



NLP

université
PARIS-SACLAY



université
PARIS-SACLAY

UVSQ
université PARIS-SACLAY



CentraleSupélec

école
normale
supérieure
paris-saclay

AgroParisTech



INRAE

Inria



ONERA
THE FRENCH AEROSPACE LAB

FM
JH
FONDATION MATHÉMATIQUE
JACQUES HADAMARD

CentraleSupélec
EXE

GUSTAVE
ROUSSY
CANCER CAMPUS
GRAND PARIS

Institut Mines-Télécom
Business School

INSTITUT
d'OPTIQUE
GRADUATE SCHOOL
ParisTech

IHES
Institut des Hautes Études Scientifiques

DATAIA PARIS-SACLAY INSTITUTE

Located within the **Paris-Saclay University** (12th Shanghai ranking), it is the **first French ecosystem in Data Sciences, AI and their societal impacts.**

MISSION

To bring together **multidisciplinary expertise and boost the collective strength of its partners** in the Paris-Saclay cluster with the aim of combining big data and AI technologies with social sciences and humanities for an AI at the service of humans.

IN FIGURES

14

DATAIA members

47

laboratories
partners

800

full-time
researchers

10

IA chairs out of
40 national

30

IA theses

450

PhD students
per year



The Industrial Affiliation Plan (PAI) aims to boost the collective strength of the Institute's academic ecosystem and its industrial members. The services offered in response to the respective needs expressed include:

- Joint actions to support research;
 - Sharing of experiences and collective needs;
 - Facilitated access to recruitment;
 - Access to training, seminars, workshops, etc.;
 - Implementation of dedicated events (hackathons, challenges, etc.);
 - Access to working places to increase exchanges.
-



The D2C system aims **upstream**, to present the priority research issues and to match them with the problems of industry. **Downstream**, to monitor contacts and opportunities for collaboration identified until they are set up and launched. It is part of the ambition to facilitate the establishment of several levels of collaboration and create a constructive dynamic:

1. Expertise / Student projects / Internships
2. Research collaborations / CIFRE theses
3. Joint laboratories / Joint teams
4. Multi-partner chairs

OBJECTIVES & PROGRAM



The main objectives of this D2C are focusing on :

- Identification of bio-marker signatures on treatment and disease progression;
- Optimization of biomarker selection by unsupervised analysis on multi-omics data;
- Integration of data from biomedical databased to study gene function.

2PM - 2:10PM

Introduction

2:10PM - 3PM

DATAIA researchers pitches on prospective research topics followed by industrialists on related issues

3PM - 4PM

One-on-one meetings with a view to setting up new collaborations

4PM - 4:10PM

Conclusion and action plan

DATAIA RESEARCHERS



Extraction, normalization, acquisition of ontologies

Claire Nédellec (INRAE, MalAge)

Integration, querying and classification of biomedical database



Collection of social network data on a private database

Fabian Suchanek (Télécom Paris, LCTI)

Data mining, automated reasoning and knowledge bases



Linguistic analysis on knowledge processing and dialogue

Ioana Vasilescu (Paris-Saclay University, LISN)

Language analysis, speech recognition and automatic language processing



Cross-reference textual and visual information to define better multimodal representations

Bianca Vieru (CEA-List, LASTI)

Multimedia content extraction, classification and semantic analysis



Study and application of medical texts

Pierre Zweigenbaum (Paris-Sorbonne University, LISN)

Automatic French language processing

DATAIA RESEARCHERS



Knowledge extraction from patents

Sylvain Desroziers (IFPEN)

Natural language processing and deep learning



Structured prediction and introducing linguistic knowledge in neural networks

Caio Filippo-Corro (Paris-Saclay University, LISN)

Natural language processing and automatic learning



Immediate memory representation, link with neural models

Patrick Paroubek (Paris-Saclay University, LISN)

Evaluation in natural language analysis, text sentiment, chatbots



The problem of acquiring knowledge from texts

Fatiha Saïb (Paris-Saclay University, LISN)

Semantic data integration and enrichment, reference reconciliation



Machine translation using statistical methods

François Yvon (Paris-Saclay University, LISN)

Symbolic and statistical learning for language and speech processing

DATAIA CLUB PAI COMPANIES



GE Healthcare

Build a conversational bot to interact and respond to clinicians' enquiries

Diana Batista - Data Scientist

Benoît Doreau - Software engineering manager

**Renault
Group**

Explicability of an NLP model based on a pre-entrained language model

Vincent Feuillard - Data Scientist

Ayhan Uyanik - Data Scientist

sanofi

Improving the management of ontologies through databases of clinical trial reports or scientific publications

Maali Minsari - Data Scientist

Marc Bianciotto - Data Scientist

Kun Mi - AI expert in knowledge graph and DL

SERVIER 
moved by you

NLP tools for processing patent databases

Thierry Dorval - Head of Data Science Lab

Sofia Lofti - Data Scientist

GUEST COMPANIES



**Abstractive text summarization
on contractual/technical documents**

Thomas Bourgeois - Data Scientist
Mehdi Rahim - PhD Senior R&D Scientist



**Natural language for semantic
interoperability between all objects**

Thierry Grenot - CEO

LETRIA

**Knowledge structuring and graphs
in support on ontologies**

Charles Borderie - CEO

OPSCIDIA

The value of knowledge

**Navigation and retrieval in
the scientific literature**

Sylvain Massip - CEO



**Conversational search capabilities
to multi-languages**

José Marcos-Rodriguez - CEO

INSTITUTIONAL PARTNERS





Eric TORDJEMAN

Head of Industrial Partnerships @DATAIA

eric.tordjeman@universite-paris-saclay.fr

 [eric-tordjeman](https://www.linkedin.com/in/eric-tordjeman)

DATAIA Paris-Saclay Institute

Université Paris-Saclay - Campus CentraleSupélec
3 rue Joliot Curie
91190 Gif-sur-Yvette

Communication Department

com-dataia@inria.fr

 www.dataia.eu

 [@institut-dataia](https://www.linkedin.com/company/institut-dataia)