



Inria Chair of Junior Professor

Supporting institution/organization: Inria center at the University of Lille
Head of the institution/organization: Mireille REGNIER
Site concerned: *ISITE-ULNE*
Academic Region: Hauts de France

Partner institutions/organizations: *CHU Lille, Inserm, Université de Lille*

Project name:

ENDotyping Chronic Inflammatory Diseases: how to obtain a unified representation of patients from heterogeneous data?

Acronym: ENDOMIC

Scientific topic: digital health

Keywords:

Autoimmune diseases; endotypes: clinical heterogeneity; unsupervised analysis; data science

Target duration: 6 years

Financial overview: 1 070k € for the project

Section (s) CNU/CoNRS/CSS corresponding: CNU 26/27, CoNRS 6/41- CSS 5/6

Contact : cpi-lille@inria.fr

Establishment strategy:

Digital Life Sciences and Health is a major area of research and innovation at Inria, and the recent program to create joint teams with Inserm is giving it new impetus. A strategic challenge is the structuring and processing of massive health data from hospitals while ensuring interoperability and preserving confidentiality. This national scientific axis of Inria is consistent with the policy of the Lille I-site. The Lille University Hospital has set up an institutional clinical data warehouse (CDW) called **INCLUDE** (INtegration Center of the Lille University hospital for Data Exploration), in which Inria Lille Nord Europe and the University of Lille are partners. This CDW (of which Professor Vincent Sobanski is a co-founder) gathers 1.6 million patients received at the Lille University Hospital since 2008 (CNIL GDPR-compliant authorization obtained in 2019). INCLUDE is part of the national dynamic of the *Health Data Hub* consisting of setting up **regional Hubs** bringing together data producers and operating skills. A major asset is its membership in the European Health Data & Evidence Network (EHDEN), which harmonizes data sources in OMOP interoperable format.

Strategy of the host laboratory:

The Inria center and the University of Lille (UMRs Cristal and Paul Painlevé) are developing health applications with the University Hospital of Lille (UMR INFINITE) in a close collaboration that has intensified in 2020. This project aims to **strengthen and affirm the "Digital Health" axis** as a structuring and strategic axis for the Inria Lille - Nord Europe center. A short-term objective is the **creation of a new Inria-INSERM-CHU-University of Lille project team**. It is integrated into the "Precision Health" and "Digital Transition for Humans" hubs of excellence of the Lille-Nord Europe University I-site. ENDOMIC is part of **the I-site "PHENOMIX cluster" research program**, which aims to develop a set of tools, models and software components to identify homogeneous groups of patients from massive data. It is **consistent with the scientific projects** of joint Inria and University of Lille teams: patient pathways (UMR Paul Painlevé), health data protection (Cristal) including the FLAMED Exploratory Action led by Aurélien Bellet in collaboration with INCLUDE on federated computing within hospital networks (2 joint publications) and man-machine interface (Cristal).

Summary of the scientific project:

Chronic inflammatory diseases (CID) are characterized by significant **clinical heterogeneity** which complicates patient management. A key concept in personalized medicine is "endotype" as a disease subtype defined by a molecular mechanism or a particular response to treatment. The "ENDOMIC" team led by Prof. Vincent Sobanski (Lille University and University Hospital of Lille, Junior IUF member) aims at revealing endotypes of CID through an integrative multidimensional analysis of highly phenotyped patients. It is based on their clinical and follow-up data (low dimensional information) and on multi-omics profiles generated specifically for this project (high dimensional information). Selected among the five national winners of the AMI "Inserm-Inria Digital Health 2020 project-team", this project constitutes a major advance towards an innovative classification because it integrates a very large number of variables by taking into account **the temporal dimension** and aims to design **sequential validation** tools to verify the relevance of such an integrative classification. An important originality of this project is to work on making the endotypes intelligible and exploitable by developing a **man-machine interface** allowing the actors (researchers, practitioners and even patients) to appropriate this new way of classifying a disease. The relationship between caregiver and patient will be profoundly modified, which will be the subject of an in-depth evaluation.

Summary of the teaching project:

The candidate will participate in the teaching related to **precision health** (Health and Digital Hubs of the ULNE I-SITE in the framework of the Experimental Public Establishment project) by bringing his/her **specific expertise** (data science, artificial intelligence, unsupervised analysis). This "precision health" graduate program is open to biologists, physicians and pharmacists, as well as to (bio)computer scientists of national and international origin, in order to **train the next generation of research leaders** for academia, industry and the biotechnology world. It includes a Master's program in Biology and Health (M2) and a PhD program, **open to the entire scientific community**. It is structured around interactive seminars on current cutting-edge technologies and approaches (e.g., big data, OMICS, next-generation sequencing, systems biology and integrated bioinformatics, AI, mathematical modeling of biological systems, innovative therapies, cohort management, health economics and ethics). He/she will be able to develop specific teaching or join the teaching team in order to structure and lead this **strategic training offer for the University of Lille**.

Scientific diffusion:

The chairholder will be required to promote the results obtained through publications in **recognized scientific journals** (target rank A) and in **international conferences** of reference for the computer science and health disciplines. The transdisciplinary nature of the research conducted will lead to the publication of articles in **generalist journals with a large audience** (core medical journals) and to the writing of chapters or educational books. He/she should also have a certain appetite for **scientific popularization**, particularly with the general public (university open days, high school students, French science days).

Open Science:

The man-machine interface will allow **navigation in the virtual representation of endotypes**. Healthcare professionals will be able to enter certain variables in a query module that will allow them to predict the potential assignment of a given patient to the endotypes. In addition, they will be able to implement their data in the form of an **interoperable database** that can be integrated into the algorithm and thus enrich the definition of endotypes (**open science**). Patients and health professionals will be asked about their **needs**. They will be involved in **co-creation activities** to finally obtain specifications that meet their needs and expectations. The specifications of this interface will also integrate the possibility for a health professional to position a given patient in the endotypes from a few variables entered in the model, the interface providing a **degree of confidence for this classification** (probabilistic projection). The algorithm should also be able to suggest to the healthcare professional the examinations to be performed (additional variables) to improve the classification of his patient, or even to guide his management, as proposed by the DETECT algorithm implemented a few years ago for the decision to perform a right heart catheterization in patients with systemic scleroderma suspected of having pulmonary hypertension.

The work done on the interoperability of data in OMOP format within the Lille University Hospital will allow the tool to be open "by design". The data necessary for the development of the algorithm will be available according to the **principles of open science**, in accordance with our recent publications (we are members of the EHDEN initiative (European Health Data & Evidence Network)).

Science and society:

The appropriation of this new tool by the actors will be ensured by publications in **open access journals**, participation in **scientific and patient congresses**, **documentation on the website**, communication on **social networks** with professionals and patient associations (e.g. in the networks in which we are already involved: FAI2R, EUSTAR) and **scientific seminars**.

We are very involved in **scientific communication, especially with the general public** (Prof. Vincent Sobanski is Vice-Dean of Communication at the University of Lille's Health and Sports Sciences Department). The ULNE I-SITE Taskforce has initiated a series of conferences for the general public on Covid-19. A YouTube platform has been set up, whose videos have been seen 27,000 times and widely relayed in the media and labelled by the Ministry of Higher Education and Research.

Indicators:

Primary indicators	
Short term (<2 years)	<ol style="list-style-type: none"> 1. Recruitment of engineers, doctoral students and post-doctoral fellows constituting the Chair 2. Participation in the submission of national or international "health" projects (ANR, ERC, European network...) in the field of ENDOMIC
Mid term (2-4 years)	<ol style="list-style-type: none"> 1. Publications in high level scientific journals 2. Establishment of a specific teaching program such as a university diploma or certificate graduate program (structured by the chairholder) 3. Submission by the chairholder of "computer science" projects of national or international scope (ANR, ERC, European network...) in the thematic of the ENDOMIC team
Long term (4-6 years)	<ol style="list-style-type: none"> 1. Open science dynamic (actions in the framework of EHDEN and AEx FLAMED, making public datasets and codes available, integration of new FAIR data in the integrative classification algorithm) 2. Transposition of the classification tool to other medical fields (new collaborations, creation of a start-up)
Secondary indicators	
Short term (<2 years)	<ol style="list-style-type: none"> 1. Publications in intermediate level computer or medical journals 2. Participation in international conferences of scientific societies
Mid term (2-4 years)	<ol style="list-style-type: none"> 1. Integration into research networks, particularly European ones (international visibility) 2. Organization of an international scientific seminar
Long term (4-6 years)	<ol style="list-style-type: none"> 1. Academic collaborations, in particular capacity of the ENDOMIC team to send/receive students/researchers in international mobility 2. Establishment of partnership contracts