

DNA Drive La Révolution de l'ADN

Première mondiale :
dépôt d'archives numériques
encodées sur ADN
aux Archives nationales



SOMMAIRE

1
Introduction : les limitations du stockage numérique

2
Vers une autre forme de stockage ?

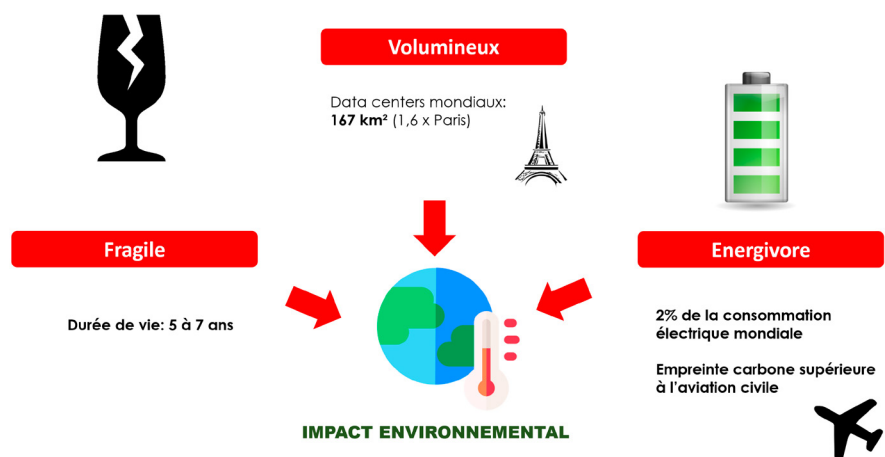
3
**DNA Drive : une stratégie bio-inspirée
de stockage sur ADN**

4
Le projet « La Révolution de l'ADN »

5
L'équipe projet et les partenaires

1 – Introduction :

Les limitations de nos technologies de stockage numérique



Chaque seconde, 29 000 giga-octets (Go) d'informations sont publiées sur internet, soit 2,5 exa-octets par jour ou 912,5 exa-octets par an¹. Le volume de la datasphère, l'ensemble des données numériques mondiales, est aujourd'hui estimé à 45 zetta-octets (Zo) et, au vu de sa croissance vertigineuse, devrait atteindre 175 Zo en 2025². Un zetta-octet est un volume considérable de données correspondant à 10²¹ octets soit 1 milliard de téra-octets (To). A titre d'exemple, avec une connexion rapide (fibre optique, 100 Mbit/s), il faudrait 2,5 millions d'années pour télécharger 1 Zo ! Parmi l'ensemble des données stockées dans le monde, 60 % sont des données « froides », c'est-à-dire des archives qui ne sont pour la plupart jamais relues ou très rarement³. Ces données froides sont stockées en général sur des supports moins performants et moins coûteux tels que les bandes magnétiques. Ces données froides peuvent être remises en ligne ou récupérées à la demande. Le stockage et l'archivage des données numériques sont des enjeux stratégiques critiques pour l'économie, la pérennité et la sécurité de nos sociétés. Cependant, ils se heurtent aujourd'hui à la concomitance de trois facteurs de limitation majeurs :

- la faible **durée de vie des supports** de stockage
- l'**expansion vertigineuse des données** qui induit une demande de stockage largement supérieure à nos capacités
- la **quantité d'énergie** gigantesque requise pour ce stockage, induisant un coût économique et un impact environnemental considérables⁴.

Les supports numériques actuels des *data centers*⁵ sont fragiles, volumineux et énergivores. Qu'il s'agisse de supports optiques (DVD, Blu-Ray), de bandes magnétiques, de disques durs ou de mémoire flash, leur durée de vie est limitée et ils sont donc remplacés tous les 5 à 7 ans en moyenne. Ainsi, les données doivent être régulièrement copiées sur des supports neufs fiables. De plus, leur fonctionnement à température et humidité constantes implique une infrastructure coûteuse associée à une empreinte environnementale colossale. Les *data centers* consomment 2% de la production électrique mondiale et leur empreinte carbone a dépassé celle de l'aviation civile. En outre, la surface occupée par les *data centers* ne cesse de croître car les supports sont volumineux. Si l'on stockait les 175 Zo prévus pour 2025 sur des disques Blu-Ray, le support optique le plus dense dont on dispose, la pile de disques obtenue représenterait 23 fois la distance de la Terre à la Lune.

Outre cette empreinte environnementale, les supports actuels ne sont pour autant largement plus suffisants pour stocker nos données. Depuis 2010, la demande est supérieure à l'offre de stockage⁶ et les systèmes classiques de stockage ne sont pas soutenables à terme : « *Si nous sommes aujourd'hui capables de stocker environ 30% des informations que nous générons, dans seulement 10 ou 20 ans, nous ne pourrions plus en stocker qu'environ 3%* », déclarait Dr. Karin Strauss de Microsoft Research, en octobre 2018⁷.

Ainsi, la transformation numérique de nos sociétés (intelligence artificielle, *Big Data*, objets connectés, voitures autonomes, informatique quantique...) impliquera une évolution technologique majeure de nos systèmes de stockage de données.

2 – Vers une autre forme de stockage ?

Il existe dans la nature une forme de stockage des données qui n'a pas été inventée par l'homme, mais qui se perfectionne depuis près de 4 milliards d'années : l'ADN⁸, le support de l'information génétique. La plupart des êtres vivants utilisent une double hélice d'ADN (deux molécules différentes entrelacées) pour stocker l'information en utilisant quatre briques appelées nucléotides et dont la partie variable est symbolisée par les lettres A, T, C et G (l'adénine, la thymine, la cytosine et la guanine). Le génome humain contient 3,2 milliards de nucléotides, ce qui correspond à l'équivalent de 700 Mo de données dans chaque cellule de notre corps.

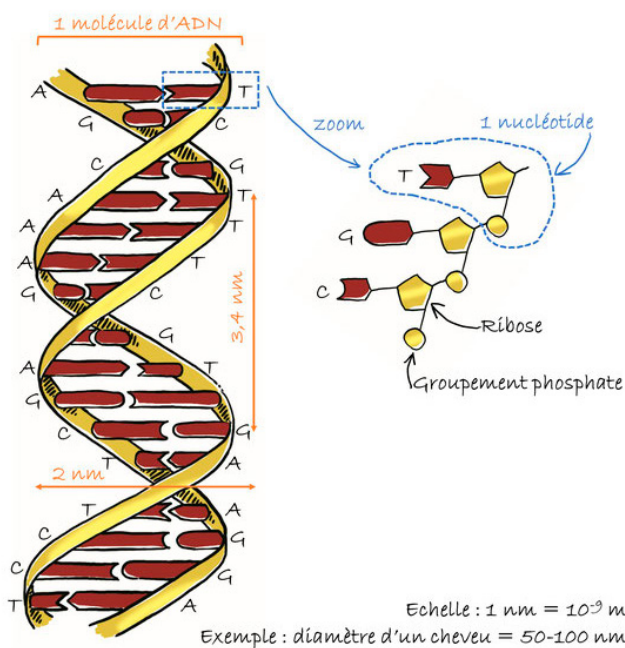


Schéma de la structure de l'ADN
Illustration de Jeanne Le Peillet.

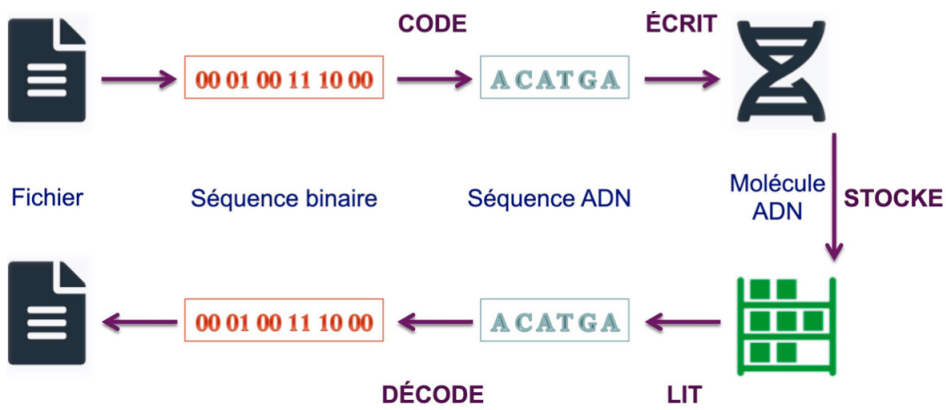
Le stockage d'information numérique sur ADN est une technologie émergente qui constitue une solution :

- **durable** : la stabilité de l'ADN se compte en dizaines, voire en centaines de milliers d'années, ce qui est incomparable avec celle des supports numériques actuels. Un génome complet a par exemple été obtenu à partir d'une défense de mammoth âgée de plus d'un million d'années⁹.

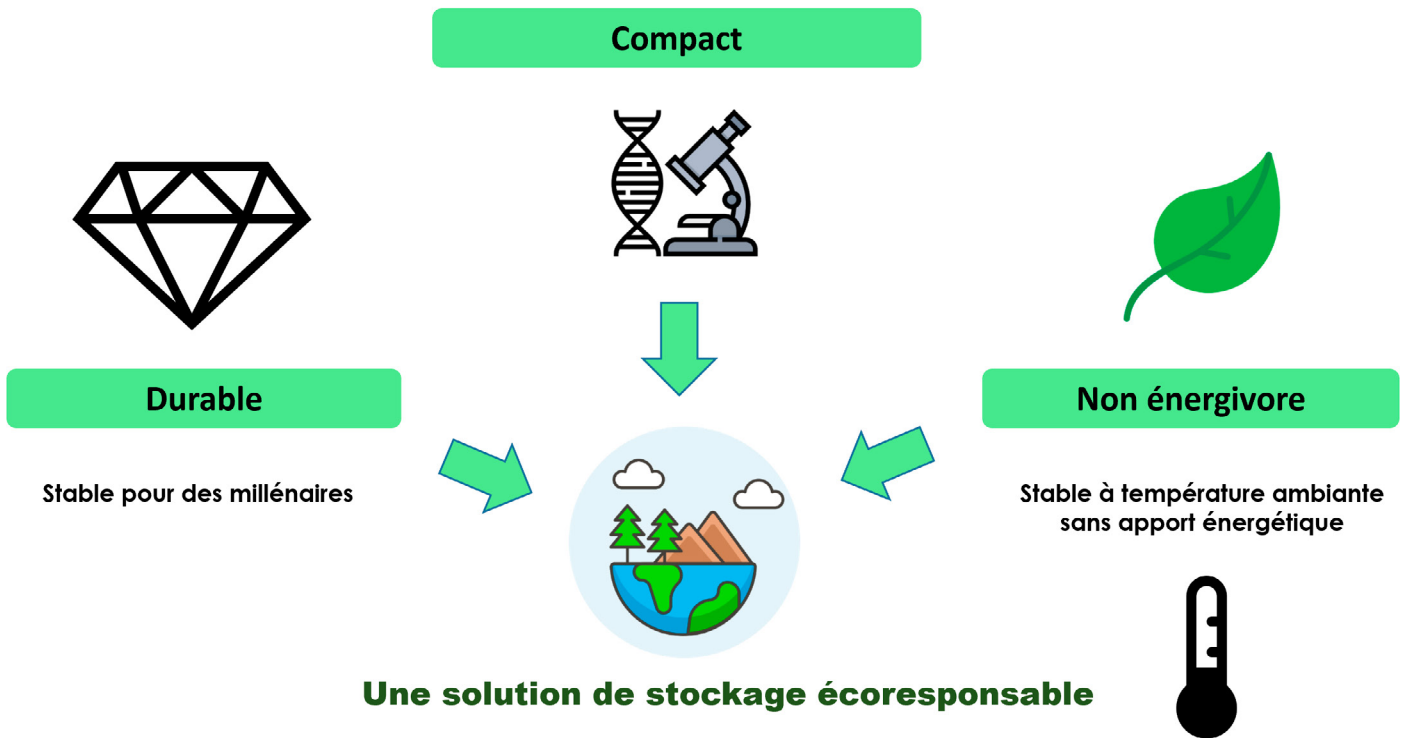
- **non énergivore** : l'ADN est stable à température ambiante sans aucun apport d'énergie s'il est conservé dans des conditions adéquates (sans eau, ni air, ni lumière).

- **extrêmement compacte** : avec une densité maximale de 450 millions de To par gramme d'ADN (0,45 Zo/g), l'intégralité des données mondiales pourrait tenir dans 100 g d'ADN, soit le volume d'une tablette de chocolat.

L'idée d'utiliser l'ADN comme support d'information numérique avait été suggérée dès 1959 par Richard Feynman, lauréat du prix Nobel de physique de 1965¹⁰. Il aura fallu cependant attendre 2012 pour que la première démonstration significative d'une telle utilisation soit publiée par Georges Church (*Harvard University*)¹¹. Les algorithmes d'encodage qu'il a utilisés ont été améliorés (correction d'erreur, accès direct, compression) plus récemment¹². Microsoft Research, l'université de Washington et la société Twist Bioscience (un des leaders mondiaux de la synthèse d'ADN) ont combiné ces technologies pour stocker 200 Mo de données sur ADN¹³. L'ensemble de ces travaux ont tous utilisé la même méthodologie consistant en la transformation d'une donnée numérique binaire (0 ou 1) en donnée quaternaire (A, T, C, G). À cette étape, un code est utilisé et peut varier d'une étude à l'autre (par exemple : 0 = [A ou C] et 1 = [T ou G] pour un code à 1 bit/base ou A=00, C=01, T=10 et G=11 pour un code à 2 bits/base). La séquence ainsi générée est ensuite synthétisée chimiquement sur des petits fragments d'ADN appelés oligonucléotides¹⁴. Ces molécules d'ADN simple brin¹⁵ ont une taille limitée à environ 200 nucléotides en raison des limitations de la technologie utilisée pour la synthèse chimique de l'ADN. Du fait de leur taille réduite, il est nécessaire d'utiliser des millions d'oligonucléotides différents pour stocker l'information (13,4 millions d'oligonucléotides pour 200 Mo encodé sur ADN¹⁶). Des index sont intégrés dans chaque oligonucléotide afin de pouvoir les remettre dans l'ordre après les avoir lus et ainsi pouvoir retranscrire les données.



Méthodologie générale de stockage d'information numérique sur ADN



Une solution de stockage écoresponsable

Ces molécules d'ADN peuvent être conservées de différentes manières (sèches sur du papier, dans un tube plastique, dans une capsule métallique, etc.). La lecture de l'information peut être effectuée avec des séquenceurs d'ADN, similaires à ceux utilisés en biologie et en médecine pour séquencer les génomes des êtres vivants. Une fois la succession de lettres obtenue, il ne reste qu'à la reconvertir en données binaires, en utilisant le même code que celui ayant servi à l'écriture, pour récupérer l'information numérique. Alors que les technologies basées sur des oligonucléotides ont permis de démontrer la faisabilité du stockage numérique sur ADN, elles présentent des limitations, notamment un coût d'écriture élevé et des possibilités d'édition et de copie limitées et coûteuses.

Afin de lever ces limitations et de développer des systèmes de stockage sur ADN plus performants, l'équipe de Stéphane Lemaire, directeur de recherche CNRS au Laboratoire de biologie computationnelle et quantitative (CNRS/Sorbonne Université) cherche à tirer parti du potentiel offert par le vivant et utilise la biologie pour répondre aux enjeux du stockage d'information sur ADN. En effet, alors que les technologies de stockage publiées sont toutes basées sur des méthodes chimiques, physiques et mathématiques, la piste biologique n'a pas encore été explorée. Des technologies très précises de manipulation de l'ADN existent pourtant dans le vivant (lecture, copie, édition, correction d'erreurs, accès direct, amplification du signal, etc.). La stratégie de l'équipe consiste à domestiquer et adapter ces technologies du vivant *via* des approches de biologie synthétique¹⁸ pour développer de nouvelles technologies de stockage sur ADN.

3 – DNA Drive : une stratégie bio-inspirée de stockage sur ADN

Des travaux de recherche menés dans des laboratoires du CNRS et de Sorbonne Université ont permis de développer la technologie DNA Drive. Il s'agit d'une stratégie bio-inspirée d'archivage, en rupture avec les technologies existantes, puisqu'elle utilise des technologies biologiques et n'est pas basée sur des oligonucléotides. L'information est stockée, comme dans le vivant, sur des longs fragments d'ADN en double hélice, appelés plasmides ou chromosomes. La lecture de l'information peut être effectuée, comme pour les oligonucléotides, avec des séquenceurs d'ADN, et notamment ceux de dernière génération (Next-Gen Sequencing, 3^e génération), capables de lire des longs fragments. Par exemple, le séquenceur de poche MinION (Oxford Nanopore Technologies), dont la taille est comparable à une clé USB, permet une relecture nomade d'un DNA Drive.

Les molécules d'ADN du DNA Drive sont conçues pour être biocompatibles, c'est à dire manipulables par des cellules vivantes, par exemple pour copier ou éditer l'information numérique. De plus, elles sont biosécurisées, c'est-à-dire que l'ADN est crypté afin de ne porter aucune information génétique potentiellement dangereuse pour l'être humain ou pour l'environnement. La conversion des données numériques est assurée par un algorithme propriétaire permettant de générer des séquences ADN au format DNA Drive, biocompatibles et biosécurisées. Grâce à ce format, les données peuvent être stockées sur des grandes molécules d'ADN double brin, compatibles avec leur manipulation biologique, à l'inverse des oligonucléotides qui sont des petites molécules simple brin non compatibles avec les systèmes biologiques.

« Le DNA Drive peut être copié biologiquement. Cela reviendrait presque à transformer du sucre en disque dur »

Stéphane Lemaire,
directeur de recherche CNRS.

Le DNA Drive permet également d'organiser physiquement les données numériques à la manière de l'organisation physique d'un disque dur qui comprend plusieurs plateaux, contenant chacun un ensemble de pistes circulaires (*tracks*) où l'information est organisée en blocs de données appelés secteurs. De la même façon, les données du DNA Drive sont organisées physiquement en secteurs d'ADN consécutifs regroupés sur une très grande molécule d'ADN circulaire correspondant à une piste. Des milliers de ces molécules sont rassemblées au sein d'un « *pool* », équivalent à un plateau, et 96 « *pools* » sont rassemblés pour former un « *array* » (équivalent à un disque de 96 plateaux). Il est possible d'ajouter autant « *d'arrays* » que nécessaire pour étendre la capacité de stockage. Cela rend la capacité du DNA Drive modulable et illimitée.

Le DNA Drive est compatible avec tout type d'information numérique, puisque l'algorithme d'encodage utilise des données binaires. Il est donc possible d'encoder tout type de systèmes de fichiers avec des tables d'allocations, des répertoires, des métadonnées, etc. De plus le DNA Drive permet un accès direct (*random access*) aux données car les secteurs ADN sont bordés de blocs d'index, permettant de relire spécifiquement chaque secteur sans avoir besoin de relire l'intégralité des données.

L'ensemble de ces propriétés permettent de définir un nouveau standard pour le stockage d'information numérique sur ADN. Un brevet protégeant la technologie DNA Drive a été déposé à l'Office européen des brevets et est en cours d'extension mondiale¹⁹.

4 – Le projet

« La Révolution de l'ADN »

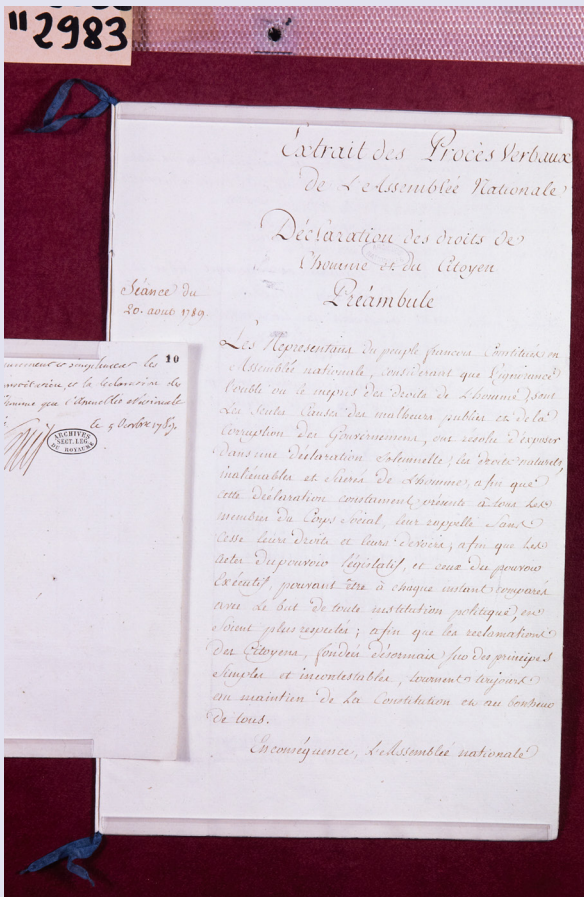
La naissance du projet racontée par **Stéphane Lemaire**, directeur de recherche CNRS au Laboratoire de biologie computationnelle et quantitative (CNRS/Sorbonne Université), membre de l'équipe projet :
« Tout a commencé par un article publié en 2018 dans le journal de l'association étudiante Alma Mater. Dans ce journal, distribué sur le campus de Sorbonne Université, il y avait un article sur la technologie de stockage sur ADN. En lisant cet article, j'ai indiqué aux étudiants de cette association que mon équipe maîtrisait les technologies de manipulation de l'ADN et pourrait donc encoder de l'information sur ADN. Ils m'ont alors mis au défi de le prouver. Je leur ai répondu que j'étais prêt à le faire mais uniquement pour encoder des données importantes et ayant du sens. Ils ont donc réfléchi et sont revenus me voir après quelques semaines pour me dire : nous avons trouvé, il faut encoder la Déclaration des droits de l'homme et du citoyen de 1789. Je leur ai répondu que c'était une excellente idée mais qu'il fallait aussi encoder la Déclaration des droits de la femme et de la citoyenne de 1791. »

Ainsi est né le projet « La Révolution de l'ADN ». Dans un premier temps il a permis de rassembler une petite équipe de biologistes et d'informaticiens qui a conçu et breveté la technologie DNA Drive avec des financements d'amorçage de Sorbonne Université et du CNRS puis de la SATT Lutech. L'équipe a ensuite poursuivi un double objectif :

- obtenir une preuve de concept de l'efficacité de cette technologie de stockage, le DNA Drive ;
- encoder sur ADN des textes fondateurs français à forte valeur symbolique et historique.

Le projet « La Révolution de l'ADN », profondément pluridisciplinaire, a impliqué des historiens, philosophes, informaticiens, et biologistes. L'un des textes sélectionnés a été choisi parmi ceux inscrits au programme Mémoire du monde de l'UNESCO : la Déclaration des droits de l'homme et du citoyen de 1789. Il a semblé essentiel à l'équipe de l'accompagner de la Déclaration des droits de la femme et de la citoyenne, texte non institutionnel rédigé par Olympe de Gouges²⁰ et publié en 1791. L'association de ces deux déclarations véhicule des valeurs humanistes auxquelles l'équipe est attachée.

« Pouvoir mettre le texte d'Olympe de Gouges au même niveau que la Déclaration des droits de l'homme et du citoyen, 230 ans plus tard, est un réel succès sur les plans de l'égalité femme-homme, de la reconnaissance du travail d'Olympe de Gouges et corrigeant une erreur de l'Histoire », se félicite Pierre Crozet, maître de conférences à Sorbonne Université, également membre de l'équipe projet.



La Déclaration des droits de l'homme et du citoyen

Adoptée le 26 août 1789 et placée en préambule de la Constitution de 1791, la Déclaration des droits de l'homme et du citoyen est née des discussions de l'Assemblée constituante. Profondément lié au contexte révolutionnaire, ce texte fondateur abolit l'Ancien régime et pose les bases de la société française et des différents régimes politiques qui se succèdent. Universel, il connaît un retentissement international et s'impose définitivement à la postérité.

La Déclaration des droits de l'homme et du citoyen de 1789 est inscrite en 2003 par l'UNESCO au Registre « Mémoire du monde » qui recense le patrimoine documentaire présentant un intérêt international et une valeur universelle exceptionnelle.

La Déclaration des droits de l'homme et du citoyen est présentée aux Archives nationales, jusqu'au 3 janvier 2022, dans le cadre du nouveau cycle *Les Essentiels*.

Déclaration des droits de l'homme et du citoyen : extrait du procès-verbal de la séance de l'Assemblée constituante du 2 octobre 1789 et billet de la main du roi Louis XVI du 5 octobre 1789. Archives nationales, AE/II/2983 et AE/II/2982. Inscrite sur le Registre « Mémoire du monde » de l'UNESCO depuis 2003

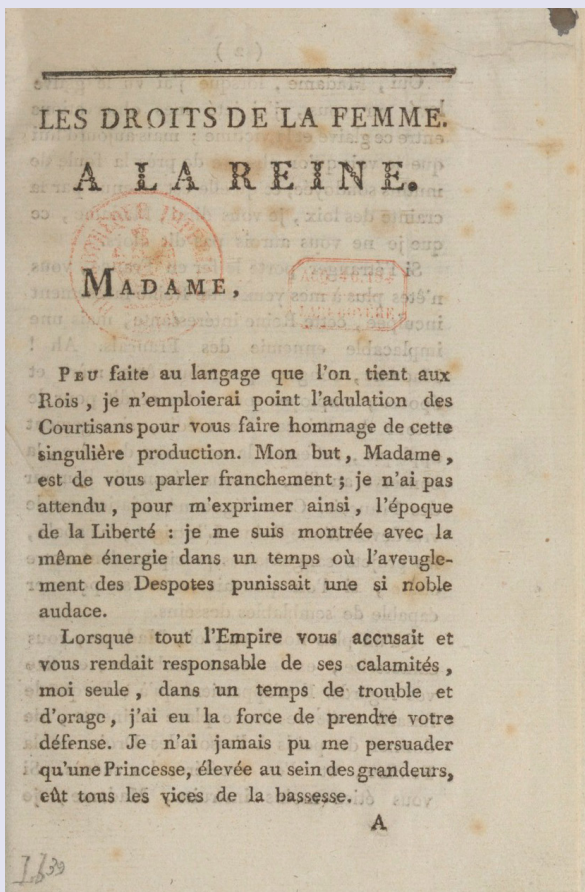


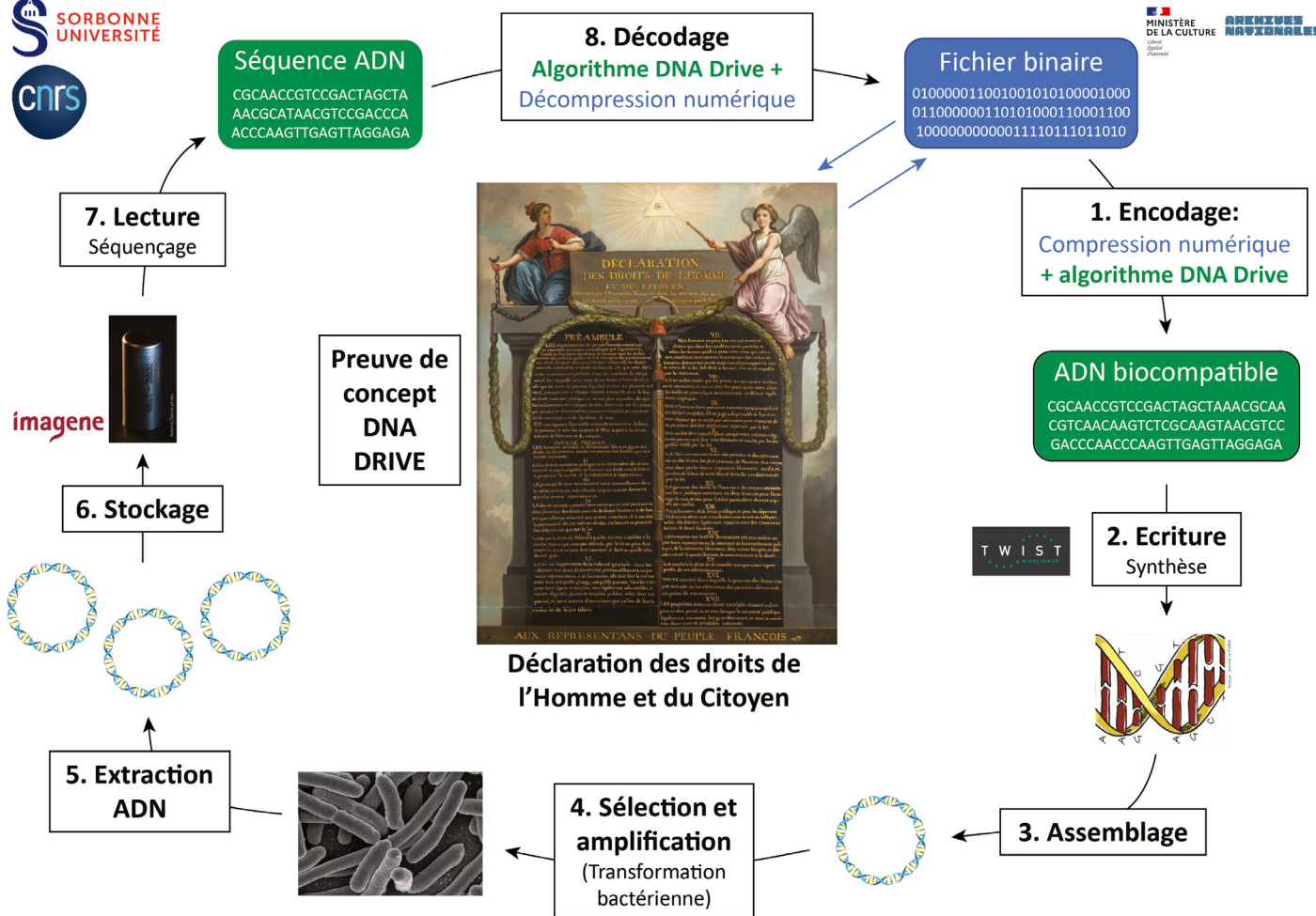
La Déclaration des droits de la femme et de la citoyenne

C'est à la mi-septembre 1791 qu'Olympe de Gouges a publié sa Déclaration des droits de la femme et de la citoyenne. Elle repose sur l'idée que les femmes, qui possèdent toutes les facultés intellectuelles, ont par nature les mêmes droits que les hommes. La Nation étant définie comme « la réunion de la femme et de l'homme » (article 3), elle en déduit que « la Constitution est nulle si la majorité des individus qui composent la nation n'a pas coopéré à sa rédaction ». La déclaration d'Olympe de Gouges passa presque inaperçue et les écrits féministes des années suivantes, comme ceux du XIX^e siècle, ne s'y référeront pas. Mais la forme de ce texte, celle d'une déclaration des droits, est unique à son époque et lui confère une force qui expliquera son succès tardif dans la seconde moitié du XX^e siècle

Portrait d'Olympe de Gouges, attribué à Alexander Kucharsky, fin du XVIII^e siècle (collection privée)

Première page de la brochure *Les droits de la femme* dans laquelle est publiée la *Déclaration des droits de la femme et de la citoyenne*, par Olympe de Gouges, septembre 1791 (Gallica.bnf.fr / Bibliothèque nationale de France)





Procédé du DNA DRIVE appliqué à la Déclaration des droits de l'homme et du citoyen

Ce projet a été mené dans le cadre d'un partenariat étroit entre l'équipe de recherche et les **Archives nationales**, qui conservent l'original de la Déclaration des droits de l'homme et du citoyen. L'objectif était de permettre aux Archives nationales d'enregistrer officiellement ces deux textes encodés sur ADN, une première mondiale pour une institution publique.

Le projet a été également mené en partenariat avec **Twist Bioscience**, entreprise américaine spécialiste de la synthèse d'ADN, et **Imagene**, entreprise française spécialiste de la conservation à long terme de l'ADN.

Les textes ont d'abord été encodés sur ADN avec l'algorithme DNA Drive après une compression numérique, similaire à celle utilisée pour les fichiers zip. Twist Bioscience a ensuite synthétisé des fragments d'ADN qui ont été assemblés et organisés par l'équipe de recherche sur des grandes molécules d'ADN double brin. Ces molécules ont été amplifiées biologiquement pour en obtenir plus de 100 000 milliards de copies. Les molécules d'ADN ont ensuite été extraites, purifiées puis encapsulées sous forme lyophilisée dans des capsules DNASHell© par Imagene.

L'extérieur de ces capsules est en acier inoxydable, l'intérieur en verre, et elles permettent d'encapsuler l'ADN sous atmosphère inerte. Cette technologie permet de préserver l'ADN des dommages du temps en le protégeant de la lumière, de l'air et de l'oxygène, et de le conserver sans aucun apport d'énergie à température ambiante. La durée de conservation de l'ADN est estimée à plus de 50 000 ans dans ces capsules. Chaque capsule peut contenir une quantité d'ADN correspondant à 5000 To de données numériques.

L'ouverture d'une capsule permet de récupérer l'ADN en y ajoutant une goutte d'eau, afin de le réhydrater. Pour relire l'information, un peu de l'ADN est déposé dans un séquenceur. Une fois la séquence d'ADN relue, l'algorithme DNA Drive est utilisé pour la convertir en information binaire qui est ensuite décompressée numériquement afin de retrouver le ou les fichiers d'origine.



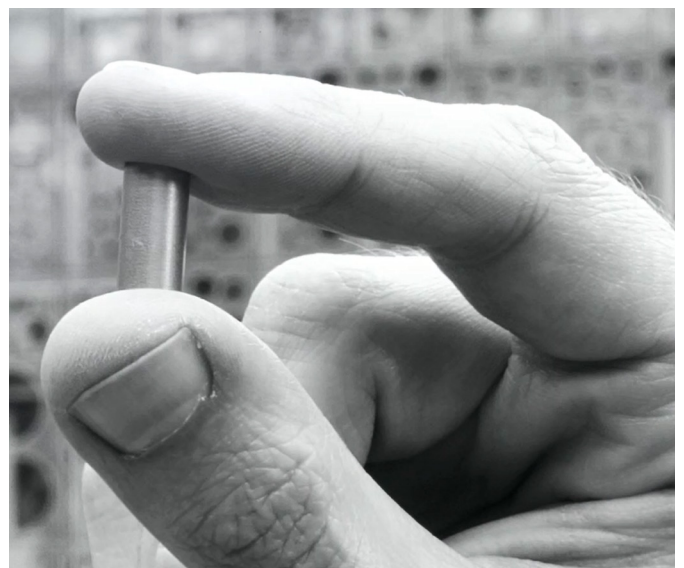
Dans le cadre du projet « La Révolution de l'ADN », deux DNA Drive différents ont été créés : un pour stocker la Déclaration des droits de l'homme et du citoyen de 1789 et l'autre pour la Déclaration des droits de la femme et de la citoyenne de 1791. Chaque DNA Drive a été stocké dans des capsules DNAshell© distinctes. Chaque capsule contient plus de 100 milliards de copies du fichier sous forme d'ADN. À plusieurs reprises, les chercheurs ont pu ouvrir des capsules, séquencer l'ADN contenu et vérifier que l'on pouvait récupérer le fichier avec une fidélité de 100%. Ces deux textes historiques sont donc désormais conservés pour des milliers d'années.

Capsules contenant les deux textes encodés sur ADN. Photo : Stéphane Lemaire / CNRS – Sorbonne Université

Ainsi, le projet a permis d'obtenir une preuve de concept de la fonctionnalité de la technologie DNA Drive. Les capsules contenant les deux DNA Drive ont été exposées lors de la conférence de presse du 23 novembre 2021, et officiellement été enregistrées par les Archives nationales comme premières archives sous forme ADN. Elles sont désormais conservées dans l'Armoire de Fer qui contient les plus précieux documents des Archives nationales.

« Je suis persuadé que le DNA Drive est la preuve que la biologie synthétique va changer notre relation au monde, et pour le meilleur »

Pierre Crozet



Chaque capsule peut contenir une quantité d'ADN correspondant à 5000 To de données numériques.
Photo : Philippe Tran

5 – L'équipe projet et les partenaires

L'équipe de recherche

L'équipe de Stéphane Lemaire, au Laboratoire de biologie computationnelle et quantitative (CNRS/Sorbonne Université) développe des approches de biologie synthétique depuis plusieurs années. Elle a récemment développé un kit de clonage modulaire (MoClo) composé de 119 biobriques d'ADN pour la microalgue *Chlamydomonas reinhardtii*²¹. Ce kit permet la construction rapide de souches reprogrammées génétiquement à la fois pour la recherche fondamentale et les approches biotechnologiques basées sur des microalgues. Le kit permet un assemblage directionnel de multiples brins d'ADN en fragments plus grands, qui peuvent eux-mêmes être assemblés en mégafragments. L'équipe a mis à profit ce savoir-faire autour de la manipulation d'ADN pour initier le développement d'une nouvelle technologie de stockage numérique sur ADN.

L'équipe projet

Les recherches sur le stockage d'information numérique ont été mises en œuvre par une équipe projet coordonnée par Stéphane Lemaire, directeur de recherche au CNRS, et Pierre Crozet, maître de conférences à Sorbonne Université, travaillant dans la même équipe.

Ces travaux de recherche sont menés au Laboratoire de biologie computationnelle et quantitative du CNRS et de Sorbonne Université, dirigé par Alessandra Carbone au sein de l'Institut de biologie Paris-Seine mais ont été initiés en 2018 au Laboratoire de biologie moléculaire et cellulaire des eucaryotes (Sorbonne Université/CNRS), sous la direction de Teresa Teixeira, où travaillaient alors les deux chercheurs.



STÉPHANE LEMAIRE

Directeur de recherche CNRS

Expert en biologie moléculaire, biologie synthétique, coordination et management de programmes scientifiques

Parcours professionnel

Passionné par la biologie et l'informatique, Stéphane s'est orienté vers la génétique et la biologie moléculaire au cours de ses études universitaires. Il a obtenu un doctorat en biologie végétale puis, après un post-doctorat à l'université de Genève, a été recruté au CNRS en 2001. Depuis 15 ans, il dirige une équipe de recherche et est devenu un spécialiste mondialement reconnu de la biologie synthétique, de la photosynthèse et des microalgues. Stéphane est l'auteur de plus de 100 publications de rang international et a été distingué par la médaille de bronze du CNRS en 2006. Il a exercé de nombreuses responsabilités administratives et scientifiques. Il a coordonné de multiples projets scientifiques et obtenu de nombreux financements français et européens pour développer ses travaux de recherche. Depuis 2015, Stéphane a choisi de réorienter ses travaux de recherche vers la biologie synthétique. Dans ce cadre, il a développé des projets innovants pour le stockage numérique sur ADN qui ont abouti à la technologie brevetée DNA Drive et à la création de la start-up Biomemory.

Motivations

L'émergence du champ disciplinaire de la biologie synthétique est un tournant majeur pour la biologie du XXI^e siècle. C'est également un tournant industriel majeur et Stéphane est convaincu que la biologie synthétique permettra la prochaine révolution industrielle, à même de répondre aux défis sociétaux et environnementaux des prochaines décennies. La technologie DNA Drive s'inscrit complètement dans ces perspectives technologiques à même de changer profondément nos sociétés pour apporter une solution éco-responsable aux problèmes environnementaux et énergétiques posés par le stockage numérique.



PIERRE CROZET

Maître de conférences, Sorbonne Université
Expert en biologie moléculaire, biologie synthétique,
formation et management de personnels scientifiques

Parcours professionnel

Passionné par la recherche depuis toujours, Pierre s'est vite orienté vers la biologie mais a conservé un vif intérêt pour l'histoire. Après un doctorat en biologie végétale, il est parti à l'étranger (Portugal) pour s'enrichir tant sur le plan personnel que professionnel (chercheur pendant 5 ans à *Instituto Gulbenkian de Ciência*, au Portugal). Il est ensuite rentré en France pour développer la biologie synthétique des microalgues avec Stéphane. Depuis 2017, il est maître de conférences à Sorbonne Université et couple ses activités de recherche, dont l'encodage d'information numérique sur ADN, avec un enseignement varié allant de l'agroalimentaire à la biologie synthétique. De par ses activités, il collabore avec les cofondateurs des startups de biologie synthétique, Gourmey et Neoplants. Il est membre du conseil scientifique de Neoplants. Aux côtés de Stéphane Lemaire, il est co-fondateur de la start-up Biomemory.

Motivations

Pierre est convaincu que la biologie synthétique va changer le monde, comme la chimie de synthèse l'a fait il y a deux siècles. Il pense que les aspects les plus préoccupants du monde d'aujourd'hui sont le changement climatique et l'instabilité énergétique, et travaille pour tenter d'y répondre. Il est sûr que le stockage d'information sur ADN est la solution pour répondre de façon écologique au défi sociétal que représente le stockage des données numériques.

Les autres membres de l'équipe projet

L'équipe projet a impliqué, en plus de Stéphane Lemaire et Pierre Crozet, des chercheurs et chercheuses, ainsi que des ingénieurs et ingénieures en biologie et en bioinformatique, membres des deux laboratoires : Clémence Blachon, Nicolas Cornille, Mariette Gibier, Achille Julienne, Jeanne Le Peillet, Alexandre Maes et Zhou Xu.

La start-up Biomemory

Afin de développer la technologie DNA Drive, Stéphane Lemaire (Chief Science Officer) et Pierre Crozet (Chief Technical Officer) ont cofondé la start-up Biomemory avec Erfane Arwani (Chief Executive Officer), entrepreneur du numérique. Biomemory est un spin-off de Sorbonne Université et du CNRS créée en juillet 2021, au sein du laboratoire des cofondateurs, le Laboratoire de biologie computationnelle et quantitative (CNRS/Sorbonne Université) sur le campus Pierre et Marie Curie de Sorbonne Université.

ERFANE ARWANI

Erfane Arwani, entrepreneur du numérique, a précédemment fondé 3 start-ups et il est désormais CEO et co-fondateur de Biomemory.

Background

Issu d'une famille de biologistes, de chimistes et d'entrepreneurs, Erfane a commencé sa carrière au sein de la société Ariba à Sunnyvale (Silicon Valley, États-Unis). Il a ensuite principalement travaillé en finance de marché pour des grands groupes bancaires français. À partir de 2013, il rejoint l'écosystème des start-up parisiennes afin de co-fonder la Deeptech Nanocloud, puis SharePlace avec Fabrice Bonan (Talend NASDAQ: TLND) et enfin l'insurtech Osaka avec Founders Factory.

Motivation

Convaincu que la biologie synthétique est le nécessaire avenir de l'informatique, Erfane a décidé de rejoindre Stéphane et Pierre afin de créer Biomemory pour révolutionner durablement notre façon de construire nos centres de données.

Les partenaires institutionnels



Une institution citoyenne au service de la mémoire collective
Collecter, conserver, communiquer, faire comprendre et mettre en valeur leurs fonds, favoriser l'apprentissage de la citoyenneté auprès des jeunes publics sont les missions fondamentales des Archives nationales. Créées pendant la Révolution française, les Archives nationales conservent les archives publiques des différents régimes politiques du VII^e siècle jusqu'à nos jours, ainsi que les minutes des notaires parisiens et des fonds d'archives privées. Avec la loi du 7 messidor an II [27 juin 1794], l'accès gratuit aux archives de l'État est établi comme une garantie du régime démocratique et de la transparence de l'État. Ce sont près de 373 kilomètres linéaires d'archives sur tous supports, qui sont conservés, ainsi que des millions de fichiers nativement numérisés. Parmi ces documents, certains symbolisent des étapes majeures de l'histoire de France : les papyri mérovingiens, le procès des Templiers, le journal de Louis XVI, le serment du Jeu de paume, la Déclaration des droits de l'homme et du citoyen, le testament de Napoléon, les Constitutions successives de la France, ou encore le premier recensement numérique de la population en 1962.

www.archives-nationales.culture.gouv.fr



Sorbonne Université est une université pluridisciplinaire de recherche intensive de rang mondial. Structurée en trois facultés, elle couvre les champs des lettres, de la médecine et des sciences. Ancrée au cœur de Paris et présente en région, Sorbonne Université est impliquée dans la réussite de sa communauté étudiante. Elle s'engage à répondre aux grands enjeux sociétaux et à transmettre les connaissances issues de ses laboratoires et de ses équipes de recherche. Grâce à ses 52 000 étudiantes et étudiants, 6 400 personnels d'enseignement et de recherche et 3 900 personnels administratifs et techniques, Sorbonne Université se veut diverse, créatrice, innovante et ouverte sur le monde. Avec le Muséum national d'Histoire naturelle, l'Université de Technologie de Compiègne, l'INSEAD, le Pôle Supérieur Paris Boulogne-Billancourt et France Education International, elle forme l'Alliance Sorbonne Université favorisant une approche globale de l'enseignement et de la recherche, promouvant l'accès au savoir, et développant des programmes et projets de formation. Sorbonne Université est également membre de l'Alliance 4EU+, un modèle novateur d'université européenne.

www.sorbonne-universite.fr



Le Centre national de la recherche scientifique est une institution publique de recherche parmi les plus reconnues et renommées au monde. Depuis plus de 80 ans, il répond à une exigence d'excellence au niveau de ses recrutements et développe des recherches pluri et inter disciplinaires sur tout le territoire, en Europe et à l'international. Orienté vers le bien commun, il contribue au progrès scientifique, économique, social et culturel de la France. Le CNRS, c'est avant tout 32 000 femmes et hommes et 200 métiers. Ses 1000 laboratoires, pour la plupart communs avec des universités, des écoles et d'autres organismes de recherche, représentent plus de 120 000 personnes ; ils font progresser les connaissances en explorant le vivant, la matière, l'Univers et le fonctionnement des sociétés humaines. Le lien étroit qu'il tisse entre ses activités de recherche et leur transfert vers la société fait de lui aujourd'hui un acteur clé de l'innovation. Le partenariat avec les entreprises est le socle de sa politique de valorisation. Il se décline notamment via près de 200 structures communes avec des acteurs industriels et par la création d'une centaine de start-up chaque année, témoignant du potentiel économique de ses travaux de recherche. Le CNRS rend accessible les travaux et les données de la recherche ; ce partage du savoir vise différents publics : communautés scientifiques, médias, décideurs, acteurs économiques et grand public.

www.cnrs.fr



La mission de LUTECH, en tant que Société d'Accélération du Transfert Technologique*, est d'aider les chercheurs et les laboratoires de l'Alliance Sorbonne Université à transformer leurs recherches scientifiques en innovations et/ou à développer une start-up, dans l'optique d'une mise sur le marché local, national ou international. LUTECH se positionne comme un bâtisseur d'innovations et de technologies, dans l'univers large de la Deeptech, au service des grands défis du monde de demain.

* Les 13 SATT, Sociétés d'Accélération du Transfert de Technologies, ont été créées dans le cadre du Programme des Investissements d'Avenir (PIA). Elles financent le développement technologique des innovations issues de la recherche publique française grâce à un fonds d'investissement exclusif de 856 M€.

www.sattlutech.com



Alma Mater est une association pluridisciplinaire et interuniversitaire. Son journal étudiant se subdivise en trois pôles : la rédaction papier, la rédaction web et la rédaction vidéo. Sa ligne éditoriale se voulant apaisante et pluridisciplinaire, Alma Mater rassemble ses membres autour des divers campus parisiens tout en permettant à chacun de traiter une pluralité de sujets.

Les partenaires industriels



Twist Bioscience est une société de biologie synthétique et de génomique de premier plan en pleine croissance, qui a développé une plate-forme de synthèse d'ADN révolutionnaire pour industrialiser l'ingénierie de la biologie. Le cœur de la plate-forme est une technologie propriétaire qui met au point une nouvelle méthode de fabrication d'ADN synthétique en « écrivant » l'ADN sur une puce de silicium. Twist tire parti de sa technologie unique pour fabriquer une large gamme de produits à base d'ADN synthétique, notamment des gènes synthétiques, des outils pour la préparation de séquençage de nouvelle génération (outils NGS) et des bibliothèques d'anticorps pour la découverte et le développement de médicaments. Twist recherche également des opportunités à plus long terme dans le stockage de données numériques dans l'ADN et la découverte de médicaments biologiques. Twist fabrique des produits destinés à de nombreux secteurs, notamment la santé, les produits chimiques industriels, l'agriculture et la recherche universitaire.



La technologie DNAsell® d'Imagene est basée sur le confinement d'ADN dessiqué sous une atmosphère inerte, anoxique et anhydre à l'intérieur de capsules en acier inoxydable hermétiques, scellées et gravées au laser. Ainsi protégé des facteurs d'altération que sont l'eau, l'oxygène et la lumière, l'ADN naturel ou synthétique peut être conservé à température ambiante pendant une durée quasi illimitée. Cette technologie permet donc un stockage et une distribution, sûrs, fiables et autonomes de matériel sans coût énergétique. De plus, la grande capacité de stockage d'ADN de ces capsules permet de conserver d'énormes quantités de données dans un petit volume. La technologie DNAsell® d'Imagene a été étendue avec succès à la préservation d'ARN, de sang et de micro-organismes pour des applications pour les biobanques et le diagnostic moléculaire.

www.imagene.fr



Biomemory (spin-off de Sorbonne Université et du CNRS)

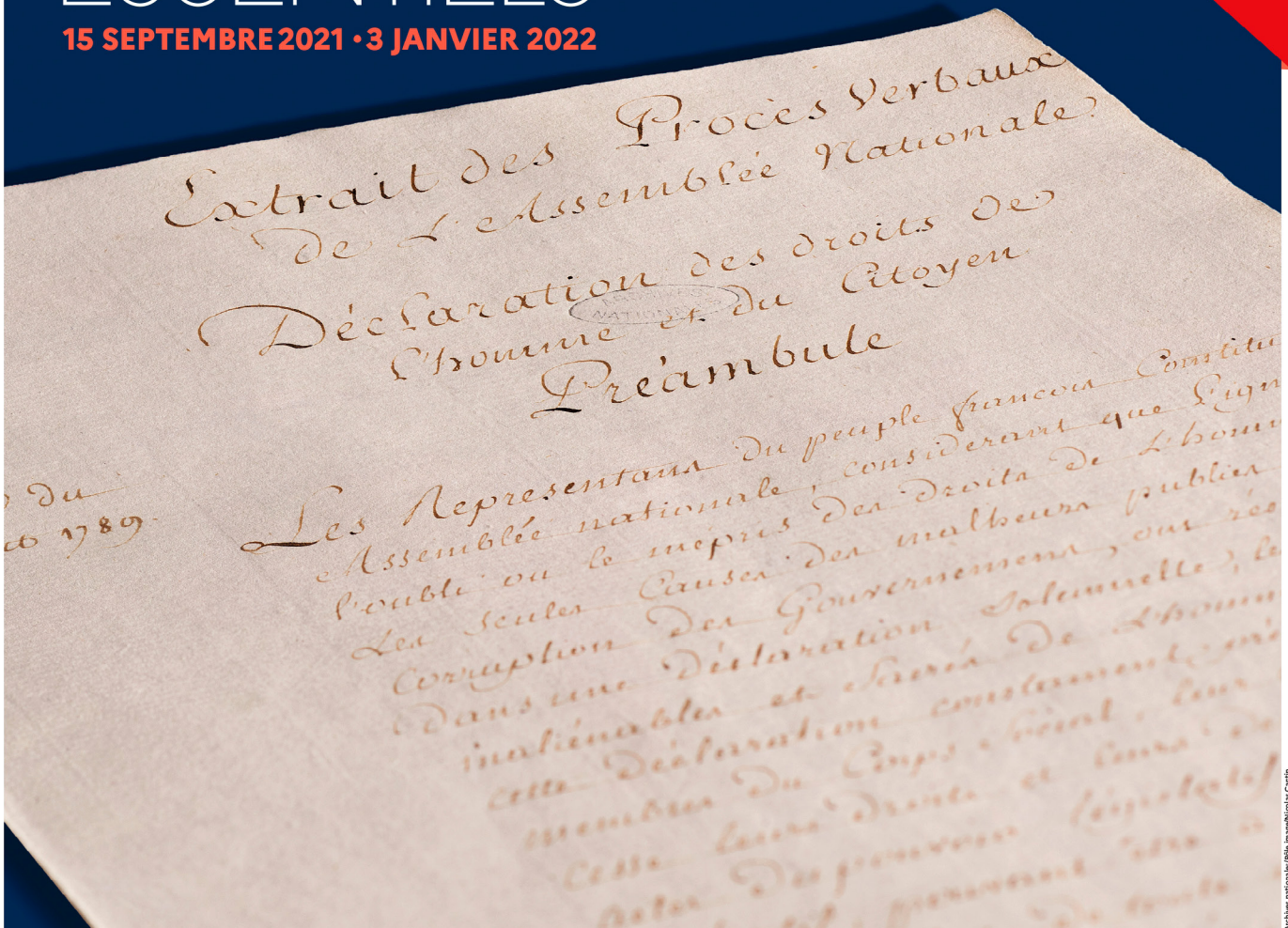
Biomemory est une société de biotechnologie spécialisée dans le stockage de données numériques sur ADN. Son objectif est de contribuer à relever le défi mondial de la croissance du volume de données. Pour cela, Biomemory travaille notamment avec de grands hébergeurs et des fabricants de matériel informatique. Biomemory conçoit des solutions de stockage de données avec la plus petite empreinte carbone possible en s'appuyant sur la domestication et l'adaptation des processus biologiques.

www.biomemory-labs.com

LES ARCHIVES NATIONALES ESSENTIELS

15 SEPTEMBRE 2021 • 3 JANVIER 2022

ENTRÉE
GRATUITE



EXPOSITION

DE LA

DÉCLARATION DES DROITS DE L'HOMME ET DU CITOYEN

DE 1789

ARCHIVES NATIONALES
60 rue des Francs-Bourgeois
75003 Paris

Du lundi au vendredi de 10h à 17h 30
Samedi et dimanche de 14h à 17h 30 - Fermeture le mardi

www.archives-nationales.culture.gouv.fr



Références

- 1) <https://www.planetoscope.com/Internet-/1523-informations-publiees-dans-le-monde-sur-le-net-en-gigaoctets-.html>
- 2) Data Age 2025. The digitization of the world. From edge to core. IDT Nov 2018, updated May 2020 <https://www.seagate.com/files/www-content/our-story/trends/files/dataage-idc-report-final.pdf>
- 3) <https://amplify.nabshow.com/articles/massive-data-growth-drives-need-for-cold-storage-strategy/>
- 4) Data Age 2025. The digitization of the world. From edge to core. IDT Nov 2018, updated May 2020 <https://www.seagate.com/files/www-content/our-story/trends/files/dataage-idc-report-final.pdf>
- 5) Un centre de données, ou centre informatique est un lieu regroupant des équipements constituant du système d'information d'une ou plusieurs entreprises. Il peut être interne et/ou externe à l'entreprise, exploité ou non avec le soutien de prestataires.
- 6) Future of DNA Data Storage, Potomac Institute for Policy Studies, 2018
- 7) <https://www.microsoft.com/en-us/research/podcast/storing-digital-data-in-synthetic-dna-with-dr-karin-strauss/>
- 8) Acide DésoxyriboNucléique
- 9) van der Valk et al. Nature, 2021 591:265
- 10) <http://www.zyvex.com/nanotech/feynman.html> Richard Feynman, Classic Talk on December 29th 1959 at the annual meeting of the American Physical Society at California Institute of Technology, transcript published in Caltech Engineering and Science 23:5, 1960
- 11) Church et al. 2012 Science 337:1628
- 12) Goldman et al. 2013 Nature 494:77 ; Blawat et al. 2016 Erlich and Zielinski 2017 Science 355:950 ; Yazdi et al. 2017 Sci Rep 7:5011 ; Organick et al. 2018 Nat. Biotechnol. 36:242
- 13) Organick et al. 2018 Nat. Biotechnol. 36:242
- 14) Les oligonucléotides sont des molécules d'ADN de plusieurs nucléotides, en général de taille inférieure à 200 nucléotides.
- 15) Une molécule d'ADN est dite simple brin lorsqu'elle n'est pas appariée à une autre molécule d'ADN complémentaire. Les chromosomes de nos cellules sont composés de molécules doubles brins, structurées en double hélice.
- 16) Organick et al. 2018 Nat. Biotechnol. 36:242
- 17) Rapport de l'Académie des technologies : Archiver les mégadonnées au-delà de 2040 : la piste de l'ADN. François Képès. Octobre 2020.
- 18) La biologie synthétique est une nouvelle approche de la biologie visant à comprendre des objets biologiques en les construisant ou à développer des systèmes biologiques nouveaux ayant des applications technologiques. Pour cela, des concepts et méthodologies issus de l'ingénierie, tels que l'amélioration continue ou la standardisation, sont utilisés. Endy D. 2005 Nature
- 19) EP 193062478, 01/10/2019; PCT EP2020/077497, 01/10/2020. CNRS, Sorbonne Université.
- 20) Marie Gouze dite Olympe de Gouges est une femme de lettres française, devenue femme politique. Considérée comme l'une pionnières du féminisme français, elle a rédigé la Déclaration des droits de la femme et de la citoyenne en 1791.
- 21) Crozet et al. 2018 ACS Synth Biol, 7 (9): 2074

CONTACTS PRESSE

Marion Valzy

06 14 02 20 51

marion.valzy@sorbonne-universite.fr

Claire de Thoisy-Méchin

06 74 03 40 19

claire.de_thoisy-mechin@sorbonne-universite.fr

Service presse des Archives nationales

communication.archives-nationales@culture.gouv.fr

Service presse du CNRS

01 44 96 51 51

presse@cnrs.fr