

STOCKAGE DES DONNÉES : LA RÉVOLUTION ADN

Orateur : Marc ANTONINI, Directeur de recherche au CNRS - Laboratoire I3S à l'Université Côte d'Azur | France

Interrogé par **Casimiro VIZZINI**, Médecin, Chargé des Relations Institutionnelles à la Fondation Une Santé Durable pour Tous | France

Le stockage de données sur ADN exploite les potentiels de la nature pour entreposer des données numériques

Le stockage des données sur l'ADN représente une révolution informatique et biologique. Cette technologie ne vise pas à recueillir des informations génétiques mais à exploiter l'intensité et la robustesse de l'ADN pour entreposer des données numériques en son sein. Le processus d'encodage des données dans l'ADN démarre par la conversion des données numériques de format binaire sous un format quaternaire. Ce dernier correspond aux bases azotées de l'ADN : adénine, guanine, cytosine, thymine. Une fois converties, les données sont stockées dans des molécules d'ADN, elles-mêmes placées dans des mini-capsules protégées de l'eau, de la lumière et de l'oxygène. Afin de récupérer les données, l'ADN est réhydraté, amplifié par PCR et séquencé. En sortie de séquenceur, l'information sous forme de code quaternaire est de nouveau formatée en code binaire pour pouvoir être analysée. Cette technologie met la biologie au service de l'informatique et implique donc des travaux pluridisciplinaires.

Le stockage de données sur ADN permet leur archivage et la traçabilité des produits

Des projets de recherche sur cette thématique sont en cours aux Etats unis, dans l'Union Européenne et plus particulièrement en France, qui pourrait d'ailleurs être en capacité de construire une filière souveraine de bout en bout sur le territoire national. Cette technologie peut être utilisée pour stocker des données dites « froides », c'est-à-dire les données d'archivage sans besoin d'accès fréquent ou rapide, qui doivent souvent être conservées pour des raisons légales. Moyennant des accréditations spécifiques, le marquage moléculaire peut aussi être un outil de traçabilité des produits (médicament, matériel, peintures...).

Le stockage sur ADN permettra de dépasser les limites des méthodes de stockage numérique

Le stockage de données sur ADN offre de nouvelles opportunités pour la conservation à long terme de l'information. En effet, l'ADN et les données qu'il contient peuvent être conservés, dans un milieu favorable, pendant des centaines voire des milliers d'années sans altération. A titre de comparaison, la durée de vie d'un disque dur est de 5 à 10 ans, celle d'une bande magnétique de 20 ans. Une mini-capsule est très compacte et peut stocker les informations d'au moins un centre de données. L'ensemble des données de la planète pourrait donc théoriquement être stocké dans un espace équivalent à la surface d'une boîte à chaussure. C'est pourquoi cette méthode de stockage est également plus respectueuse de l'environnement. Enfin, le stockage est pérenne. L'ADN étant stable et robuste, il n'est pas nécessaire de migrer les données d'un support à un autre pour éviter le risque de pertes d'informations.

Des défis technologiques importants à relever pour généraliser l'utilisation de cette technologie

Les freins principaux à la généralisation de cet outil relèvent de son automatisation et de son passage à l'échelle. La synthèse de l'ADN est encore très coûteuse et considérablement plus lente que celle d'un centre de données. Stocker davantage de données numériques, dont le volume est d'ailleurs en croissance exponentielle nécessite donc une accélération du processus de synthèse. Le projet de recherche MolecuArXiv vise à accélérer la vitesse de synthèse de l'ADN synthétique d'un facteur 100 (soit 1 mégabit par seconde) d'ici 3 ans. Un autre objectif à horizon 2030 serait de rattraper la vitesse d'écriture des disques durs. L'amélioration du séquençage de la donnée, afin qu'elle soit récupérable, intègre et lisible est aussi un enjeu important. La problématique de l'accès aléatoire à n'importe quelle donnée stockée doit également être résolue : comment accéder à un fichier en particulier sans séquencer l'ensemble des données de la capsule ?

La capsule n'a pour le moment aucun accès à internet. Cependant dans 5 à 10 ans, des enjeux importants de sécurité devront être pris en compte lorsque des systèmes de stockage sur ADN automatiques connectés à internet émergeront. Cela nécessitera une cryptographie du signal binaire ou des données quaternaires voire la mise en place d'un cloud privé interne avec accès limité aux membres de l'organisation propriétaire des données.

Enfin, les enjeux éthiques sont peu nombreux concernant cette technologie. En effet, l'ADN étant synthétique, seules des informations binaires codées peuvent y être stockées contrairement à des séquences génomiques ou des virus.