Funded by the European Union

Marie Skłodowska-Curie Actions

Green Field Data

Exceptional
36-months Double Degree PhD Scholarship
Position-P
Phd in Computer Science (France)

PhD in Land, Environment, Resources and Health (Italy)

Agricultural AI data integration and management based on LLM

EU Recruiting institutions	Université Toulouse Capitole, IRIT – Institut de Recherche en Informatique de Toulouse, Toulouse, France (18 Months), Supervisor: F. Ravat, R. Tournier, J. Song. Università degli Studi di Padova, TESAF – Department of Land Environment, Agriculture and Forestry, Padova, Italy (18 Months), Supervisor: F. Marinello, A. Cogato.
Keywords	Data Science, ML & AI, Digital Agriculture, Remote Sensing (IoT and IoRT), Sustainability, Impact assessment, IoT data, Data lakes, Large Language Models

Exceptional benefits at a glance

- ***International PhD training excellence (here)***
- ***Renowned supervisors & top-tier labs***
- ***Interdisciplinary & multi sectoral research***
- ***Competitive MSCA salary & allowances***
- ***Global academic & industrial network***
- ***Non-academic secondments***

Salary

	Living Allowance	Mobility Allowance*	Family Allowance**
Gross amount	**EUR 4736**	**EUR 710**	**EUR 495**

Long Term leave allowance (if applicable)

Special needs allowance (if applicable)

*private mobility-related costs (e.g. travel and accommodation costs), not their professional costs under the action

**doctoral candidate has or acquires family obligations during the action duration, i.e. persons linked to him/her by (i) marriage, or (ii) a relationship with equivalent status to a marriage recognised by the legislation of the country or region where this relationship was formalised; or (iii) dependent children who are actually being maintained by the researcher, the family allowance must be paid to him/her as well

GreenFieldData Project at glance

GreenFieldData: "IoRT Data management and analysis for Sustainable Agriculture" is a project funded under the action HORIZON Marie Sklodowska-Curie Action (MSCA) Joint Doctoral Network. **GreenFieldData** will train a new generation of researchers able to tackle digital and green transition

challenges using a human-centric approach to ensure the robustness and relevance of the solutions responding to the specific needs of the EU market in a context of climate change and increasing socio-economic constraints. At a policy level, **GreenFieldData** outcomes will feed in directly to the aims of the HE Strategic Plan 2025-2027, EU Partnership Agriculture of Data and Digital EU Program. **GreenFieldData** proposes a high-level interdisciplinary, inter-sectoral and international (triple 'i') research project and training network on new IoRT (Internet of Robotic Things) based solutions for sustainable agriculture. **GreenFieldData** will mobilize 14 Doctoral Candidates (DCs) enrolled in Double Degree Doctorate programmes with 12 academic main beneficiary partners, across 7 EU countries. Moreover, 21 non-academic associated partners, and 3 academic associated partners will provide support to the DCs. The partners form a high quality network, where Academic partners have previous research collaborations as outlined in a common vision paper. The ambitious project will provide the DCs with a unique toolbox of cutting-edge knowledge, tools and strategies which will boost their employability and benefit the next generation operational workforce (researchers, Digital Technologies (DTs) and agricultural stakeholders). The project results will also benefit EU innovation as the human-centric IoT devices & robotics, and data-based solutions tailored to EU context will enable the agricultural sector to assess and mitigate the impacts of climate change, and define new sustainable low input practices, thus increasing resilience and competitiveness.

PhD Position P – Agricultural AI data integration and management

Context: Agriculture and agronomy generate a wide variety of data (connected equipment, weather, environmental sensors, livestock, crop monitoring, etc.) [1]. Despite a lot of progress, the value of this large quantity of data remains complex to develop, often not well automatized and underexploited. This data suffers from a strong disparate and disseminated nature (heterogeneity). As a consequence, managing, reusing and crossing different datasets is often complex.

Current Approaches: Managing this data currently rests on conventional approaches using databases (mainly RDBMS—Relational Database Management Systems). However, new approaches have emerged in the past few years, notably consisting in storing the data in separate better suited systems (polystores or multistores [2,3]) by combining different data management paradigms (relational, document-oriented noSQL, graphs databases, columns stores, etc.). Moreover, to favor democratization of approaches using artificial intelligence based on Machine Learning (statistical AI), data lakes have also been proposed [4,5].

Objectives: Monitoring farming and agronomical activities is based on very different types of multi-scale and multi-modal information (weather, soil conditions, crop growth, farming operations, etc.) and is critical in farm management. This information holds a potentially high value exploitable for supporting farming decisions. The problem is further complicated by the wide variety not only of farming conditions (in terms of crops, soils, mechanization, weather...) but also in terms of priorities for the end user: the farmer might target the maximum yield or profit, or minimum environmental impact, or product quality. Such variability not only increase the complexity of possible scenarios but also increases the uncertainty on post processing conditions, constrains and expected results.

This research will produce a systematic classification of different alternative scenarios and will investigate best approaches allowing an efficient and effective combination and integration of heterogenous data. It will require investigating how to define a common reference for all information sources and to provide the relevant analysis of all the available variables in different scenarios conditions. In order to reach this goal, deep learning-based algorithms will be implemented allowing to recognize, summarize, translate, predict,

and generate contents using medium to large or very large datasets: in particular new approaches based on Large Language Models (LLM) will require an adapted access to agricultural data with, for example, adapted data querying operators. Such activity will allow the doctoral candidate to develop an approach which can identify efficiently data similarities, synthesize the most crucial variable combinations, and provide them as a key output to the farmer.

Objectives can be listed as follows:

1. Investigate an approach using a common reference typical of polystore for all information sources available and provide the relevance analysis of the available variables in different farming scenarios conditions provided by the partner (**Tasks 1.3, 2.1**).
2. Explore potential use of AI algorithms (such as hybrid operators using Large Language Models—LLM—or specifically adapted data such as distance precomputing or even vector data managements) algorithms to synthesize the most crucial combination of variables and provide them as a key output to the user (the farmer) (**Task 2.1, 4.2**).
3. Provide an evaluation framework for identifying the relevant combinations of variables for farming management (**Task 4.1**).

Work plan:
The objectives will be achieved using the following work planning to grant their feasibility:

1. Conduct a literature review on data management (multi-store), AI algorithms (such as machine learning or clustering).
2. Familiarize with the multi-level data and how to model them in a polystore architecture [2] or similar (Month 1 – 12).
3. Develop an environment to test the new operators and integrate the available data [Month 13 – 27]; Validate experiments and develop the data analysis process for extracting the combination of crucial variables representing the key output to the user [Month 28– 36].

Expected Results

1. Systematic classification and characterization of data available in farming practice.
2. With the possible help of a master student, a polystore multi-model approach for integrating, combining and representing data from heterogeneous sources (satellite imagery, plant monitoring, management…).
3. New algebraic operators adapted to AI requirements needed for the development of decision support systems.
4. Testing in a real farming environment.

References

[1] Ahmed Kayad, Marco Sozzi, Dimitrios S. Paraforos, Francelino A. Rodrigues, Yafit Cohen, Spyros Fountas, Medel-Jimenez Francisco, Andrea Pezzuolo, Stefano Grigolato, Francesco Marinello. How many gigabytes per hectare are available in the digital agriculture era? A digitization footprint estimation. Comput. Electron. Agric. 198: 107080 (2022).

[2] Zhenzhen Gu, Francesco Corcoglioniti, Davide Lanti, Alessandro Mosca, Guohui Xiao, Jing Xiong, Diego Calvanese: A systematic overview of data federation systems. Semantic Web 15(1): 107-165 (2024)

[3] Rana Alotaibi, Bogdan Cautis, Alin Deutsch, Moustafa Latrache, Ioana Manolescu, Yifei Yang: ESTOCADA: Towards Scalable Polystore Systems. Proc. VLDB Endow. 13(12): 2949-2952 (2020)

[4] Franck Ravat, Yan Zhao. Data Lakes: Trends and Perspectives. DEXA (1) 2019: 304-313, (2019).

[5] Yan Zhao, Imen Megdiche, Franck Ravat, Vincent-nam Dang. A Zone-Based Data Lake Architecture for IoT, Small and Big Data. IDEAS 2021: 94-102 (2021).

PRACTICAL INFORMATION

Recruiting and host institutions	• Université Toulouse Capitole, IRIT – Institut de Recherche en Informatique de Toulouse, Toulouse, France (18 Months) (Recruiting institution) • Università degli Studi di Padova, TESAF – Department of Land, Environment, Agriculture and Forestry, Padova, Italy (18 Months)
Doctoral schools	• MITT @ University of Toulouse, France • Doctoral School Land, Environment, Resources and Health of the University of Padova, Padua, Italy
Supervisors	• F. Ravat, R. Tournier (Université Toulouse Capitole, Toulouse, France) • F. Marinello (Università degli Studi di Padova, Padova, Italy)
Non-academic mentors	• G. Didonna (UpToEarth, Italy) • A. Hassan (CGI, France)
Secondments (1 to 6 hosting months)	• UpToEarth Italia S.r.l., Italy, 6 months, Within the farming environment, experiment on the newly developed solutions with Uptoearth's network. • CGI, France, 3 months, Work on big data and AI technologies architectures as complementary solutions with the multi-model approach.
Contact information	• franck.ravat@irit.fr, ronan.tournier@irit.fr • francesco.marinello@unipd.it; alessia.cogato@unipd.it

RECRUITMENT CRITERIA

General criteria

- MSCA Mobility Rule: researchers must not have resided or carried out their main activity (work, studies, etc.) in **France** for more than 12 months in the 36 months immediately before their date of recruitment
- All researchers recruited in a DN must be doctoral candidates (i.e. not already in possession of a doctoral degree at the date of the recruitment)
- An applicant must have received the equivalent of 300 ECTS with a major in computer science, from which at least 60 ECTS corresponds to a master degree. The master degree must be granted by a university recognized by the International Association of Universities.
- Scientific excellence to fit the PhD project
- Fluent (oral and written) English skills as the project operates in English language
- Knowledge of the language of the host country may be considered a merit
- Team-mindedness

Required skills

- Advanced databases and programming skills
- Modern data stores architectures and querying is a plus
- Modern noSQL database systems is a plus
- Machine Learning is a plus
- Interdisciplinary work
- Master degree (or equivalent)

APPLICATION

How to apply?

- **All information are provided here**

Deadline: **15th April 2026**

Other information

Université de Toulouse

The **Université de Toulouse (Univ Toulouse)** is a large, public research university system based in Toulouse, France, and one of the oldest in Europe, originally founded in **1229**.

- **Structure:** While historically a single entity, the modern University of Toulouse is a collective entity (a *Communauté d'universités et établissements* or *ComUE*) that federates several autonomous universities and institutions of higher education in the Toulouse region, including:
  - **Toulouse 1 Capitole University** (Law, Economics, Management), which will administratively host the PhD**,**
  - **University of Toulouse - Jean Jaurès** (Arts, Literature, Languages, Humanities, Social Sciences),
  - **Toulouse III - Paul Sabatier University** (Sciences, Health, Sport).
- **Scale and Focus:** It is a major university area in France, educating over **100,000 students** across a wide spectrum of disciplines, including aerospace, computer science, health, and social sciences.

IRIT: Toulouse Research Institute in Information Technology

The **Institut de Recherche en Informatique de Toulouse (IRIT)** is a highly influential **Joint Research Unit (UMR 5505)** in France, dedicated to **Information Technology (IT)** and **Computer Science**.

- **Nature:** It is one of the largest and most significant research laboratories in the field in the Occitanie region and nationally.
- **Supervision (Multi-Tutelle):** It operates under the joint supervision of several major institutions, including the **CNRS** (French National Center for Scientific Research) and the various Universities of Toulouse (Toulouse INP, UT Capitole, UT Jean Jaurès, and UT Paul Sabatier).
- **Focus Areas:** IRIT's research covers a broad range of core IT topics, structured around key scientific themes and strategic application domains.
  - **Scientific Topics** include the design and construction of systems, numerical modeling of the real world, concepts for cognition and interaction, autonomous adaptive systems, and transforming raw data into intelligible information.
  - **Strategic Applications** include Health, Autonomous Systems, and Well-being; Smart Cities; Aeronautics, Space, and Transportation; and Digital Social Ecosystems (Social Media).

- **Size:** It comprises a large team of over six hundred researchers, administrative staff, and PhD students.

University of Padova (UNIPD)

The University of Padova is one of the **oldest and most prestigious universities in Europe**, founded in **1222**.

- **History and Prestige:** It is the second-oldest university in Italy and the fifth-oldest surviving university in the world, renowned for its motto, *Universa Universis Patavina Libertas* (Paduan Freedom is Universal for Everyone), emphasizing academic freedom. Key historical figures like **Galileo Galilei** and **Nicolaus Copernicus** were associated with the university, and it houses the world's oldest anatomical theatre and the oldest academic Botanical Garden (a UNESCO World Heritage Site).
- **Size and Scope:** Located in Padua, Italy, it is a large public research university with around 60,000-70,000 students, offering a wide range of multidisciplinary programs across **eight schools** and **32 departments**.
- **Excellence:** UNIPD consistently ranks among the best Italian universities, particularly excelling in **teaching quality** and **research**.

Department of Land, Environment, Agriculture and Forestry (TESAF)

The Department of Land, Environment, Agriculture and Forestry (TESAF - *Dipartimento Territorio e Sistemi Agro-Forestali*) is a key part of the University of Padova's School of Agricultural Science and Veterinary Medicine.

- **Focus:** TESAF is a **multidisciplinary structure** dedicated to the **understanding, effective management, and sustainable use of renewable agricultural and forest resources**.
- **Scope:** Its research and teaching activities cover a broad range of disciplines, including:
    - Forestry (Silviculture and Forest Engineering)
    - Agricultural and Forest Economics and Policy
    - Land and Environmental Appraisal and Planning
    - Hydrology and Water Resources Management
    - Phytopathology
    - Remote Sensing and GIS (Geographic Information Systems)
- **Location and Collaboration:** TESAF is primarily located on the **Agripolis Campus** in Legnaro (PD), a specialized center for agricultural and veterinary sciences, where it collaborates closely with other departments in the sector.
- **International Presence:** The department is highly active internationally, involved in numerous European and international research projects and hosting several international Master's and PhD programs, often with a focus on sustainable forest and land management in Europe and the Mediterranean.