

CHAT POINT NET

Comment épargner l'environnement en conservant ses données dans l'ADN

François Képès

MOTS-CLÉS

Biologie de synthèse ; ADN ; Polymères ; Données numériques ; Archivage de données ; Environnement.

INTRODUCTION

En tout lieu, nous sommes habitués à stocker et retrouver en un clic nos données, vidéos, photos, livres, sur n'importe lequel de nos terminaux informatiques, téléphone, ordinateur. Ce service ne demande aucune compréhension des mécanismes sous-jacents, que ce soit le fonctionnement d'un terminal informatique, d'un centre de données, d'un système de télécommunication. Or l'impact environnemental du stockage et archivage de nos données numériques, déjà problématique, deviendra insupportable d'ici 2040. Notre propos ici concernera des pistes pour réduire drastiquement cet impact environnemental.

ARCHIVAGE DES DONNÉES NUMÉRIQUES

Vous avez sauvegardé cette émouvante photo de votre famille sur le "cloud" (en français, le "nuage"). Comme vous y êtes souvent — dans le nuage — vous avez peine à vous remémorer où elle est stockée, alors même que vous souhaiteriez la montrer à des amis auxquels vous rendez visite. Votre dernier souvenir est de l'avoir engrangée sur votre ordinateur qui se trouve à 100 km de là. Il vous semble avoir un jour coché la case "cloud", quoi que cela puisse signifier. Sans trop y croire, vous lancez une application nommée "Photo" sur votre téléphone, et en deux clics voici la photo recherchée. Vous n'en revenez pas, et vos amis vous félicitent pour vos capacités techniques hors pair.

Où se trouve réellement votre photo ? Sur votre ordinateur, sur votre téléphone, dans le nuage cumulo-nimbus qui vous surplombe en cet instant ? La réponse est plus exotique : le fichier de votre photo de famille est conservé quelque part sur un disque magnétique, dans un immense hangar situé au nord de la Norvège, où vous n'avez pourtant jamais mis les pieds ; ce hangar abrite un centre de données. Par sécurité, votre fichier a été dupliqué dans un autre énorme centre situé, disons, en Estonie. En fait, des camions ont un jour transporté à travers l'Europe les disques magnétiques portant toutes les données du premier centre vers le second, afin de créer une copie de sécurité : votre photo ne sera pas perdue en cas d'incident grave dans le hangar norvégien. Jamais les réseaux informatiques n'auraient pu transporter de pareilles quantités de données dans un délai acceptable. Il y faudrait deux millénaires avec une liaison internet de qualité moyenne. C'est pourquoi des camions sont utilisés pour de tels transports.

Quelles données sont abritées dans de pareils centres de données ? Celles de nos connexions familiales, amicales et professionnelles, nos livres, vidéos et photos, nos données médicales, celles de la recherche scientifique, de l'industrie etc.

Bien plus est à venir, avec l'augmentation du nombre des capteurs connectés, la télésurveillance, la réalité virtuelle, la consultation et chirurgie à distance, les voitures autonomes ...

Pourquoi parle-t-on de données "numériques" ? La musique enregistrée s'écoutait sur des disques microsillon; ces enregistrements non numériques s'endommageaient dès les premières écoutes. Puis sont arrivés les disques compacts ; comme le streaming, il s'agit d'une technologie

numérique. Tant que le support physique reste en bon état, la musique lue ne connaît aucune dégradation. Pour la photo, la comparaison se ferait entre le procédé argentique qui est chimique donc instable, et le procédé numérique. Le numérique a donc l'avantage de la stabilité, en principe parfaite, et de permettre le traitement informatique, qui est quasi-universel.

Qu'est ce que représenterait l'ensemble des données accumulées par l'humanité, la "sphère globale des données" ? Notre unité sera le caractère, comme ceux que j'utilise ici pour écrire : une lettre, un chiffre, un symbole. Techniquement, assimilons cela à un octet, soit une suite de huit '0' ou '1' dans un fichier numérique. Eh bien, la sphère globale des données comprend actuellement autant de caractères qu'il y a d'étoiles dans l'univers observable : quelques dizaines de milliers de milliards de milliards de caractères. Pour transférer tout cela par internet, il faudrait environ 50 millions d'années. Oui, c'est long, d'autant plus que cette sphère globale double environ tous les deux ans. Comme disait Woody Allen, l'éternité c'est long, surtout vers la fin.

LIMITES DU MODÈLE ACTUEL

Enfin, du moment que votre photo est sauvegardée et disponible, nous vivons dans un monde merveilleux et vous pourrez continuer à accumuler des photos et vidéos dans le futur. Mais tout de même, ce système qui nous rend le monde merveilleux est-il merveilleux lui-même ? Revenons donc à ce gros centre de données norvégien tout à fait typique. Quelques ordres de grandeur : un exaoctet (un milliard de milliards de caractères) ; un million de serveurs (des ordinateurs sans clavier ni écran) ; des disques durs ou mémoires statiques et des bandes magnétiques qu'il faut renouveler tous les 5 à 7 ans ; un terrain de plusieurs centaines de milliers de m² ; une consommation électrique supérieure à celle d'une ville de cent mille habitants, dont 40% pour le refroidissement des serveurs (d'où l'intérêt de positionner ces centres dans des régions froides du globe) ; des milliers de tonnes de métaux chers, de plastiques issus du pétrole, et des terres rares, des ressources coûteuses à produire; un investissement de quelques milliards d'euros ; une durée de vie de 20 ans. Sans compter les réseaux de connexions qui le relie au reste du monde, également consommateurs de ressources. N'en jetez plus, la coupe est pleine.

Ce n'est là qu'un seul centre. Si la sphère globale des données était entièrement contenue dans des centres de cette taille, il en faudrait 50 000 à ce jour. En fait, ces gros hangars coexistent avec de plus petits bâtiments, pour un total mondial de quelques millions de centres de données. Ensemble, tous ces centres couvrent actuellement un millionième des terres émergées de la planète. Ils consomment 2 à 4% de l'électricité dans les pays avancés, soit, si tous ces centres formaient un seul pays par hypothèse, alors ce pays serait le cinquième plus gros consommateur d'électricité au monde, entre Inde et Japon.

Enfin, du moment que votre photo est sauvegardée et disponible ! Oui aujourd'hui, mais qu'en sera-t-il demain ? Répétons le, nous vivons approximativement un doublement de la sphère globale des données tous les deux ans; donc, un facteur 1 000 en 20 ans. La mauvaise nouvelle est que la construction de hangars dédiés, ainsi que la fabrication de coûteuses ressources comme par exemple le silicium de qualité électronique, ne suit pas du tout ce doublement. Dans les 20 ans à venir, sera produit seulement 1% du silicium de qualité électronique dont a besoin l'industrie. Continuons à jouer avec la loi exponentielle de ce doublement tous les deux ans. En 2040, hors la faible marge d'optimisation qu'il reste à gagner aux méthodes conventionnelles, les centres de données couvriraient un millième des terres émergées. Je vous laisse calculer la consommation électrique des centres et de leurs réseaux en 2040.

Mais alors, votre photo ? Faudra-t-il choisir entre la photo de famille et celle du chat ? Vous ne le croyez pas, et puis ce serait trop dur d'en éliminer une. Eh bien, il a été calculé que, déjà en 2020, nos capacités de stockage en centre de données correspondent seulement à 60% des besoins. Ce

pourcentage va baisser rapidement. Même en ignorant l'impact environnemental, il reste donc que le système actuel ne remplit même plus son office.

Pour toutes ces raisons, le modèle actuel des centres de données n'est pas soutenable, et il deviendra insupportable d'ici 2040. Outre l'optimisation du modèle actuel, qui sera limitée, il faut donc trouver rapidement une autre issue au problème des mégadonnées ou "big data".

Vous renoncez à votre photo de chat ? Merci pour votre sacrifice, et merci de la part de notre environnement. Mais ce ne sera pas suffisant. "On n'a pas besoin de toute cette information." Là, je vous arrête. Vous parlez probablement des informations des autres, mais les autres, eux, ne sont pas forcément passionnés par vos photos, par exemple votre photo de chat, d'autant qu'elle est floue — je ne voulais pas être désagréable et vous le faire remarquer, mais vous m'y obligez. Plus globalement, il convient de rappeler que l'information a été le moteur de la croissance socio-économique de la civilisation depuis ses débuts. C'est un fait que la production d'information corrèle avec le bien-être social et la croissance économique. Or, la production et l'utilisation de l'information ont connu une croissance exponentielle dans les dernières décennies.

PISTES POUR L'AVENIR

Qu'alors y faire (les serveurs chauffent beaucoup) ? Plusieurs choses :

1. économiser notre espace de stockage de données ;
2. distinguer les données méritant d'être conservées ;
3. compresser et *interpréter* les données ;
4. éduquer les petits et les grands ;
5. changer radicalement le modèle des centres de données.

Développons calmement tout ceci à l'aide de petits exemples.

1. Économiser notre espace de stockage de données

Par exemple, telle photo de chat trop chou mais un peu flou est sur votre mémoire d'ordinateur ou de téléphone. Mais vous n'êtes pas seul à l'apprécier. Des millions de fans de chats l'ont gardée aussi, et leurs données informatiques sont sauvegardées en de multiples exemplaires dans divers centres de données, souvent à leur insu. Si ces millions de fans avaient conservé et échangé le lien internet vers cette photo de chat point net (www.xyz.chat.net — quelques caractères), plutôt que le fichier de la même photo (quelques 100 000 à 2 000 000 caractères), le gain d'espace aurait été d'un facteur 5 000 à 100 000.

Et Internet ne regorge pas que d'une seule photo de chat. Et il n'y pas que des photos de chats. Et il n'y a pas que des photos. Et il n'y a pas qu'Internet.

C'est un cas clair pour lequel la discipline et la compréhension de chacun fera une différence gigantesque à l'arrivée.

2. Distinguer les données méritant d'être conservées

Bien sûr, toutes les données ne sont pas égales. De sublimes films anciens ont été numérisés et archivés, mais le coût de cet archivage les met actuellement en danger ; or ils font partie du patrimoine le plus précieux de l'humanité. Mais le statut de votre brosse à dents connectée l'an passé n'en fait pas partie, même si vos dents font partie de votre patrimoine personnel le plus précieux.

Schématiquement, il faudrait distinguer trois types de données.

- d'usage immédiat et d'intérêt nul dans le futur (par exemple venant de capteurs de vitesse et de pente sur votre voiture, qui ont déterminé son passage automatique en 3^e vitesse) ;
- méritant un archivage longue durée (par exemple résultant des expériences passées de physique nucléaire, utiles à transmettre entre générations de physiciens) ;

- intermédiaire (par exemple émanant des données de télésurveillance, qui ne sont conservées au-delà d'un jour que si une enquête est diligentée).

Dans le cadre professionnel, ces distinctions sont souvent opérées, et permettent à l'actuel déficit en espace de stockage de se creuser, pour l'instant sans douleur. Pour poursuivre sur la lancée, il serait nécessaire d'étendre les domaines d'activités qui opèrent une distinction soigneuse entre ces différentes durées de stockage.

3. Compresser et interpréter les données

Le terme "compression" recouvre en général des techniques informatiques permettant de réduire la taille d'un fichier en perdant peu ou pas d'information utile. Par exemple si une image contient un grand à-plat de couleur constante, cette zone sera décrite plus concisément en indiquant une seule fois sa couleur, et en précisant géométriquement ses frontières, plutôt qu'en répétant la valeur de sa couleur pour chacun de ses nombreux pixels. Beaucoup d'entre nous connaissent les standards de compression d'images comme par exemple PNG qui préserve toute l'information, ou JPEG qui en perd une part.

Ces techniques sont essentielles pour la réduction des volumes d'information à sauvegarder. Pourtant, j'estime que des approches plus radicales doivent y être combinées au vu du fardeau grandissant des données. Il s'agit ni plus ni moins de ce qu'on appelle en langage courant *l'interprétation des informations, des données, des résultats*.

Par exemple, supposons que nous souhaitions connaître les réactions d'une cellule vivante en culture dans le laboratoire, à un traitement médicamenteux. Pour cela, supposons que nous ayons mesuré l'évolution temporelle et spatiale de la production de ses dizaines de milliers de protéines, grâce à une série d'images à très haute résolution. Finalement, supposons que nous décrivions notre expérience et ses résultats dans un article scientifique. Nous concluons par exemple que ce traitement induit la baisse de production des protéines P1 et P2 qui, dans ce contexte, sont très pertinentes pour telle raison. C'est là l'interprétation du plus haut niveau possible. Il est donc possible de remplacer des centaines de lourds fichiers d'images (des millions de millions de caractères) par une phrase contenant quelques caractères : "Si telle cellule subit tel traitement, alors la production de sa protéine P baisse de 40% en 65 minutes." En pareil cas, les lourds fichiers pourraient être automatiquement supprimés un an après publication de cet article par exemple. Au-delà, seule serait conservée l'interprétation signifiante, c'est-à-dire la phrase contextualisée par l'article scientifique.

Il sera indispensable d'introduire cette "compression par l'interprétation" (suivie bien sûr de compression informatique conventionnelle). Je suggère de réfléchir dès maintenant à ses modalités selon les cas qui se présentent dans les très divers champs de l'activité humaine.

4. Éduquer les petits et les grands

Le point 1 portant sur l'économie de l'espace de stockage pourrait être enseigné de manière ludique et visuelle dès l'âge des premiers terminaux informatiques, souvent le téléphone mobile confié à un enfant. Le point 2 concernant la durée de conservation des informations pourrait faire l'objet de sensibilisation progressive durant les cycles secondaires d'enseignement, en s'appuyant concrètement sur les activités habituelles de cet âge. Le point 3 portant sur la compression et son avatar ultime, l'interprétation, pourrait tôt faire l'objet d'une prise de conscience, pour ensuite évoluer et se spécialiser au gré des activités professionnelles de chacun.

Avec toutes ces améliorations en vue, vous vous demandez si finalement vous ne pourriez pas garder votre photo de chat ? Non, gardez plutôt le pointeur web vers cette photo. Allez faire votre stage de civisme internautique. Éducation, vous dis-je.

5. Changer radicalement le modèle des centres de données.

Il serait utile à toutes les personnes connectées d'avoir quelques notions du fonctionnement des centres de données pour bien motiver et diriger leurs efforts d'économie. Or, 75% des humains

seront connectés en 2025, et chacun interagira avec des données toutes les 18 secondes en moyenne.

Pour le stockage et l'archivage des données, les centres reposent sur des solutions admirables au plan technique, mais n'offrant plus de marges suffisantes d'optimisation pour faire face au déluge des données. En outre, leur impact environnemental est déjà problématique. Un chantier de 20 ans se dresse devant nous, pour révolutionner ces technologies.

L'Académie des technologies s'est penchée sur cette question durant deux années (2018-2020). Son groupe de travail s'est focalisé sur la piste de l'ADN. L'ADN serait ici utilisé comme un composé chimique en dehors du vivant, et d'ailleurs pourrait être remplacé par une autre fibre du type plastique. Je vais résumer pour vous les conclusions de ce groupe de travail que j'ai animé.

Mais revenons à nos chats. Imaginez une molécule toute en longueur, dépourvue de branchements. Imaginez de plus que cette molécule soit constituée de l'enchaînement de deux sortes de briques différant par leurs couleurs, par exemple saphir (S) ou nankin (N). Comme mon fichier numérique est constitué d'un enchaînement de "0" et de "1", il suffit de le coder en décidant que "0" vaut "S", et "1" vaut "N" (ou l'inverse), selon le petit tableau de correspondances. Ainsi, l'octet "00 10 00 11" serait recodé en "SS NS SS NN". Il ne reste plus qu'à écrire la molécule (polymère) portant l'enchaînement correct des briques (monomères) "S" et "N". Cette écriture consiste en une synthèse chimique, qui doit être contrôlée afin de positionner les briques dans le bon ordre.

Binaire Binaire

S	0
N	1

Vous dites ? Oui, cela marche aussi avec du Lego, du moment que vous avez des briques d'au moins deux couleurs. Mais avec du Lego, vous ne risquez pas de réduire le volume des données, car ses briques sont énormes à l'échelle d'un disque dur. Or l'intérêt du stockage à l'échelle de la molécule (du polymère), c'est justement que ses briques sont bien plus petites que l'espace dédié à un "0" ou un "1" sur un disque dur.

Où en sommes-nous ? Nous avons codé puis écrit la molécule. Maintenant il faut encore la stocker sans qu'elle se dégrade. Enfin, le moment venu, elle sera sortie de stockage et soumise au processus inverse : lecture et décodage. La lecture consiste à lire la séquence des briques dans l'ordre : S, puis S, puis N etc. Le décodage consiste à transformer la séquence " SS NS SS NN " en "00 10 00 11" : voilà le fichier numérique d'origine reconstitué.

Il devrait être possible d'encore densifier le support des données numériques, en utilisant des alphabets plus fournis que {0,1} ou {S,N}. Au lieu de ces alphabets binaires, prenons-en un qui soit quaternaire, comme {A, C, T, G}, avec les correspondances du petit tableau. Ainsi, l'octet "00 10 00 11" serait recodé en "AGAT", donc seulement quatre briques au lieu de huit avec l'alphabet binaire {S,N}, d'où un gain en espace.

Quaternaire Binaire

A	00
C	01
G	10
T	11

Nous pourrions imaginer des codages / décodages plus sophistiqués, mais celui-ci tient la route.

L'ADN, le constituant de nos chromosomes, est un polymère linéaire (sans branchement) qui comporte quatre briques, ou monomères {A, C, T, G}. Il pourrait donc faire l'affaire en tant que composé chimique utilisé en dehors du vivant. Un avantage décisif de cette molécule est que les technologies développées par l'homme pour lire et écrire l'ADN depuis les années 1980 sont bien avancées, car elles sont portées par les très rapides progrès du domaine biomédical. Une petite complication pratique avec l'ADN est qu'en fait il se présente généralement comme une hélice faite de 2 brins courant en sens opposé. Mais les 2 brins sont complémentaires, en face

d'un A il y a toujours un T, et en face d'un C un G. Aussi, en ne considérant qu'un seul des brins et en le déroulant, nous ne perdons aucune information et nous n'avons rien à changer aux considérations ci-dessus.

Petite parenthèse sur les ordres de grandeur, notre génome, qui prend la forme de 46 chromosomes dans le noyau de chacune de nos 100 mille milliards de cellules, fait au total 3,5 milliards de briques de long. Il aurait donc le potentiel de coder 0,8 milliards de caractères, soit un film en basse résolution. Mais rappelons qu'ici nous considérons l'ADN comme un composé chimique hors du vivant.

Avant de quitter le domaine des polymères "numériques" (ceux qui permettent de stocker des informations numériques), il faut signaler que d'excellents chimistes estiment que l'ADN n'est pas le support idéal pour les données informatiques. Ils proposent, soit de le "rectifier" par des retouches plus ou moins profondes, soit de s'en éloigner pour choisir des polymères très différents, mais répondant aux critères définis plus haut.

Revenons à la piste de l'ADN puisque sa faisabilité a été démontrée : il peut servir de polymère numérique. Quel principe guide l'usage de l'ADN pour stocker des données numériques ? Il suffit de 5 étapes : coder, écrire, stocker, lire, décoder. Coder et décoder sont des opérations informatiques simples. Nous venons d'examiner comment coder avec au moins 2 briques (4 pour l'ADN). Décoder est l'opération inverse. Écrire et lire sont des opérations chimiques. Écrire, c'est synthétiser chimiquement l'ADN ou un autre polymère. Lire, l'opération inverse, c'est le séquencer, *i.e.* déterminer la séquence de ses briques ou monomères. Stocker, c'est protéger la molécule dans un conteneur. Le procédé actuellement le plus performant permet de conserver l'ADN durant des dizaines de milliers d'années à température ambiante. L'ADN est protégé en atmosphère inerte dans un conteneur en verre et acier de la taille d'une petite pile bouton. Ce mode de stockage ne consomme ni énergie ni autre ressource. Il existe depuis 2019 un prototype automatisé qui enchaîne l'ensemble des opérations sur un milliard de caractères, soit la taille d'un film, en 24 heures.

Quels gains peut-on attendre de ce stockage moléculaire, par rapport à la situation actuelle de stockage informatique ?

Le volume requis pour stocker des données serait diminué d'un facteur 10 millions. La longévité du support serait augmentée d'un facteur 10 mille. La consommation électrique pourrait être réduite d'un facteur mille. Il deviendrait beaucoup plus facile de réaliser des copies des données numériques, et beaucoup plus sûr de les détruire volontairement. Enfin, l'obsolescence du support ADN ne se produira pas tant que l'homme disposera des technologies nécessaires à l'écriture et à la lecture de l'ADN, qui font partie intégrante de la médecine moderne.

Alors pourquoi votre photo de chat n'est pas déjà stockée sous forme d'ADN ? Non, la bonne réponse n'est pas que votre chat stocke de l'ADN, son ADN. Je parlais de sa photo qu'il s'agit de stocker sur de l'ADN en dehors du chat, en dehors du vivant. Or, la preuve de principe du stockage de données sur l'ADN existe, mais les technologies associées à l'ADN doivent encore progresser fortement avant de devenir compétitives par rapport aux solutions conventionnelles. Aux plans du coût comme de la vitesse, la lecture de l'ADN doit encore progresser d'un facteur mille, et l'écriture d'un facteur 100 millions.

Comment cela, revenez m'en parler dans un siècle ? Depuis plusieurs décennies, ces technologies progressent au rythme d'un facteur 2 tous les 6 mois, donc d'un facteur mille tous les 5 ans, ou encore un million tous les 10 ans. Autrement dit, la lecture sera compétitive dans 5 ans, et l'écriture dans moins de 15 ans. Cela arrivera donc avant 2040, sans compter les autres pistes prometteuses actuellement moins avancées, comme des chimies différentes de l'ADN.

Pour les geeks d'entre vous, je glisse une petite parenthèse pour signaler que dans la pyramide des types de mémoire (plus on va vers le bas, plus la capacité et le temps d'accès augmentent), la bande magnétique est tout en bas. Au vu de ce qui vient d'être dit, il semble donc probable que

l'ADN ou un autre polymère numérique arrive en complément de la bande magnétique, tout en bas de la pyramide ; puis qu'il finisse par entrer en compétition.

CONCLUSION

Étant donné que le handicap du stockage moléculaire se situe dans la lenteur et le coût des procédés de lecture et surtout d'écriture, on peut prédire que cette solution sera mise sur le marché progressivement, depuis des marchés de niche sous 5 à 10 ans, vers des marchés plus globaux sous 10 à 20 ans. Des exemples de marchés de niche pourraient être le patrimoine culturel (films, livres, monuments scannés en 3 dimensions), le patrimoine scientifique (résultats de physique nucléaire comme de biologie moléculaire), les données bancaires qui doivent règlementairement être conservées durant plusieurs années.

Le monde (du numérique) est sauvé, ouf !

RÉFÉRENCES

- Académie des technologies (2020). Rapport "Archiver les mégadonnées au-delà de 2040 : la piste de l'ADN" — <https://www.academie-technologies.fr/blog/categories/publications-de-l-academie/posts/archiver-les-megadonnees-au-dela-de-2040-la-piste-de-l-adn>
- Jean-François Lutz, Université de Strasbourg (2021). Travaux sur les polymères à séquences contrôlées — https://fr.wikipedia.org/wiki/Jean-François_Lutz
- Usbek & Rica (2021). "Et si demain nous stockions nos informations dans l'ADN ?" — <https://usbeketrica.com/fr/article/et-si-demain-nous-stockions-nos-informations-dans-l-adn-10-chiffres-pour-comprendre>
- Science&Vie (2021). "ADN : il peut stocker toutes nos données !" — <https://www.science-et-vie.com/technos-et-futur/adn-il-peut-stocker-toutes-nos-donnees-61968>
- Radio Télévision Suisse (2021). "Stocker nos données sur de l'ADN" — <https://pages.rts.ch/la-1ere/programmes/cqfd/11890013-stocker-nos-donnees-sur-de-ladn-26-01-2021.html?mediaShare=1>
- France Info (2020). "Nouveau monde. Une solution pour faire face à l'explosion des données numériques : le stockage sur ADN" — https://www.francetvinfo.fr/replay-radio/nouveau-monde/nouveau-monde-une-solution-pour-faire-face-a-lexplosion-des-donnees-numeriques-le-stockage-sur-adn_4162983.html
- Les Echos (2020). "Stockage de données : pourquoi l'ADN va devenir incontournable" — <https://www.lesechos.fr/idees-debats/sciences-prospective/stockage-de-donnees-pourquoi-ladn-va-devenir-incontournable-1271212>

CV DE L'AUTEUR

François Képès est un biologiste cellulaire dont les approches mêlent les biologies moléculaire, des systèmes et de synthèse, et les microscopies.

François Képès est membre de l'Académie des technologies, et membre correspondant de l'Académie d'Agriculture de France. Il est co-fondateur du Programme d'Épigénomique (Genopole), et fondateur de l'institut de Biologie des Systèmes et de Synthèse (iSSB — Genopole, CNRS, Univ. Évry) et du master éponyme (mSSB — Université Paris-Saclay). Normalien, il fut Professeur associé à l'École Polytechnique et au Politehnica de Bucarest, Professeur invité au Collège Impérial de Londres et Directeur de recherche au CNRS.

Auteur de plus de 130 articles ou chapitres scientifiques, et auteur ou éditeur de plus de 25 livres, François Képès a de 2003 à 2018 organisé ou présidé 3-7 manifestations scientifiques internationales par an. Il est éditeur de quatre journaux scientifiques internationaux dont "Synthetic Biology". Il est expert auprès d'agences de moyen européennes et nord-américaines.