



Journée de lancement

Jeudi 15 Février 2018



PRÉSENTATION DE LA FEUILLE DE ROUTE DE INSTITUT DATAIA

Institut Convergence 17-CONV-0003 INSTITUT DATAIA (I2-DRIVE)

Nozha Boujema
Directrice de L'Institut DATAIA
Directrice de Recherche Inria
nozha.boujema@inria.fr

inria
informatics mathematics

Février 2018

université
PARIS-SACLAY



Instituts Convergences : Contexte

Ambition de l'action « *Instituts Convergences* » :

- **Regrouper et structurer des expertises pluridisciplinaires de grande ampleur et de forte visibilité** pour mieux répondre à des enjeux majeurs **des sciences des données et de l'intelligence artificielle**, à la croisée des défis sociétaux et économiques et des questionnements de la communauté scientifique.
- Intégration poussée **recherche-formation**
- Couplage efficace avec **le monde socio-économique et les relations industrielles**



Institut DATAIA : Contexte

- **2 appels et 10 projets sélectionnés** toutes disciplines *confondues* (santé-médecine, climat, agriculture, migration, etc.) ;
- **Evaluation du jury international :**
 - > *Nous désigne comme le centre français des sciences des données et des ses interfaces*
- **Objectif :** être parmi les 5 premiers à l'international ;
- **Budget :** 10 M€ (180 M€ en global), 130 enseignants-chercheurs (trajectoire: 300 à 3 ans) ;
- **Date de fin :** 31/12/2025 (8,5 années) ; **Evaluation 1^{er} semestre 2022.**



Les membres fondateurs



- L'Institut DATAIA est porté par l'Université Paris-Saclay et dirigé par le centre de recherche Inria Saclay – Île-de-France :



- Le consortium rassemble des Universités, des Instituts de recherche nationaux et des Grandes Ecoles :



L'écosystème du site

IDEX



IRS NUTRIPERSO





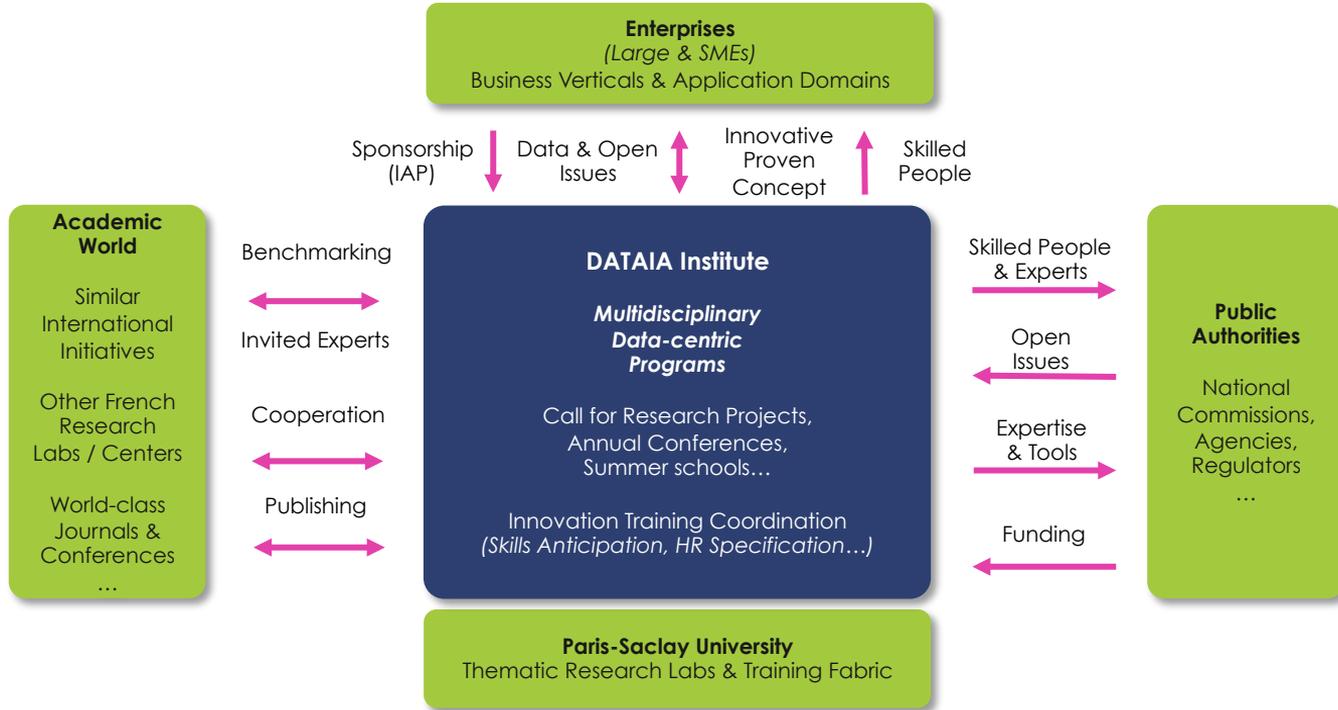
Les partenaires internationaux

- Alan Turing (UK)
- IVADO (Canada)
- Advanced Core Technologies for Big Data Integration (Japan)
- DSI (Data Science Institute – Columbia University)
- MIT IDSS (Institute for Data Science, Systems and Society)

**The
Alan Turing
Institute**



L'Institut DATAIA et son écosystème



Motivation

« 2 sides of the same coin »



Données & Algorithmes

- Les **données sont omniprésentes** dans les environnements personnel et professionnel
- Les **algorithmes** de gestion et d'analyse de ces données sont **de plus en plus présents** dans les services numériques
- Les bénéfices croissants des **technologies Big Data et des services IA** ont un impact important sur notre **économie** et notre **organisation sociale** : voitures autonomes, agents conversationnels, santé et bien-être, etc.

Motivation

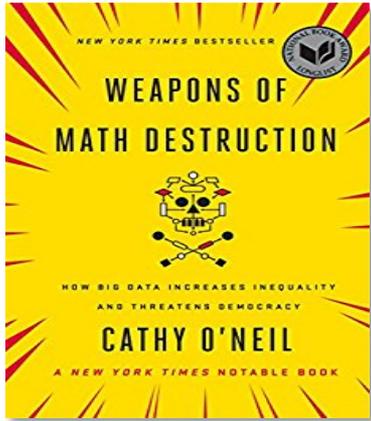
Données & Algorithmes

« 2 sides of the same coin »



- La **transparence** et l'**éthique** de tels **systèmes algorithmiques** (données et algorithmes) deviennent critiques pour la **confiance** et l'**appropriation** des services numériques.
- L'analyse des données évolue de la description du passé à l'analyse **prédictive** et **prescriptive** pour l'aide à la décision.
- Importance de remédier à l'**asymétrie informationnelle** entre le **producteur** du service numérique et son **consommateur**, qu'il soit citoyen ou professionnel - B2C ou B2B => **droits civiques, concurrence, souveraineté**

A l'international



Weapons of Math Destruction
How Big Data Increases Inequality and Threatens Democracy



Federal Trade Commission
January 2016



Office of Science and Technology Policy - WH
May 2016



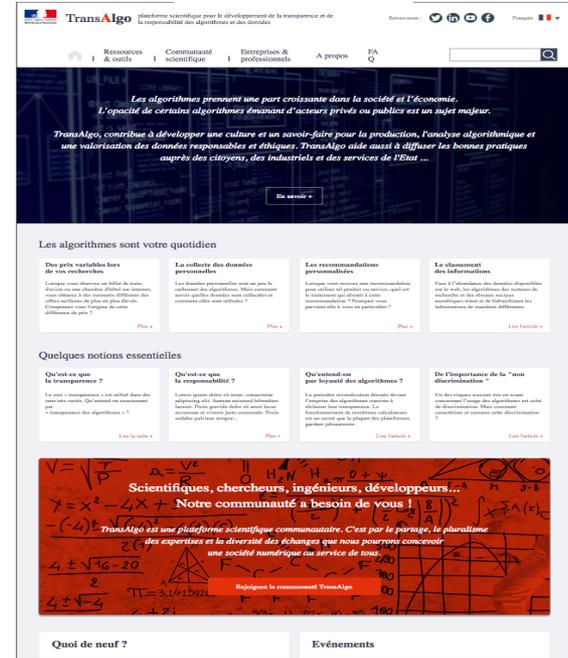
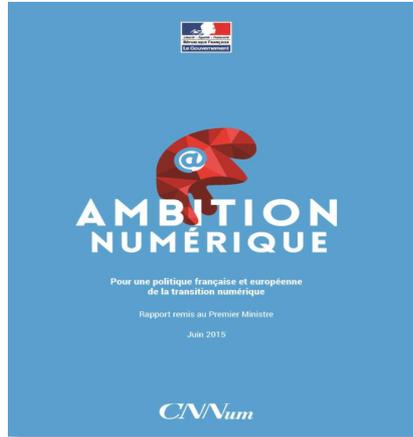
Amazon, Apple, Google, Facebook, IBM, Microsoft



« *Explainable AI* »
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY



En France



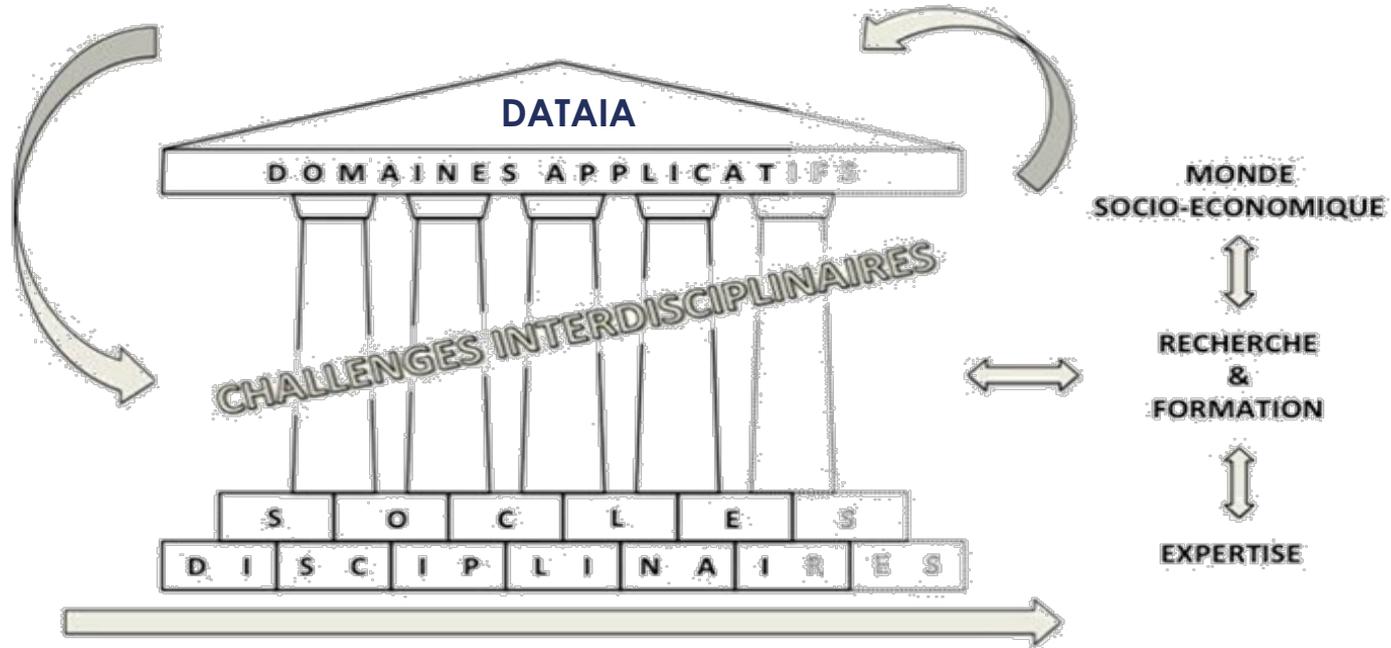


DATAIA - Sciences des données, Intelligence & Société

- La transformation numérique par les sciences de la donnée présente des **verrous scientifiques, technologiques, juridiques, économiques et éthiques totalement interdépendants**.
 - La richesse interdisciplinaire des **différents acteurs** et **établissements du plateau de Saclay** offre **un terrain unique sur le territoire français** pour la mise en place d'un Institut Convergences autour des **sciences des données et de leurs interfaces**: disciplinaires et enjeux socio-économiques.
 - Vocation à développer un **réseau scientifique national**
- => Fer de lance de la stratégie française** en sciences des données et IA et leurs impacts socio-économiques



DATAIA : Axes de travail





Les défis interdisciplinaires

- **Apprentissage et Intelligence Artificielle**
 - Apprentissage automatique innovant et IA : sens commun, adaptabilité, généralisation
 - Apprentissage profond et apprentissage adversarial
 - Apprentissage automatique et hyper-optimisation
 - Optimisation pour l'apprentissage, e.g. améliorations des méthodes de gradient stochastiques, optimisation Bayésienne), optimisation combinatoire
 - Lien apprentissage-modélisation, intégration d'à priori dans l'apprentissage
 - Reproductibilité et apprentissage robuste
 - Inférence statistique et validation
 - Compositionnalité des architectures profondes.



Les défis interdisciplinaires

- **Des données à la connaissance, des données à la décision**
 - Données hétérogènes, complexes, incomplètes, semi-structurées et/ou incertaines
 - Fast big data : structuration de la donnée pour pouvoir l'exploiter
 - Apprentissage en ligne, méthodologie pour les données massives, méthodes efficaces
 - Amélioration du stockage, calcul et estimation pour la science des données
 - Modélisation des interactions entre agents (humains ou artificiels) par théorie des jeux.
 - Représentation et algorithmes multi-échelle et multimodaux
 - Analyse théorique de méthodes heuristiques (théorie de la complexité, géométrie de l'information, théorie des chaînes de Markov)
 - Coévolution Humain-Machine dans les systèmes autonomes : agents conversationnels, voitures, robots sociaux



Les défis interdisciplinaires

- **Transparence, responsabilité et éthique**
 - Responsabilité-by-design, Explicabilité-by-design
 - Transparence-by-design, équité-by-design
 - Audit des systèmes algorithmiques : non-discrimination , loyauté, biais techniques, neutralité, équité
 - Mesure de la confiance et de l'appropriation du numérique
 - « Progressive user-centric-analytics » (monitoring interactif des systèmes de décisions : dataviz, dashboards, IHM)
 - Responsabilité du traitement de l'information et de la prise de décision : contrôle d'usage des données et fact-checking
 - Découverte causale, traçabilité des inférences à partir des données sources, interprétabilité des architectures profondes



Les défis interdisciplinaires

- **Protection, régulation et économie des données**
 - « Privacy-by-design », GDPR
 - Apprentissage respectueux de la vie privée (« *differential privacy* »)
 - Développement de méthodologies éthiquement responsables, et de technologies pour réguler la collection, l'usage et le traitement des données personnelles, et l'exploitation des connaissances tirées de ces données
 - Sécurité informatique des chaînes de traitement de données
 - Sécurité/crypto : block-chain et tiers de confiance



Socles disciplinaires

- Statistiques, science des données et décision
 - Identification du thème « infrastructures basse consommation d'énergie » pour le moyen terme
- Sciences sociales
- Stratégie, management et économie
- Sciences juridiques



Exemples de domaines d'application

- Internet des humains et des objets (*Média, Réseaux sociaux...*)
- Urbanisation 4.0 et mobilité (*smart city, véhicules autonomes...*)
- Santé, Nutrition et Bien-être
- Sécurité et justice
- Finance, banque, assurance...
- Optimisation de la production et de la gestion d'énergie
- e-Sciences



Formation et recherche

- Trois formations doctorales de l'Université Paris-Saclay : EDMH, ED STIC et ED SHS.
- Renforcer le **croisement math-info** dans les formations sciences des données, nouveaux **parcours interdisciplinaires** plus ouverts sur les **SHS : sensibilisation à la responsabilité** des systèmes algorithmiques, aux modèles économiques, aux droits et usages des données.
- **AAP recherche** – 3 ans, 2 bourses de thèses (ou thèse + un post-doc/ingénieur).
- **Mobilités internationales** étudiantes (entrantes et sortantes) avec 2 bourses de thèse (bourses d'excellence) par an.
- **Semestres thématiques** pour MSc / PhD /E-C, Conférence biannuelle, symposium annuel d'**auto-évaluation**, Workshops, **Challenges**, Junior conférence, Summer-school.



Programme d’Affiliation Industrielle*

- Apports : soutien à la recherche, données et uses-cases.
- Participation à la définition, à la sélection et au suivi des programmes.
- Participation à la définition de la vision stratégique à long terme.
- Workshops, sessions d’échange de travaux S&T, brainstorming sessions (problèmes ouverts), etc.
- La PI suivra les règles définies dans un accord de consortium.
- Droit de premier regard sur la PI.

* *A partir de ce qui se fait dans les grandes universités américaines (Modèle de Stanford)*

Co-working

- Des espaces de travail sont disponibles pour les équipes affiliées à l'Institut DATAIA dans le bâtiment Alan Turing, lieu emblématique :
 - 1800 m² dont environ 300 m² pour les nouvelles équipes
 - 8 équipes sur place
 - 800 m² d'espaces de réunions
- Mise en place d'écrans de télé-présence en cours.



MERCI DE VOTRE ATTENTION

nozha.boujemaa@inria.fr



Scientific actions

- PhD and post doc grants program
- Bis annual Conference, Annual symposium of self assessment, Workshops (10/year)
- Thematic quarters
- **Interdisciplinary training curricula** and 3 MOOCs
- International mobility for master students
- Eight Junior Conferences
- Summer School
- Internships
- Challenges (1 per year)

INSTITUT

DATAiA

Science des données, Intelligence & Société

