



**Offer #2024-07437**

## **Post-Doctoral Research Visit F/M Postdoctoral Researcher Position in Artificial Intelligence for Omics Data Analysis**

**Contract type :** Fixed-term contract

**Level of qualifications required :** PhD or equivalent

**Fonction :** Post-Doctoral Research Visit

### **About the research centre or Inria department**

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regional economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

### **Context**

Recent advances in computational mass spectrometry-based metabolomics have unleashed a massive amount of chemical knowledge from metabolomics analysis. Yet, there is a need for a comprehensive computational framework that can better integrate the information derived from both experimental and computational analyses and help chemists analyse their results.

We invite applications for a Postdoctoral Researcher position at the Nice Chemistry Institute (ICN) and the Wimmics team at Inria both hosted at the Université Côte d'Azur. The selected candidate will be jointly embedded and interact with PhD students and research engineers of the Interdisciplinary Institute for Artificial Intelligence (3iA) TechPool, and operates in the context of international collaboration between France and Switzerland. The position is initially offered for one year, with the possibility of extension up to three years.

Key contacts are:

- Louis-Felix Nothias (project leader), Nice Chemistry Institute (ICN), Université Côte d'Azur, at [Louis-Felix.NOTHIAS@univ-cotedazur.fr](mailto:Louis-Felix.NOTHIAS@univ-cotedazur.fr)
- Fabien Gandon, Inria Université Côte d'Azur, at [Fabien.Gandon@inria.fr](mailto:Fabien.Gandon@inria.fr)

### **Assignment**

The selected candidate will contribute to the development of a proof of concept obtained at University Côte d'Azur for accessing the content of a metabolomics knowledge graph (KG) with a large language model. It is Python prototype of a metabolomics assistant available as conversational agent that:

1. represents the core experimental information from processed and annotated mass spectrometry data results using the standardized Resource Description Framework (RDF).
2. enables advanced data mining queries using the SPARQL query language.
3. provides a natural language-based interface to perform these queries on the knowledge graph using a large language model (LLM).

### **Main activities**

The topic of natural language access to knowledge graphs is gaining a growing attention both in top international academic conferences and in international industrial conferences. The ICN has a background in knowledge graphs representation and processing for mass spectrometry and metabolomics. The Wimmics team specializes in different AI techniques for knowledge graph providing open-source tools and has a long history in coupling natural language techniques and knowledge graphs.

The R&D programme for that position includes several tasks:

- *Generalize and abstract the bot from specific large language models (LLMs) and specific knowledge graph (KGs):* (1) We will survey and study the impact of using different LLMs on the quality of the results and the potential cost/benefit trade-off in choosing models of different sizes, availability, freshness, etc. (2) We intend to propose and evaluate declarative approaches to loosely-couple the workflow of the bot to the knowledge graph and maximize domain-independence with the goal of incrementally moving from a specific chemistry knowledge graph dedicated prototype to a domain-independent solution.
- *Design a generic and declarative method for tool integration, selection and automated use by the bot:* (1) We propose to identify and implement a library of tools to perform important generic tasks on the knowledge graph including: name-entity recognition against a specific knowledge graph, knowledge extraction from schemata and graph data for context/prompt generation, query solving, etc. (2) We will compare the approaches for integrating a library of tools and their calls to interactions with an LLM in the context of an application to the task of question-answering (Q&A) on a knowledge graph.
- *Go beyond the actual textual interface to support richer interactions:* (1) We propose to study the realization of more complex tasks than textual answers generation to include the generation of graphical widgets and data visualization means when appropriate to an answer. (2) We will consider the possibility to generalize our approach by considering this as a special family of tools for the bot for its task when it comes to express results.
- *Support longer dialogical interactions:* (1) We will investigate the different alternatives for encoding the context of the users' questions in terms of background knowledge, previous interactions, available tools, etc. (2) We intend to leverage the increasing context and prompt size to design a chaining mechanism that supports dialogs and follow-up queries.

A longer-term perspective of the project is to consider other tasks than the support to accessing the content of graph. Methodologically, we imagine extending the previous steps to consider tasks such as contributing, maintaining, validating or semantically enriching a knowledge graph.

## Skills

- Technical skills: knowledge graphs, semantics web, Linked Data, LLMs, ChatBot
- Programming: Python, RDF, SPARQL.
- Relational skills: Ability to work in an interdisciplinary and international network of collaborators
- Other valued appreciated: autonomous, proactive, focused on the research program, deliver on time
- Languages: English

## Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

## Remuneration

Gross Salary: 2788 € per month

## General Information

- **Theme/Domain :** Data and Knowledge Representation and Processing Information system (BAP E)
- **Town/city :** Nice
- **Inria Center :** [Centre Inria d'Université Côte d'Azur](#)
- **Starting date :** 2024-09-02
- **Duration of contract :** 2 years
- **Deadline to apply :** 2024-09-07

## Contacts

- **Inria Team :** [WIMMICS](#)
- **Recruiter :**  
Gandon Fabien / [fabien.gandon@inria.fr](mailto:fabien.gandon@inria.fr)

## About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more

than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

## The keys to success

We are looking for candidates who have (i) a solid engineering experience in AI or bioinformatics or (ii) a PhD in computational biology / computational metabolomics or proteomics.

The candidate should also be autonomous in conducting software development, compiling surveys, conducting experiments and planning the work.

Essential qualities in order to fulfil this assignment are feeling at ease in a dynamic international scientific environment.

The candidate must be passionate about innovation, with expertise in knowledge graphs and large language models.

**Warning :** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

## Instruction to apply

### **Defence Security :**

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

### **Recruitment Policy :**

As part of its diversity policy, all Inria positions are accessible to people with disabilities.