

Fiche projet – DNAStore - Stockage d'images sur ADN et polymères artificiels

Après échanges avec l'institut INS2I du CNRS, cette fiche vise à soutenir le portage par le CNRS d'une Chaire de Professeur Junior en 2022 sur le sujet du stockage sur ADN et polymères, avec atterrissage sur un poste DR CNRS. L'Idex UCA-JEDI apportera un budget d'environnement à cette CPJ portée par le CNRS et stratégique pour Université Côte d'Azur.

Établissements/organismes/entités partenaires envisagés : CNRS, Université Côte d'Azur, Inria

Nom du projet : Stockage d'images sur ADN et polymères artificiels (DNAStore)

Mots-clés : Stockage d'information, données massives, compression d'information, codage, traitement du signal et des images

Thématique scientifique : Compression/codage des signaux et des images

Section (s) CNU/CoNRS/CSS correspondante(s) : 61, 27

Positionnement de la thématique dans la stratégie de l'établissement : *décrire en quoi la thématique est en lien avec la stratégie de l'établissement (15 lignes maximum)*

Cette demande rentre dans le cadre du projet MolecuArXiv (stockage massif de données sur l'ADN et les polymères artificiels) dont la direction scientifique est assurée par un membre du laboratoire I3S. Ce projet est financé à hauteur de 20 M€ sur 7 ans dans le cadre des Programmes et Equipements Prioritaires de Recherche exploratoires (PEPR) du Programme et Investissement d'Avenir PIA4.

MolecuArXiv favorisera la synergie et stimulera le travail collaboratif au sein d'Université Côte d'Azur, en proposant de véritables outils transdisciplinaires, qui seront étendus à une communauté plus large, nationale et internationale. Plusieurs collaborations sont déjà en place entre le laboratoire I3S et des acteurs de Sophia Antipolis tels que l'IPMC (Plateforme UCA-Genomix) et EURECOM.

Il favorisera aussi l'animation des communautés françaises et européennes jusqu'à l'émergence d'une proposition européenne FET-flagship.

Positionnement de la thématique dans la stratégie du laboratoire d'accueil : *décrire en quoi la thématique est en lien avec la stratégie du laboratoire d'accueil (15 lignes maximum)*

L'ambition est de créer un centre de référence multidisciplinaire dans le traitement des données numériques de l'ADN, axé sur le développement de l'avenir du stockage des données numériques à long terme. Ce centre de référence a le potentiel d'attirer davantage de partenaires industriels et universitaires internationaux.

Impliqué dans cette étude, le groupe SIS/MediaCoding à I3S a déjà commencé des travaux relatifs au stockage d'images sur ADN synthétique depuis plus de 4 ans (1 thèse soutenue et 2 thèses en cours, 2 postdocs et 1 ingénieur). Le projet de valorisation « Paircode » développé par l'équipe propose une solution de stockage des données numériques dites « froides » (rarement consultées) dans l'ADN synthétique en réponse à la croissance exponentielle des données et aux limitations des systèmes actuels (HDD, bande magnétique). Cette solution innovante s'appuie sur une technologie de codage brevetée (CNRS et UCA) qui profite des propriétés biologiques de la molécule d'ADN pour assurer une fiabilité de stockage pendant des centaines d'années tout en permettant une maîtrise du coût de la synthèse moléculaire. « Paircode » est accompagné par le programme i-PhD de BPI France ainsi que par RISE CNRS Innovation et la SATT Sud-est.

Un vivier d'excellents candidats a déjà été identifié.

Résumé du projet scientifique : 15 lignes maximum

La mémoire de l'humanité repose sur notre capacité à gérer des quantités de données de plus en plus importantes, sur des périodes de temps allant de quelques années à plusieurs siècles. Les outils actuels ne sont plus suffisants et il est nécessaire d'envisager des solutions qui puissent devenir rapidement opérationnelles. L'une des plus prometteuses est fournie par un stockage de l'information sous forme d'ADN, à l'instar de ce que réalisent les êtres vivants pour le codage de leur génome. L'ADN fournit en effet un moyen de stockage très stable sur de très longues périodes avec des conditions simples de mise en œuvre. Les enjeux autour du stockage ADN se positionnent aujourd'hui à l'échelle mondiale. Plusieurs verrous scientifiques en lien avec le stockage moléculaire restent encore à lever. En particulier, une attention particulière doit être apportée à la compression de l'information (informatique, nanotechnologies) pour réduire les coûts de synthèse de polymère et rendre le décodage robuste au bruit induit par le processus biochimique. L'objectif de ce projet scientifique est de quantifier les contraintes et la dégradation des signaux apportées par les différents procédés biotechnologiques. Pour cela, nous pouvons considérer le processus de stockage de l'ADN comme un canal de transmission (comme en communication numérique) qui génère des erreurs fournissant un signal bruité et nous modéliserons précisément le type d'erreurs (synthèse, conditionnement de stockage, dégradation à long terme de l'ADN, sélection des molécules, séquençage), pour finalement concevoir des codes correcteurs d'erreurs adaptés.

Résumé du projet d'enseignement : 15 lignes maximum

La personne recrutée pourrait s'investir dans la création et la direction d'un parcours de Master 2 orienté traitement du signal et des images, en coordination étroite sur les trois départements disciplinaires concernés, afin de renforcer cette thématique enseignement sur Sophia Antipolis, en adéquation avec les activités recherches de l'équipe SIS.

Diffusion scientifique : préciser les résultats attendus en termes de diffusion scientifique (publications, communications...)

- Revue à comité de lecture (public universitaire). Nous prévoyons des publications dans des revues du domaine ayant un facteur d'impact élevé telles que IEEE Transactions on Image Processing, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Information Theory, Nature etc.
- Conférences internationales (public de recherche universitaire et industrielle). La publication et la présentation lors de conférences du domaine telles que IEEE ICIP, IEEE ICASSP, Eusipco, ICPR, etc. sont particulièrement importantes pour le projet, car les conférences sont le moyen privilégié de diffusion.

Science ouverte : *le projet s'inscrit-il dans une démarche de science ouverte ? Si, oui décrire sa mise en œuvre.*

Nous nous attendons à ce que les connaissances liées au travail de recherche soient générées sous la forme de nouvelles techniques expérimentales ou de nouveaux outils logiciels. Les méthodes et les logiciels prototypes seront généralement disponibles gratuitement après publication. Le logiciel de production pourra être concédé sous licence aux parties intéressées. Les mises en œuvre prototypes pourront être protégées par des brevets.

Science et société : *le projet envisage-t-il une communication auprès du grand public ? Si oui : préciser de quelle manière et à quelle échéance*

- Site web du projet : L'objectif du site web est de fournir des moyens de diffusion du projet vers le monde extérieur, sa structure, ses résultats et sa documentation, des vidéos en ligne de nos technologies et des résultats de la recherche.
- Médias sociaux : Nous allons mettre en place des outils de diffusion tels que LinkedIn, Twitter ou Research Gate afin de développer une activité de diffusion de l'information avec ces nouveaux canaux.
- Médias : nous prévoyons d'utiliser les services de communication de l'UCA et du CNRS pour diffuser nos résultats et les rendre accessibles au grand public par les canaux traditionnels.
- Animation de la communauté internationale : contacts, présentations du projet, afin d'aller vers la création d'un projet européen.

Indicateurs : *préciser les indicateurs de suivi du déploiement du projet et la méthodologie de leur suivi*

Différents indicateurs scientifiques et technologiques peuvent être envisagés tels que :

- Qualité / impact des publications (revues, congrès)
- Qualité de l'encadrement (nombre de thèses soutenues, postdocs)
- Participation à des conférences, diffusion de l'information
- Effet levier : contrats ANR, contrats Européens, contrats industriels, et autres
- Brevets, valorisation
- Mise en place d'Unités d'Enseignement