

D2C

DATAIA CLUB CONNECTION

NLP

April 7, 2021

université
PARIS-SACLAY

INSTITUT DATAIA
Science des données, Intelligence & Société

The DATAIA Paris-Saclay Institute

Located within the Paris Saclay University (16th in the Shanghai ranking, 1st in mathematics), 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



12

academic members



42

partner laboratories



1200

professors-researchers



10

IA Chairs out of 40 national



18

research projects launched



450

PhD Students

Industrial Affiliation Plan (PAI)

The Industrial Affiliation Plan (IAP) 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.

DATAIA Club Connection (D2C)

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 and program

The main objectives of the D2C « *NLP* » to address are:

- 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 databases to study gene function.

2pm - 2:10pm

Introduction

2:10pm - 3pm

DATAIA researcher pitches on prospective research topics followed by industry presentations 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

Etude et application sur les textes médicaux



Pierre Zweigenbaum (Université Paris-Sorbonne, LISN)
Research: traitement automatique de la langue médicale

Extraction, normalization, acquisition of ontologies



Claire Nedellec (INRAE, MaIAGE)
Research: integration, querying and classification of biomedical database data

Collection of social network data on a private database



Fabian Suchanek (Télécom Paris, LCTI)
Research: data mining, automated reasoning and knowledge bases

Linguistic analysis on knowledge processing and dialogue



Ioana Vasilescu (Paris-Saclay University, LISN)
Research: language analysis, speech recognition and automatic language processing

Croiser les informations textuelles et visuelles pour définir de meilleures représentations multimodales



Bianca Vieu (CEA-List, LASTI)
Research: extraction de contenus multimédia, classification et analyse sémantique

DATAIA researchers

The problem of acquiring knowledge from texts



Fatiha Saïs (Paris-Saclay University, LISN)

Research: semantic data integration and enrichment, reference reconciliation

Machine translation using statistical methods



François Yvon (Paris-Saclay University, LISN)

Research: symbolic and statistical learning for language and speech processing

Structured prediction and introducing linguistic knowledge in neural networks



Caio Filippo Corro (Paris-Saclay University, LISN)

Research: natural language processing and automatic learning

Knowledge extraction from patents



Sylvain Desrozier (IFPEN)

Research: natural language processing and deep learning

Intermediate memory representation, link with neural models



Patrick Paroubek (Paris-Saclay University, LISN)

Research: evaluation in natural language analysis, text, sentiment, chatbots

DATAIA Club PAI Companies

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



GE Healthcare

Diana Batista - Data Scientist

Benoît Doreau - Software engineering
manager

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

**GROUPE
RENAULT**

Vincent Feuillard - Data Scientist

Ayhan Uyanik - Data Scientist

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

SANOFI 

Maali Mnasri - Data Scientist

Marc Bianciotto - Data Scientist

Kun Mi - AI expert in knowledge
graph and DL

NLP tools for processing patent databases



Thierry Dorval - Head of data science lab

Sofia Loffi - Data Scientist

Invited companies

Abstractive text summarization
on contractuels/technicals documents



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

Natural language for semantic interoperability between all objects



Thierry Grenot - CEO

Knowledge structuring and graphs in support of ontologies



Charles Borderie - CEO

Navigation and retrieval in the scientific literature



Sylvain Massip - CEO

Conversational search capabilities to multi-languages



José Marcos Rodriguez - CEO

Institutional partners



université
PARIS-SACLAY

INSTITUT DATAIA
Science des données, Intelligence & Société

Institut Convergence 17-CONV-0003 INSTITUT DATAIA (I2DRIVE)

Institut DATAIA Paris-Saclay

Centre de Recherche Inria Saclay - Île-de-France
Campus de l'Ecole Polytechnique - Bâtiment Alan Turing
1 rue Honoré d'Estienne d'Orves
91120 Palaiseau

Service Communication
com-dataia@inria.fr

 www.dataia.eu

 [@institut_dataia](https://twitter.com/institut_dataia)