

Thème 1 : Perspectives offertes par l'Intelligence Artificielle et la Data Science (30/9 14h30-18h00)



Animatrice : Anne-Françoise Adam-Blondon. Directrice de l'Unité de Ressources Genomic-Info (URGI), INRAE. Directrice Adjointe de l'IFB-Core, INRAE, CNRS, INSERM, CEA / Heads of Node ELIXIR-FR. Coordinatrice Scientifique de BRC4Plants, le pilier Plantes de l'Infrastructure RARE

Des présentations illustrant l'intégration de l'IA dans le domaine des sciences biologiques qui se traduit par des avancées concrètes et de nouvelles pistes de recherches. Cependant, ces progrès soulèvent également des nouveaux défis scientifiques, éthiques et sociétaux qu'il faut connaître et gérer.



Hervé Monod herve.monod@inrae.fr Directeur de recherche, UR MaIAGE. Directeur du métaprogramme DIGIT-BIO. INRAE

Développements de l'IA en cours dans les recherches d'INRAE.

La modélisation et la statistique, la science des données et les technologies du numérique constituent, de longue date, un pilier pour les recherches menées à INRAE dans les sciences du vivant et de l'environnement. Avec l'intelligence artificielle (IA), leurs développements les plus récents irriguent aujourd'hui tous les domaines de recherche, et apparaissent comme un levier pour contribuer à la transition vers des modes de gestion plus durables, que ce soit pour la détection précoce de maladies, le suivi des cultures, la gestion des ressources ou celle de l'exploitation agricole. La présentation illustrera comment l'IA est mobilisée de façon croissante dans de nombreuses recherches et dispositifs d'INRAE, générant de nouveaux enjeux et soulevant des questions de stratégie, d'organisation, d'animation et de formation. Comme pour d'autres domaines d'application, il s'agit d'adapter et de s'approprier les capacités inédites qu'apportent les différentes méthodes associées à l'IA, avec discernement et en bonne connaissance des risques et limites associés.



Sharif Islam sharif.islam@naturalis.nl Senior Data Architect, Naturalis Biodiversity Center, Leiden, The Netherlands.

From Samples to Systems: Building transparent, AI-Ready infrastructures for biological resource access.

This talk explores how FAIR-aligned research infrastructures connect physical biological resources with their digital representations. Drawing on experience from EU funded projects such as DiSSCo (natural science collections), Biodiversity Genomics Europe (DNA barcoding and whole-genome sequencing), and BioDT (Biodiversity Digital Twin), I will highlight the design foundations that enable this linkage. At the core are persistent identifiers and rich metadata,

which ensure traceability, provenance, and interoperability across collections and digital objects. Metadata not only documents the origin and context of samples but also supports automated workflows for discovery and reuse. These mechanisms enable integration across domains and scales, from individual physical specimens to global data networks. Achieving AI readiness requires infrastructures to go beyond data availability toward machine-actionable, structured, and linked resources. This means building transparent systems where provenance is consistently captured and workflows themselves are FAIR. I will discuss how these principles are being operationalised across European biodiversity research infrastructures, highlighting both challenges and success stories.

Lionel Da Cruz lionel.da-cruz@sante.gouv.fr chef du Pôle recherche et accès à l'innovation

Teddy Léguillier teddy.leguillier@sante.gouv.fr Conseiller en politiques médicales et scientifiques – Ministère de la santé – DGOS

IA et biobanques.

Résumé en attente

Stéphane Pouyllau, stephane.pouyllau@huma-num.fr CNRS Cofondateur Human-Num Lab. Pr Evry-Saclay.

ISIDORE 2030 : adapter un moteur de recherche académique aux IA génératives. (visio)

Résumé en attente



David Makowski david.makowski@inrae.fr Directeur de recherche. Unité Mathématique et Informatique Appliquées Paris-Saclay. INRAE – Université Paris-Saclay. Palaiseau, France

AI tools for forecasting plant, animal, and human health: how to assess their reliability? (visio)

The prediction of diseases and health status of plants, animals, and humans has undergone a double technological revolution since the beginning of the 21st century. On the one hand, the available data characterizing health status has become considerably more diverse, increased in volume, and decreased in cost. On the other hand, new AI algorithms have been developed to use this data, relying on increasingly powerful computers. Thanks to these two revolutions, new diagnostic tools have been developed and are now available to both specialists and ordinary users. It is now possible to obtain a diagnosis quickly and at low cost, without the need to consult an expert, simply using a standard smartphone. These developments raise questions about the reliability of these new diagnosis tools, especially in comparison with human experts. A rigorous evaluation of AI forecasts is indeed essential not only from a scientific standpoint, but also and above all from a societal perspective. In this presentation, I will outline the challenges facing the scientific community in evaluating predictive tools based on machine learning. I discuss in particular the challenges posed by the acquisition of independent data necessary for the evaluation of predictive machine learning tools, by the selection of the most appropriate evaluation criteria, by the implementation of reliable and operational calculation chains, and by the risks of conflict of interest. In conclusion, I demonstrate the importance of establishing an independent and widely-recognized evaluation procedure to limit the risks associated with the use of accessible, inexpensive but unreliable tools that can lead to erroneous diagnosis and

unfortunate decisions. Our ability to make progress in the future will depend largely on our ability to distinguish between good and bad predictions.



Antoine Cornuejols antoine.cornuejols@agroparistech.fr Professeur d'Informatique à AgroParisTech. Membre de l'UMR AgroParisTech/INRAE MIA-Paris_Saclay. Responsable de l'équipe Ekinocs.

L'IA d'hier et d'aujourd'hui : que doit-on enseigner pour l'utiliser à bon escient ?

Depuis les années 2010, avec l'arrivée de l'apprentissage profond, l'IA est entrée en force dans nos laboratoires. L'abondance de données, des algorithmes nouveaux et une puissance de calcul en constant accroissement ont permis l'émergence d'une science guidée par les données. Cependant, l'irruption de l'IA générative à l'automne 2022 marque une rupture encore plus spectaculaire. Il est encore très difficile d'anticiper les changements sans doute profonds qui vont en résulter dans nos pratiques de travail et dans la société.

Mais ce qui est certain, c'est qu'il est impossible d'ignorer tant « l'IA d'avant 2022 », avec ses outils d'analyse de données très puissants, que « l'IA d'après » avec ses assistants et agents numériques désormais sollicités à une échelle stupéfiante dans la société et qui inondent toute la mémoire digitale de ses productions. Se pose alors la question de ce qui doit être enseigné pour un usage approprié de ces deux types d'IA.

Dans cet exposé, nous soulignerons, d'une part, ce qui est essentiel dans l'enseignement de l'IA d'avant, les concepts clés, les questions fondamentales et, d'autre part, ce qui caractérise l'IA d'après, ce qui la distingue de l'IA d'avant et donc ce qui doit faire l'objet d'une nouvelle attention.

Thème 2 : Outils numériques et usages (01/10 09h00-12h30)

Animateur : Valentin Loux.

Un ensemble d'exposés pour soulever la complexité des différentes étapes et les questions pratiques qui se posent aux CRB liées à l'utilisation du numériques en matière de gestion des ressources biologiques.



David Rousseau david.rousseau@univ-angers.fr Professeur des Universités, UMR INRAe IRHS

Intelligence artificielle et Centres de Ressources Biologiques : vers une nouvelle ère de valorisation des données.

Le Professeur David Rousseau présentera les apports récents et les perspectives de l'intelligence artificielle dans la gestion, l'analyse et la valorisation des données. À travers des exemples concrets, il montrera comment l'IA permet d'améliorer la qualité et l'interopérabilité des données, de faciliter l'annotation et la classification des ressources, et d'accélérer la découverte de nouvelles connaissances. Ce talk abordera également les défis associés à l'usage de l'IA – fiabilité des modèles, gouvernance des données, exigences éthiques et réglementaires – ainsi que les opportunités pour construire des stratégies numériques innovantes au service de la recherche et de l'innovation en recherche pour les sciences du vivant



Sandra Dérozier sandra.derozier@inrae.fr, Ingénieure de recherche en bioinformatique, INRAE Jouy-en-Josas, Unité MaIAGE.

Omnicrobe, an open-access database of microbial habitats and phenotypes using a comprehensive text mining and data fusion approach

The dramatic increase in the number of microbe descriptions in databases, reports, and papers presents a two-fold challenge for accessing the information: integration of heterogeneous data in a standard ontology-based representation and normalization of the textual descriptions by semantic analysis. Text mining methods offer powerful ways to extract textual information and generate ontology-based representation. The Omnicrobe application gathers comprehensive information on habitats, phenotypes, and usages of microbes from scientific sources of high interest to the microbiology community. The Omnicrobe database contains around 1 million descriptions of microbe properties. These descriptions are created by analyzing and combining seven information sources of various kinds, i.e., biological resource catalogs, sequence databases and scientific literature. The microbe properties are indexed by the Ontobiotopie ontology and their taxa are indexed by an extended version of the NCBI taxonomy. The Omnicrobe application covers all domains of microbiology. With simple or rich ontology-based queries, it provides easy-to-use support in the resolution of scientific questions related to the habitats, phenotypes, and uses of microbes.

Christine Prat christine.prat@univ-amu.fr Université Aix-Marseille.

Conséquences de la décision de Cali (DSI) pour les gestionnaires de CRB et les utilisateurs de ressources génétiques, dans le cadre de la recherche publique et des partenariats avec des entreprises privées.

Résumé en attente



Aline Bouchard, aline.bouchard@chartes.psl.eu Co-responsable de l'URFIST de Paris, conservateur des bibliothèques.

Scopus AI, WOS Research Assistant, Elicit, Consensus... : enjeux et limites de l'IA pour la revue de littérature.

ChatGPT, mode IA, recherche sémantique, deep research, assistants de recherche... La recherche bibliographique semble connaître depuis 3 ans des évolutions constantes. « Évolutions » ou « révolutions » ? Alors que l'IA générative gagne même les interfaces de recherche traditionnelles, dans quelles mesures les outils IA viennent-ils modifier la recherche bibliographique ? Sont-ils actuellement à la hauteur des promesses ?



Anne-Françoise Adam-Blondon anne-francoise.adam-blondon@inrae.fr

Directrice de l'Unité de Ressources Genomic-Info (URGI), INRAE. Directrice Adjointe de l'IFB-Core, INRAE, CNRS, INSERM, CEA / Heads of Node ELIXIR-FR. Coordinatrice Scientifique de BRC4Plants, le pilier Plantes de l'Infrastructure RARe.

BReIF and AgroDiv: success stories for data publication.

In the frame of the France 2030 programme Agroecology & Numeric, the AgroDiv project aims at harnessing the French collections of plant and domestic animal's genetic resources in order to facilitate the identification of materials in a context of increasing the diversity and the resilience of crops and animals. For that purpose, around 20K plant accessions and 7,5K animals are planned to be genotyped or sequenced by a large consortium of French public research teams supported by three French research infrastructures, RARE, INRAE Genomics and BioinfOmics/IFB. In parallel, the BRIF project gathers the three infrastructures to develop interoperable services aiming at a seamless data management, following the FAIR principles and supporting the needs of the data analysis.

Amélie Fiocca amelie.fiocca@inrae.fr ingénieure et chargée de projets pour la DipSO (Direction pour la Science Ouverte) à INRAE.

Ecosystème des PIDs et recommandations à venir.

Les PIDs (identifiants pérennes) permettent de citer de manière fiable et précise des objets, ressources et productions de la recherche comme les publications, les données, ou encore les logiciels), renforçant ainsi la reconnaissance des auteurs et des organisations.

Pour citer un des plus connus, l'identifiant ORCID est devenu un standard pour identifier de manière unique les personnels qui contribuent aux travaux de recherche.

Les PIDs assurent donc le suivi de l'évolution des concepts (personnes, structures, productions, ...) tout au long de leur existence, en conservant un historique détaillé de leurs changements.

Amélie Fiocca amelie.fiocca@inrae.fr ingénieure et chargée de projets pour la DipSO (Direction pour la Science Ouverte) à INRAE.

Les entrepôts de données et la qualité du dépôt.

Présentation de l'écosystème Recherche Data Gouv et le cas de l'entrepôt Data INRAE.

Nous présenterons quelles sont les bonnes pratiques de dépôts et ce qu'est la curation documentaire effectuée par notre équipe Data INRAE.