le nombre de data centers dans le monde. On en dénombreait seulement 500 000 il y a dix ans.

Source : Pearcode.

**L'holographie (5D)
permet de conserver
jusqu'à 500 To
sur un support aussi
fin qu'un CD**

■ ■ ■ Selon Statista, le monde a en effet produit 79 zettaoctets (Zo) de données – soit 79 milliards de disques durs d'un téraoctet (To) – rien qu'en 2021, devrait avoir consommé 97 Zo en 2022 et certainement... 281 Zo en 2025! Un torrent continu d'informations qu'il faut stocker de manière temporaire ou dans la durée. Les data centers utilisent à cet effet des supports magnétiques, de type disque dur ou bandes, et des SSD. Des supports dont on commence à percevoir les limites et qui présentent l'inconvénient de consommer quantité d'électricité. L'arrivée prochaine des disques durs de 30 To ou la promesse faite par

le laboratoire IBM Research de Zurich d'une bande magnétique capable d'enregistrer 330 To ne font que repousser un peu l'échéance et la nécessité de voir émerger des technologies de rupture.

S'inspirer de la nature

En ce sens, l'ADN pourrait constituer la solution de substitution idéale aux data centers actuels. Après tout, l'acide désoxyribonucléique est ce que la nature a inventé de plus efficace pour coder et préserver l'ensemble du matériel génétique des espèces. Il suffirait donc de maîtriser à grande échelle des techniques de stockage des informations

numériques dans des molécules d'ADN synthétique pour libérer nos sociétés de leur dépendance aux fermes de données énergivores que nous connaissons. Avec l'ADN, les informations constituant internet tiendraient aisément dans une simple boîte à chaussures! La piste est prometteuse et attire chercheurs et entrepreneurs.

En France, le Centre national de la recherche scientifique (CNRS) et un tissu de jeunes entreprises planchent sur la question. Biomemory (voir interview p. 25 et infographie p. 26), Pearcode et DNA Script développent des procédés de biologie de synthèse afin de stocker des données sur des fragments d'ADN artificiels. Outre-Atlantique, Microsoft s'est emparé du sujet en finançant le laboratoire en systèmes d'information moléculaires (MISL) de l'université de Washington. Celui-ci serait parvenu à créer un premier système d'écriture rapide de stockage de l'ADN à l'échelle nanométrique⁽²⁾. Les performances ne font toutefois qu'approcher celles des vieilles cassettes magnétiques (LTO) d'antan, rappelant que le chemin est encore long avant de disposer de solutions susceptibles de détrôner nos bons vieux supports.

Gravé dans le verre

L'ADN n'est cependant pas l'unique option envisagée pour préserver les données dans le futur. Le stockage 5D porte lui aussi de jolies promesses. Développée par des chercheurs de l'université de Southampton, au ■ ■ ■

DATA CENTERS, LE DÉFI ÉNERGÉTIQUE

Chacun des huit millions de centres de données installés à travers le monde héberge des milliers de serveurs et plus encore de disques durs ou de SSD qu'il faut alimenter et refroidir en permanence. Un centre de données de grande taille peut compter jusqu'à un million de serveurs et consommer dix terawattheures par an, soit l'équivalent d'une ville française de 100 000 habitants. Au total, ces équipements monopoliseraient environ 3 à 4 % de la consommation

énergétique mondiale. Leurs besoins en eau pour le refroidissement se révèlent également considérables. En août dernier, le quotidien néerlandais *Noordhollands Dagblad* a même révélé qu'un centre de stockage de données géant, installé par Microsoft aux Pays-Bas, avait englouti 84 millions de litres d'eau en 2021. Une réelle problématique en période de sécheresse. D'importants efforts sont toutefois menés par les acteurs du secteur pour tendre vers davantage de tempérance énergétique.

Ainsi, les infrastructures (sécurité, refroidissement, onduleurs) du centre de données PA10 d'Equinix, implanté à Saint-Denis, près de Paris, nécessitent deux à trois fois moins d'énergie que certaines installations plus anciennes. Par ailleurs, la chaleur dégagée par ces installations est transférée vers le réseau de chauffage urbain de l'opérateur Plaine Commune Énergie, qui servira à chauffer des équipements publics, des bureaux, mais aussi la future piscine des jeux Olympiques 2024!

La piscine olympique de Saint-Denis sera chauffée en partie par les data centers voisins



LA CHAMBRE FORTE MONDIALE

des semences, au cœur
de l'archipel du Svalbard,
en Norvège.

UN HOLOGRAMME POUR L'ÉTERNITÉ

Le stockage holographique, ou 5D, agite
les chercheurs depuis les années 1940.

Cette idée un peu folle, qui consiste à enregistrer
des données dans un bloc de verre au moyen d'un
laser, trouve de premières applications concrètes.

Un mystérieux bloc de béton émerge du flanc d'une montagne de grès, telle une épée plantée dans la glace de l'île du Spitzberg. Les rares visiteurs en poussent les portes blindées pour pénétrer au cœur d'un vaste coffre-fort accueillant la plus grande collection mondiale de graines de plantes et d'arbres, rassemblée dans le cadre du projet gouvernemental norvégien Global Seed Vault. Dans quelques mois, un second sanctuaire sortira de terre, consacré cette fois à la préservation du patrimoine musical mondial. Des trésors audio soigneusement numérisés et sauvegardés sur de fines plaques en verre de quartz au moyen d'un laser. Une technologie née dans les laboratoires du projet Silica de Microsoft, qui « *garantit non seulement un archivage sûr et respectueux de l'environnement, mais aussi que la diversité mondiale de la musique est à portée de main pour que les compositeurs actuels et futurs puissent apprendre et s'en inspirer* », se félicite Alfons Karabuda, directeur du conseil d'administration du Global Music Vault. Chaque plateau mesure sept centi-

mètres de côté et peut stocker jusqu'à 100 Go de données en toute sécurité.

Des supports conçus pour durer

Microsoft explore d'autres pistes pour la sauvegarde longue durée. La technologie baptisée HSD (Holographic Storage Device) exploite un cube de niobate de lithium où sont empilées des couches de données encodées comme des images 2D. Les informations sont gravées dans la structure cristalline trigonale du matériau à l'aide de deux faisceaux laser. En modifiant l'angle de l'un des lasers, il devient possible d'entreposer une multitude de données au sein de la structure du cristal, d'où l'appellation de « stockage holographique ». La lecture des images s'effectue en utilisant l'un des faisceaux laser pour diffracter l'hologramme et recréer le flux de données qui est ensuite lu par une caméra. Le support est réinscriptible, l'effacement des informations étant

Avec ces plaques,
le projet Silica
entend réduire
le volume et le coût
du stockage.

obtenu par rayonnement ultraviolet. Outre la sauvegarde de grandes quantités d'informations dans un volume réduit, le procédé vaut par sa vitesse de lecture. Le laser se déplace à la vitesse de la lumière vers l'emplacement demandé par l'utilisateur pour un accès instantané aux données. Après la scène et les concerts virtuels d'artistes disparus, les hologrammes vont-ils investir les centres de données? Réponse dans quelques années.●



MICROSOFT

La longévité des disques 5D est d'environ 13,8 milliards d'années

... Royaume-Uni, cette technologie consiste à graver les informations dans du verre de silice nanostructuré résistant à la chaleur et aux réactions chimiques à l'aide de courtes pulsations de lumière émises par un femto-laser. Le terme «5D» signifie que l'écriture prend place à la fois sur les trois dimensions spatiales, mais aussi sur deux dimensions optiques. On peut aussi citer le projet Silica⁽³⁾ de Microsoft, qui emploie pour sa part des plaques de verre de quartz où sont gravées des centaines de couches de pixels volumétriques (voxels). Une solution qui réduit sensiblement le volume et les coûts de stockage. Ces projets semblent prometteurs. Mais, en attendant, le temps presse. Or nous pouvons d'ores et déjà réduire notre consommation numérique de manière à ralentir la prolifération des data centers. Dérisoire à l'échelle d'un individu, la chasse aux données inutiles, au moindre mail stocké en trop, s'avère efficace quand elle est partagée par plus de huit milliards d'humains! ● Stéphane Philippon

(1) bit.ly/3ho3WwI - (2) bit.ly/3j39AEI - (3) bit.ly/3j3go4Q

L'ADN SYNTHÉTIQUE LE STOCKAGE DANS LES

Ce procédé de sauvegarde encore balbutiant pourrait constituer le futur du stockage. Bonne nouvelle, les laboratoires et les start-up français comptent parmi les plus avancés en ce domaine.

Et si tout le savoir du monde tenait dans notre salon? Et non plus dans les millions de data centers disséminés à travers le globe? C'est en tout cas l'avenir qu'esquissent les équipes de recherche travaillant sur le stockage de données au sein de brins d'ADN. Une technologie de rupture avec les solutions électroniques et magnétiques que nous connaissons aujourd'hui. L'avantage du stockage ADN tient à sa capacité à conserver d'immenses quantités de fichiers dans un volume réduit. Le biologiste et chercheur François Képès, de l'Académie des technologies, estime ainsi que, en 2040, l'ensemble des données de l'humanité copiées sur un support ADN trouverait place dans un simple semi-remorque! Il reste de nombreux défis techniques à surmonter pour y parvenir, à commencer par la vitesse d'accès aux informations et le coût de revient. Les laboratoires français s'y emploient activement, à l'image du programme exploratoire MolecuArXiv, piloté par le Centre national de la recherche scientifique (CNRS). Classé « programme et équipement prioritaire de recherche », le projet bénéficie d'un budget de 20 millions d'euros sur sept ans de l'État. De quoi financer seize laboratoires impli-

qués dans la chimie de synthèse de l'ADN, parmi lesquels Pearcode, une jeune entreprise établie à Sophia-Antipolis et pilotée par Marc Antonini, directeur de recherche au CNRS, et Melpomeni Dimopoulou, une scientifique lauréate du prix Jeunes Talents 2021 pour les femmes et la science décerné par la Fondation L'Oréal, l'Académie des sciences et l'Unesco.

Repenser le codage des informations

Les travaux sur l'ADN de synthèse consistent, dans un premier temps, à coder un fichier numérique (une photo par exemple) composé d'une suite de 0 et de 1 (on parle de « codage binaire ») dans une molécule d'ADN constituée, elle, de quatre blocs distincts, les nucléotides, nommés A, T, C et G (un système de codage dit « quaternaire »). Il est ainsi possible de stocker la valeur 00 dans le nucléotide A, 01 en C, 10 en G et 11 en T. Une fois le codage effectué, il reste à écrire le tout dans la molécule d'ADN. Synthétique, celle-ci partage les propriétés essentielles de l'ADN biologique, la présence de gène exceptée. La molécule artificielle constituée des informations « quaternisées » de notre image est enfin transférée dans une petite capsule qui la protège de tout

Grâce à l'ADN, la totalité des données numériques mondiales pourrait tenir dans un camion!

GÈNES

contact avec l'oxygène et l'eau. Pour extraire l'information entreposée dans ce disque dur biologique, et donc ficher la photo, la capsule doit passer au travers d'un séquenceur qui lit son contenu et effectue le codage inverse, du mode quaternaire vers le mode binaire, grâce à un algorithme spécialement élaboré afin de minimiser les erreurs de séquençage. Si le procédé fonctionne, ce n'est pour le moment qu'à petite échelle. Melpomeni Dimopoulou estime qu'il faudra au moins dix ans avant de savoir exploiter de longs brins d'ADN pour y stocker d'importants volumes de données. En raison des temps d'accès induits par le séquençage et déséquençage, l'avenir du stockage sur ADN se trouve sans doute dans la sauvegarde de données froides (archivage culturel, conservation des films, données scientifiques, etc.) plus que dans les applications nécessitant de constantes opérations de lecture et d'écriture. Une alternative durable aux disques durs, puisqu'un gramme d'ADN peut contenir autant de données qu'un data center pour une consommation électrique infinitésimale.●

ISABELLE MORISON/HANS LUCAS



« Nous exploitons des mécanismes perfectionnés par le vivant depuis quatre milliards d'années »

TROIS QUESTIONS À ERFANE ARWANI

Fondateur et PDG de Biomemory, une jeune pousse issue d'un laboratoire du CNRS travaillant au stockage des données numériques sur ADN biosourcé.

En quoi consistent les recherches menées par Biomemory ?

Nous entendons apporter une réponse aux enjeux climatiques posés par les centres de données. Nous développons pour cela de nouveaux supports de stockage de données moléculaires à l'impact carbone nul ou négligeable. Afin de ne plus « pétrosourcer » la synthèse et la copie d'ADN, nous nous appuyons sur des technologies de biologie de synthèse créées en interne et qui exploitent des mécanismes perfectionnés par le vivant depuis quatre milliards d'années. Biomemory peut ainsi produire de longs fragments d'ADN biosourcés. L'ADN est purifié avant de subir une dessiccation et d'être déposé sur une lamelle de verre et introduit dans la capsule. L'atmosphère de la capsule ne contient que de l'argon et de l'hélium, ce qui garantit une conservation supérieure à cinquante mille ans sans le moindre apport énergétique. Et dans un volume très faible, puisque ce support présente une densité un million de fois supérieure à celle d'un disque dur ou d'une bande magnétique.

Quels défis posent l'écriture et la lecture de données dans l'ADN ?

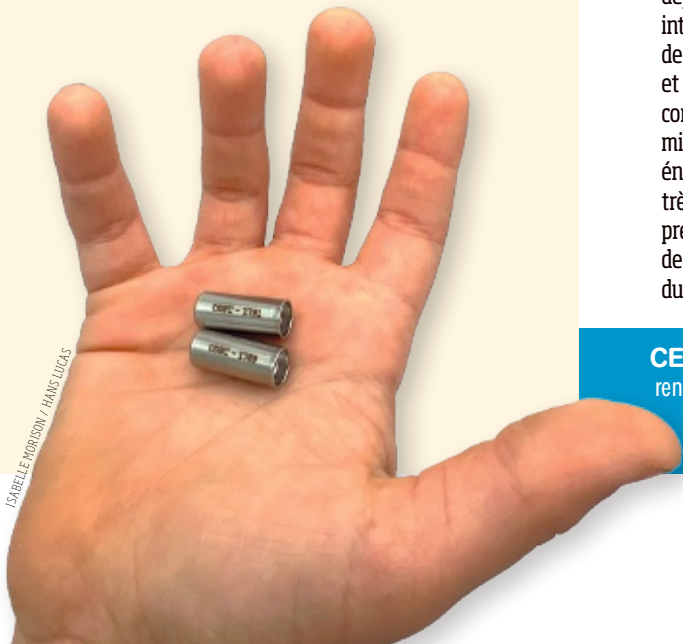
Le principal écueil tient à la production d'ADN à une vitesse et en quantités compatibles avec le stockage des mégadonnées et pour un prix comparable à celui du stockage électronique conventionnel. Pour y parvenir, Biomemory se concentre sur la miniaturisation, l'automatisation et un système d'assemblage d'ADN intégré de bout en bout. Notre solution s'attache également à la biosécurité de l'ADN. Le nôtre ne peut pas coder pour des protéines, même en introduisant des mutations. L'idée est d'empêcher la création d'une toxine potentiellement dangereuse.

Quelles applications imaginez-vous pour cette technologie ?

Biomemory est un pure player du stockage sur ADN. Notre objectif est d'adresser au plus vite les mégadonnées et d'intégrer nos solutions aux centres de données d'ici à 2030. Nous collaborons pour cela avec des acteurs engagés sur le secteur B2B dans le but de valider de façon régulière nos technologies en dehors des laboratoires. Nous nous intéressons notamment à la conservation d'œuvres d'art numériques ou au marquage d'objets à haute valeur ajoutée.

CETTE MINICAPSULE

renferme quantité de données stockées dans des brins d'ADN. Le disque dur de demain ?



ISABELLE MORISON / HANS LUCAS

UNE TECHNO À LA LOUPE

LE DNA DRIVE



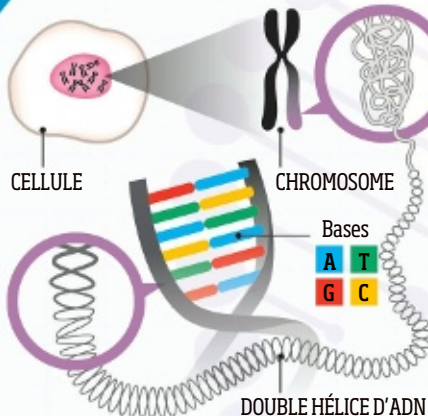
Biomemory, une jeune pousse issue du CNRS, développe une technologie de stockage des données sur ADN des plus prometteuses. Baptisée « DNA Drive », cette solution

biologique présente le double avantage d'être ultracompacte et économe en énergie.

1

POURQUOI L'ADN ?

L'ADN constitue la molécule support de l'information génétique. Il se compose de quatre bases de nucléotides : l'adénine, la thymine, la guanine, et la cytosine, notées A, T, G et C. Sachant que chaque nucléotide se compose d'environ 50 atomes, que chaque cellule détient 3,2 milliards de paires de nucléotides et que notre corps est un ensemble de milliers de milliards de cellules... Biomemory évalue la capacité de stockage du génome d'un individu à 2,7 zettaoctets⁽¹⁾ de données.



Stocké dans une petite capsule métallique, à l'abri de l'eau, de l'air et de la lumière, l'ADN peut rester stable pendant cinquante mille ans.

100 g d'ADN peuvent contenir toutes les données du monde.

LE STOCKAGE À L'HEURE DU BIG DATA ET DE L'INTELLIGENCE ARTIFICIELLE

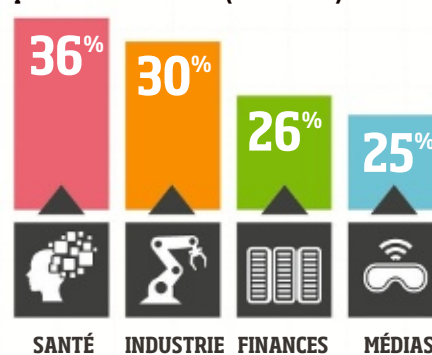
La quantité de données disponible dans le monde croît de façon exponentielle. De 45 Zo en 2020, la datasphère devrait atteindre les 175 Zo dès 2025⁽²⁾.

175 Zo...

C'est
23

fois la distance entre la Terre et la Lune en piles de Blu-ray

Croissance annuelle mondiale des données par secteur d'activité (2018-2025)⁽²⁾



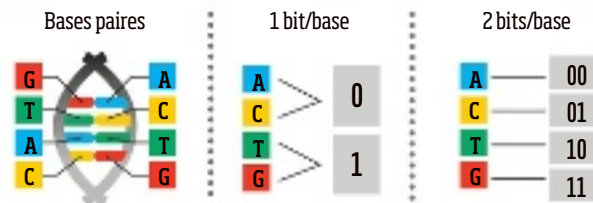
(1) Un zettaoctet (Zo) = un trilliard (10²¹) d'octets ou un milliard de téraoctets. (2) Source : Data Age 2025, IDC, 2018.

2

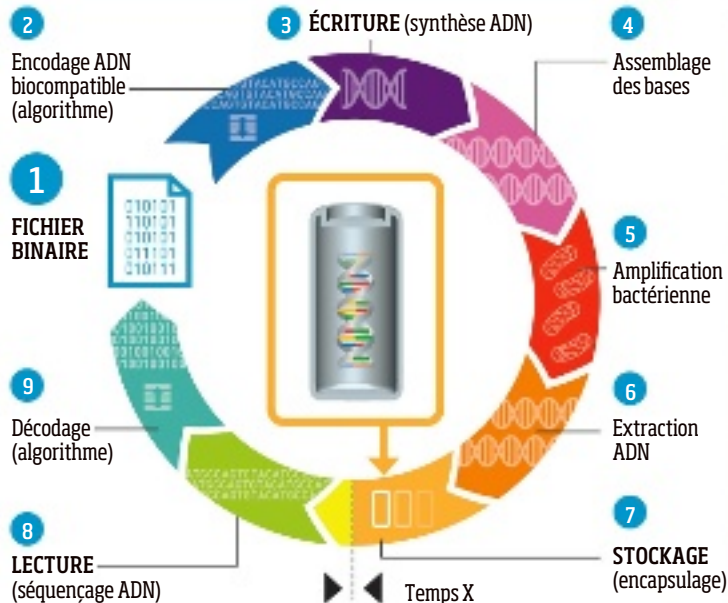
DU NUMÉRIQUE À LA GÉNÉTIQUE

Pour stocker un fichier numérique binaire (fait de 0 et de 1) sur de l'ADN, il faut d'abord le convertir en données quaternaires génétiques (G, A, T, C). Soit en encodant deux bits par base de nucléotide, soit en encodant un bit par base ($A = C = 0$ et $T = G = 1$). Reste à écrire les données sur de l'ADN, en le synthétisant chimiquement, et à encapsuler celui-ci. Un séquenceur ADN permet de lire les données.

Les principaux algorithmes de conversion pour l'encodage



Quand l'écriture et la lecture de données rejoint la biologie



4

UN CYCLE ÉCRITURE/LECTURE ENCORE PERFECTIBLE

L'écriture des données sur ADN reste à l'heure actuelle très complexe, et donc trop longue. Néanmoins, la lecture des données ne nécessite que l'administration d'une simple goutte d'eau dans un séquenceur USB. Certes, avec un SSD, l'accès se fait à la milliseconde près, mais les chercheurs semblent vraiment sur la bonne voie.



3

DEUX TYPES DE SYNTHÈSE

Les premiers essais ont été réalisés sur de petits brins d'ADN synthétisés chimiquement et limités à 200 bases (au-delà, le taux d'erreur serait trop élevé). Au mieux, Microsoft a réussi en 2018 à stocker 200 Mo sur 13,4 millions de brins. L'ADN biologique, lui, peut stocker avec un faible taux d'erreur dans des molécules de grandes tailles, compatibles avec le vivant et répliquables à bas coût.



SIMPLE BRIN

■ CHIMIQUE



DOUBLE HÉLICE

■ BIOLOGIQUE



AVANTAGES ET INCONVÉNIENTS

→ **La densité** : c'est LE point fort de l'ADN, capable de stocker des zettaoctets de données.



→ **Non énergivore** : l'ADN est stable à température ambiante, sans apport énergétique.



→ **Une durée de vie illimitée** : le génome complet le plus ancien connu a 1,6 million d'années.



→ **Un coût encore trop élevé** : tout l'enjeu de l'ADN biologique est de réduire le temps d'écriture.



Boutique > Anciens numéros > MHW N° 990



MHW N° 990

3 €
,10

J'EN PROFITE

-  Consultation sur liseuse PC/Mac gratuite
-  Consultation sur appli iOS/Android gratuite

[< Retour à la boutique](#)

Nos engagements

PAIEMENT ET
DONNÉES SÉCURISÉSSATISFAIT OU
REMBOURSÉSERVICE CLIENT EN FRANCE
FACILEMENT ACCESSIBLEPAR EMAIL
CLIQUEZ ICIPAR COURRIER
01NET SERVICE ABONNEMENTS
45 avenue du Général Leclerc
60643 Chantilly CedexPAR TÉLÉPHONE
01 70 37 31 74