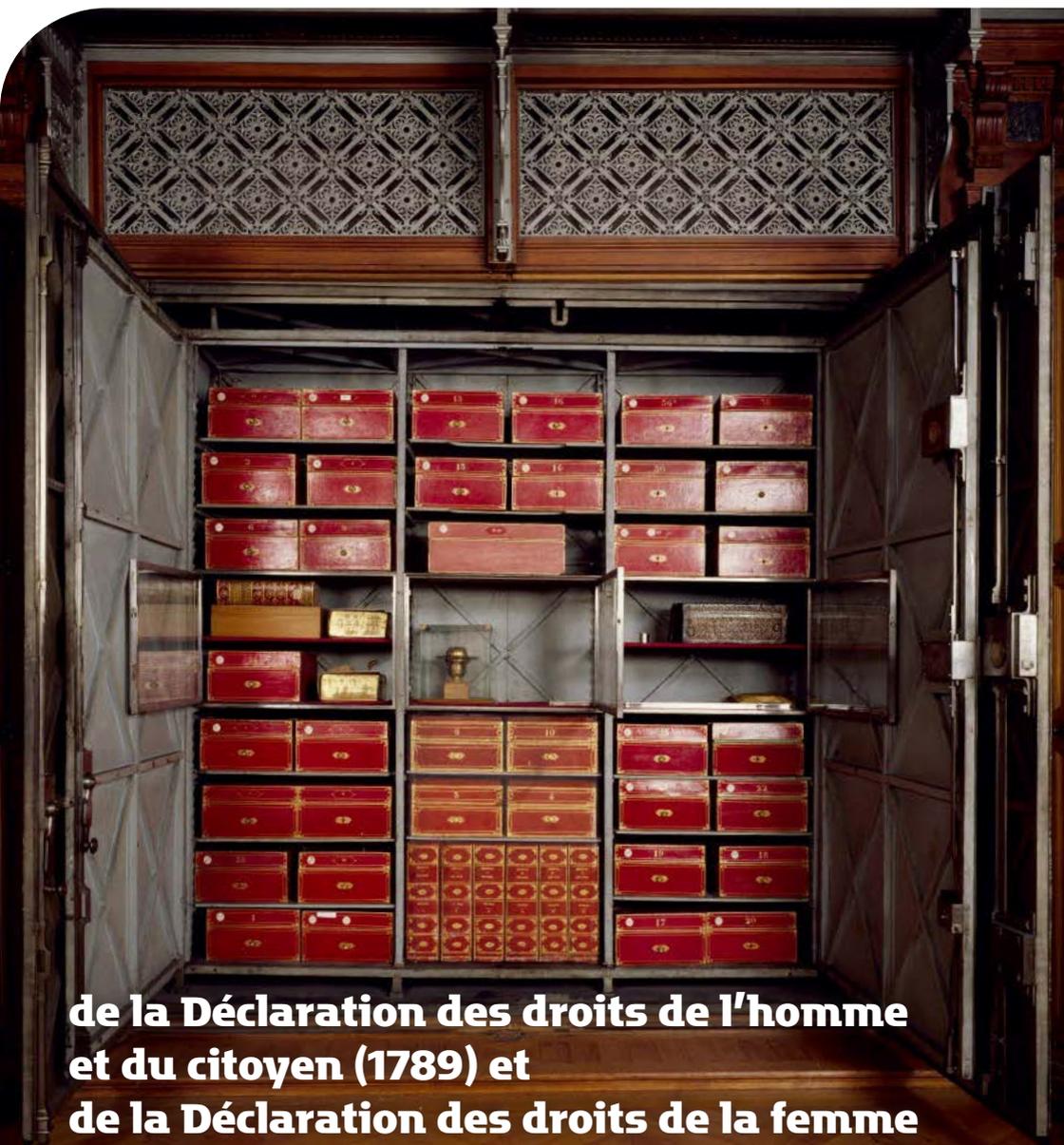


Archivage sur ADN



**de la Déclaration des droits de l'homme
et du citoyen (1789) et
de la Déclaration des droits de la femme
et de la citoyenne (1791)**

Archiver sur ADN

Les limitations de nos technologies de stockage numérique

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 considérables.

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'acide désoxyribonucléique (ADN), le support de l'information génétique. Les êtres vivants utilisent une double hélice d'ADN (deux molécules 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 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'années.
- ◇ non énergivore : l'ADN est stable à température ambiante sans aucun apport d'énergie s'il est conservé dans des conditions adéquates (sans eau, sans air, sans lumière).
- ◇ extrêmement compacte : avec une densité maximale de 450 millions de To par gramme d'ADN, l'intégralité des données mondiales pourrait tenir dans 100 g d'ADN, soit le volume d'une tablette de chocolat.

Méthodologies de stockage d'information numérique sur ADN

Les premiers travaux ont tous utilisé la même méthodologie consistant en :

- 1 - la transformation d'une donnée numérique binaire (0 ou 1) en donnée quaternaire (A, T, C, G).
- 2 - la synthèse chimique de la séquence ADN ainsi générée sur des petits fragments d'ADN appelés oligonucléotides.

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 (Sorbonne Université–CNRS) cherche à tirer parti du potentiel offert par le vivant et utilise la biologie pour répondre aux enjeux du stockage d'information sur ADN.

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

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 fondée 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 similaires à ceux utilisés en biologie et en médecine pour séquencer les génomes des êtres vivants. C'est une technologie sécurisée : l'ADN est crypté afin de ne porter aucune information génétique potentiellement dangereuse pour l'être humain ou pour l'environnement.

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. Un très grand nombre de molécules sont rassemblées pour former un DNA Drive dont la capacité est illimitée.

Le DNA Drive est compatible avec tout type d'information numérique, puisque son algorithme 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 ADN sans avoir besoin de relire l'intégralité des données.

La preuve de concept avec l'encodage de deux textes majeurs de l'histoire de France

Ainsi est né le projet « la Révolution de l'ADN ». Dans un premier temps, une petite équipe de biologistes et d'informaticiens a conçu et breveté la technologie DNA Drive avec des financements d'amorçage de Sorbonne Université et du CNRS puis de la SATT Lutec.

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, des philosophes, des informaticiens et des biologistes.

Il a bénéficié également du partenariat avec Twist Bioscience, entreprise américaine spécialiste de la synthèse d'ADN, et avec 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.

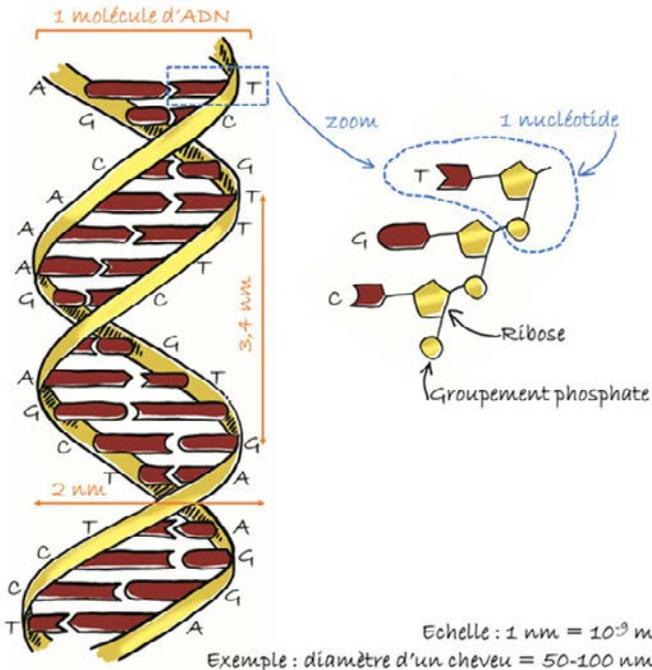
Twist Bioscience a ensuite synthétisé des fragments d'ADN qui ont été assemblés et organisés par l'équipe de recherche sur des grandes molécules d'ADN double brin. Ces molécules ont été amplifiées biologiquement pour en obtenir plus de 100 000 milliards de copies.

Les molécules d'ADN ont ensuite été extraites, purifiées puis encapsulées sous forme lyophilisée dans des capsules DNA Shell© 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 préserve l'ADN des dommages du temps en le protégeant de la lumière, de l'air et de l'oxygène et permet 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 encodée, un peu de l'ADN est déposé sur un séquenceur. Une fois la séquence obtenue, 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.

Pour le 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, l'équipe de recherche a 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 au moins plusieurs dizaines de milliers d'années.

Schéma de la structure de l'ADN :



© Illustration : Jeanne Le Peillet

Les Déclarations

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 fondements 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 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.

Les acteurs de la recherche

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 produit un kit de clonage modulaire (MoClo) composé de 119 biobriques pour la microalgue *Chlamydomonas reinhardtii*. Ce kit permet la construction rapide de souches reprogrammées génétiquement à la fois pour la recherche fondamentale et les approches biotechnologiques fondées 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 mettre en œuvre le développement d'une nouvelle technologie de stockage numérique sur ADN.

L'équipe projet

Les recherches sur le stockage d'informations numériques ont été menées par une équipe projet coordonnée par Stéphane Lemaire, directeur de recherche au CNRS, et Pierre Crozet, maître de conférences, Sorbonne Université.

Ces travaux de recherche se passent à l'Institut de biologie Paris Seine [IBPS] au Laboratoire de biologie computationnelle et quantitative (LCQB, CNRS – Sorbonne Université), dirigé par Alessandra Carbone, mais ont été lancés à l'Institut de biologie physico-chimique [IBPC] au Laboratoire de biologie moléculaire et cellulaire des eucaryotes (Sorbonne Université – CNRS), dirigé par Teresa Teixeira.

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 : Clémence Blachon, Nicolas Cornille, Mariette Gibier, Achille Julienne, Jeanne Le Peillet, Alexandre Maes et Zhou Xu.

Les partenaires



**SORBONNE
UNIVERSITÉ**

Sorbonne Université

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 des projets de formation. Sorbonne Université est également membre de l'Alliance 4EU+, un modèle novateur d'université européenne.

www.sorbonne-universite.fr



CNRS

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 interdisciplinaires 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* 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



SATT LUTECH

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 millions d'euros.

www.sattlutech.com



Twist Bioscience Corporation

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. Elle fabrique des produits destinés à de nombreux secteurs, notamment la santé, les produits chimiques industriels, l'agriculture et la recherche universitaire.

www.twistbioscience.com

immagine **Imagene**

La technologie DNAshell® d'Imagene est fondée 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 DNAshell® 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



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 créé en 2021 et hébergé au sein du laboratoire des cofondateurs, sur le campus de Sorbonne Université.

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. Elle 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

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.

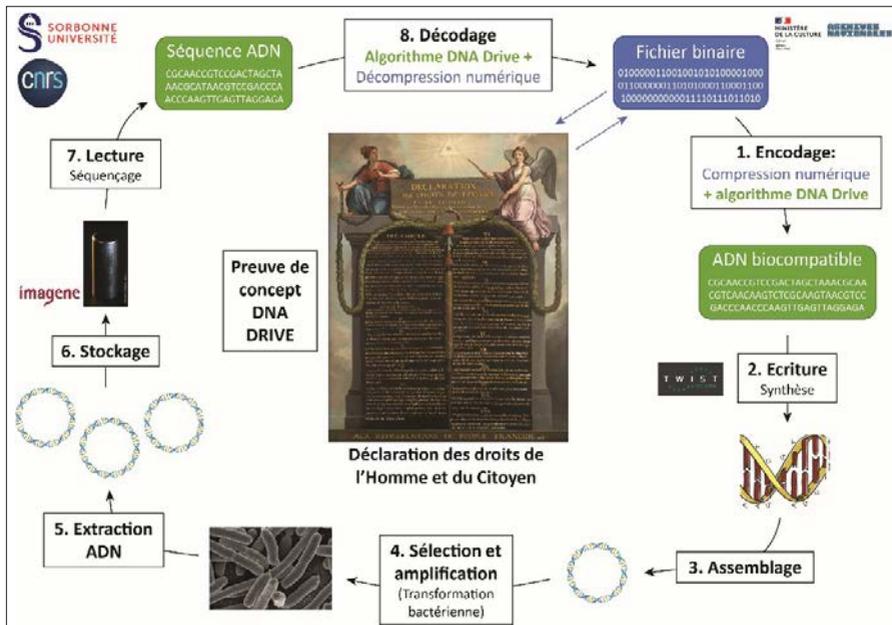
Créées pendant la Révolution française, les Archives nationales conservent les archives publiques des différents régimes politiques du ^{vii}e siècle jusqu'à nos jours, ainsi que les minutes des notaires parisiens et des fonds d'archives privées.

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 numériques.

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.





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

www.archives-nationales.culture.gouv.fr