

Colloque organisé par l'École Doctorale Galilée,  
Université Paris 13

2<sup>E</sup> ÉDITION

# Une vision transversale des sciences

Thème : La science et les données

Aloïs Brunel (Informaticien)  
Antoine Danchin (Mathématicien)  
Christel Daniel (Médecin)  
Nathalie Martial-Braz (Juriste)  
Bernard Srouf (Épidémiologiste)



**Mardi 07 mai 2019**

Institut Henri Poincaré,  
Amphithéâtre Hermite  
11 rue Pierre et Marie Curie  
75231 Paris CEDEX 05



$$V = \frac{4}{3} \pi r^3$$



**Inscription obligatoire** (mais pas de frais d'enregistrement)  
avant le 02 mai 2019 (ecole-doctorale.galilee@univ-paris13.fr)

**USPC**  
Université Sorbonne  
Paris Cité

**CAMPUS**   
**CONDORCET**  
Paris - Aubervilliers

## **Le mot de la direction**

L'École Doctorale « Sciences, Technologie, Santé - Galilée » organise cette année la deuxième édition de la journée « Une vision transversale des sciences ». Notre École Doctorale, comme son nom l'indique, est une école pluridisciplinaire. Elle couvre un spectre très large de disciplines scientifiques (Chimie, Ethologie, Informatique, Informatique biomédicale, Traitement de l'information, Mathématiques, Physique, Santé et Santé publique, Sciences des matériaux, Sciences pour l'ingénieur, Sciences de la vie et de la santé).

Au-delà de cette pluridisciplinarité, elle aspire à tisser des liens féconds entre ces disciplines, à créer de l'interdisciplinarité. Cette journée est une des initiatives menées en direction de cette action.

L'objectif principal de cette conférence est donc de montrer les synergies qui peuvent naître des travaux à l'interface de plusieurs disciplines ; montrer comment les sciences peuvent s'entrelacer pour produire de nouvelles théories et de nouveaux concepts ; comment un regard ouvert et constructif sur le monde dans sa diversité peut permettre de découvrir des horizons nouveaux. L'Union Européenne et la France ont clairement décidé de promouvoir l'interdisciplinarité en recherche. Pour témoin, les défis sociétaux de l'appel à projet Horizon 2020 sont tous intrinsèquement transdisciplinaires.

Cette journée est aussi le moment pour rappeler que la science du XXI<sup>ème</sup> siècle est et sera la science des actions en commun. Que la richesse et la profondeur de nos disciplines demandent maintenant d'interagir entre nous, car le chercheur omniscient est sans doute une espèce éteinte ! On ne peut concevoir les grands défis de l'informatique sans les mathématiques. L'apprentissage artificiel nécessitent de mutualiser les forces et les compétences des informaticien et des mathématiciens. Les nano-technologies mettent en œuvre conjointement chimie, physique et électronique. De même, le séquençage du génome allie biologie et informatique. Les prothèses et implants nécessitent des compétences en physique des matériaux, en mécanique et en biologie... Et nous serions, si nombreux sont-ils, bien incapable de faire une liste exhaustive des défis appelant à des échanges entre les disciplines. Il paraît donc aujourd'hui plus que jamais nécessaire de joindre nos forces afin de progresser sur ces questions contemporaines.

Nous avons souhaité dans cette deuxième édition construire un programme autour du thème « La science et les données » afin de mieux apprécier les interactions apodictiques et les regards croisés sur ce même objet d'étude. Les cinq exposés qui seront présentés forment cinq exemples de recherches fructueuses à l'interface de champs disciplinaires : Santé publique et Informatique (base Nutrinet), Biologie et Mathématique (séquençage du génome), Informatique et société (deep learning), Sciences et Droit (protection et propriété des données), Médecine et Informatique (dossier numérique patient).

Nous vous souhaitons une journée enrichissante et scientifiquement épanouissante !

**La direction de l'école Doctorale 146**

**Dominique Ledoux**

**Olivier BODINI**

# PROGRAMME

09h00 - 9h30 Enregistrement

09h30 - 9h40 Introduction

09h40 - 10h55 *Antoine Danchin (Mathématicien et Biologiste, membre de l'Académie des sciences) :*  
Il n'y a pas d'intelligence de la foule. L'annotation des génomes au temps des données massives.

10h55 - 11h15 Pause café

11h15 - 12h30 *Bernard Srour (Docteur en pharmacie, Epidémiologiste, INSERM) :*  
NutriNet-Santé – première cohorte électronique dédiée à la nutrition et la santé.

12h30 - 13h45 Déjeuner (non compris)

13h45 - 15h00 *Nathalie Martial-Braz (Juriste, Université de Paris) :*  
Le cadre juridique de la protection des données.

15h00 - 16h15 *Christel Daniel (AP-HP, Sorbonne Université) :*  
L'Entrepôt de Données de Santé de l'AP-HP : une opportunité pour la recherche et la médecine personnalisée.

16h15 - 16h30 Pause café

16h30 - 17h45 *Alois Brunel (Informaticien, co-fondateur de la société DeepOmatic) :*  
Deep learning : quel impact sur notre société ?

17h45 - 17h50 Mot de la fin

## **Il n'y a pas d'intelligence de la foule. L'annotation des génomes au temps des données massives**

**Antoine Danchin** (Mathématicien et biologiste)

La recherche sur les génomes repose sur l'accumulation de données massives de séquences d'acides nucléiques associée à la connaissance transmise via l'annotation des gènes (métadonnées). En réponse au déluge des données de séquençage, les pratiques d'annotation standard utilisent des procédures automatiques reposant sur des règles de majorité. On considère que ce qui est suffisamment semblable a le même rôle. Cette façon de faire entrave le progrès en raison de la génération et de la propagation massive d'erreurs, conduisant les chercheurs dans des impasses. Plus subtilement, ce processus inductif décourage la découverte de la nouveauté, qui reste essentielle en recherche biologique et reflète la nature de la biologie elle-même. Dans ce contexte, les systèmes d'annotation, plutôt que simples collecteurs de faits, devraient être des outils qui proposent plusieurs modes d'inférence. C'est en combinant déduction, induction et abduction que les chercheurs pourront produire des hypothèses fructueuses lorsque la connaissance est extraite de bases de données bien construites. Pour dépasser la phénoménologie ou le simple diagnostic, il faut, en biologie récuser la position qui veut que « la séquence dise la structure et que la structure dise la fonction » en plaçant la fonction au premier plan. Cette vision sera illustrée par quelques exemples.

## **NutriNet-Santé - première cohorte électronique dédiée à la nutrition et la santé**

**Bernard Srour** (Épidémiologiste)

L'étude NutriNet-Santé est une cohorte sur Internet (n = 164 000), coordonnée par l'Equipe de Recherche en Epidémiologie Nutritionnelle et lancée en 2009 en France, pour étudier les associations entre nutrition et santé. Des questionnaires en ligne sont régulièrement utilisés pour collecter des informations et ainsi caractériser de manière approfondie le phénotype de chaque participant-e, notamment ses facteurs sociodémographiques, anthropométriques et de mode de vie (tabagisme, alcool, activité physique), le régime alimentaire, les médicaments, l'utilisation de compléments alimentaires et l'état de santé. Des données clinico-biologiques sont également disponibles.

NutriNet-Santé est donc une plateforme innovante qui permet en plus de la mine de données qu'elle constitue, d'administrer facilement des protocoles de recherche « personnalisables ».

## **Le cadre juridique de la protection des données.**

**Nathalie Martial-Braz** (Juriste)

L'intervention aura pour objet de présenter les grandes orientations réglementaires en matière de protection des données, personnelles et non personnelles. Le droit, construit désormais à l'échelle européenne, encadre très rigoureusement la protection des données à caractère personnel dans un texte connu de tous par son intitulé (RGPD) mais qui doit être présenté plus précisément pour déterminer les marges de manoeuvre pratique sur les données, notamment lorsqu'elles sont utilisées à des fins scientifiques. Quelles sont les données concernées par les textes, quelles sont les pratiques appréhendées par les textes, qui est responsable et de quoi, sont des questions qui seront notamment abordées au cours de l'intervention.

## **L'Entrepot de Données de Santé de l'AP-HP : une opportunité pour la recherche et la médecine personnalisée**

**Christel Daniel** (Médecin)

L'Assistance publique Hôpitaux de Paris (AP-HP) a engagé depuis plusieurs années un processus de transformation numérique dans l'optique d'offrir à ses professionnels et ses patients les services les plus performants et innovants. Dans ce contexte, une plateforme Big Data a été mise en œuvre pour héberger un Entrepôt de Données de Santé (EDS) qui centralise les données administratives, médicales et sociales de plus de 8 millions de patients. Cet EDS a permis à plus de 300 professionnels de l'AP-HP de réaliser des études sur les données des patients qu'ils ont pris en charge et également d'engager plus de 40 études multicentriques de plus grande envergure dont certaines en partenariat.

## Deep learning : quel impact sur notre société ?

**Aloïs Brunel** (Informaticien)

Nous proposons dans cet exposé de comprendre l'impact présent et futur de cette technologie sur nos sociétés. Des péages autoroutiers à la radiographie, en passant par le tri des déchets, ou encore la biologie et la traduction automatique: le deep learning s'immisce dans un nombre toujours croissant de domaines extrêmement variés. Nous commencerons par rappeler ce qu'est le deep learning, comment cette technologie s'inscrit dans le domaine plus large du machine learning, et comment elle se décline en différents domaines d'application (image, son, texte, etc.). Nous en verrons ensuite les potentialités, à travers de nombreux exemples d'applications réelles, mais également les limites actuelles, ce qui sera l'occasion de démonter quelques mythes et fantasmes couramment répandus. Nous tenterons également de dresser une vision prospective du deep learning en termes scientifiques et technologiques, mais aussi de dangers potentiels pour notre société.

## **Remerciements.**

L'Institut Henri Poincaré a été inauguré en 1928 sous l'impulsion des mathématiciens Émile Borel et George Birkhoff. C'est un haut lieu, internationalement reconnu, des Mathématiques en France.

Nous remercions vivement la direction de l'Institut d'avoir accepté de nous inviter dans ces locaux.

Nous remercions bien sûr tout particulièrement les orateurs qui ont accepté de donner leur temps à notre mission de formation doctorale.