Data Platform for Agritech

1 Project proposal

Digital transformation is one of the most disruptive trends of recent years, and the synergy of information and communications technologies and precision agriculture is fostering new potential in the agritech domain. Indeed, data-driven decision support systems are now taking on a major role due to the spread of the internet of (robotic) things devices coupled with artificial intelligence and machine learning, which make it possible to precisely act in every aspect of the life cycle of agricultural assets. In this context, data platforms answer the issues related to data collection, processing, exploitation, and management.

The goal of the project is to engineer a domain-level data platform to collect, integrate, and manage every data relevant to the Agritech Spoke 3 [1]. Domain level refers to high-level agritech problems or areas of analysis that should be addressed and supported by the proposed system. Indeed, the data platform must collect and integrate data, and support the analytic workloads of the six partners of the Agritech Spoke 3.

2 Project activities

Previous activities. In previous research activities [2, 3, 4], techniques for orchestrating and profiling data collections within a data platform for precision agriculture were presented [5] and an approach to enable queries in a polyglot context was introduced [6] with the idea of exploiting the expressiveness and heterogeneity of data patterns as an added value, both in querying and interpreting the analysis results.

Proposed activities. The candidate, as part of the work carried out by the research group, will be responsible for:

- Study the state-of-the-art data platforms for precision agriculture to identify their different architectures and technological stacks.
- Acquire and store data in *stream* and *batch* modes from the partners' heterogeneous sensors and sources (e.g., satellite images and IoT sensor data).

- Define and implement a unifying conceptual schema of the data layer to integrate and standardize the different data sources.
- Identify, deploy, and orchestrate the optimal set of services necessary to support all the case studies from the partners of the Agritech Spoke 3.
- Study and apply the Lakehouse paradigm to the platform to ease the storage and analysis of heterogeneous data.

References

[1] Agritech. `https://agritechcenter.it/`. accessed 2023-10-31.

[2] WeLaser, 2020. H2020 "Sustainable weed management in agriculture with laser-based autonomous tool" — n. Grant Agreement 101000256 e n. Cup J32F20001250006. `https://welaser-project.eu/`.

[3] Agro.Big.Data.Science. `http://agrobigdatascience.it/`.

[4] MoReFarming. `http://www.morefarming.it/`.

[5] Matteo Francia, Enrico Gallinucci, Matteo Golfarelli, Anna Giulia Leoni, Stefano Rizzi, and Nicola Santolini. Making data platforms smarter with MOSES. *Future Gener. Comput. Syst.*, 125:299–313, 2021.

[6] Chiara Forresi, Matteo Francia, Enrico Gallinucci, and Matteo Golfarelli. Optimizing execution plans in a multistore. In *European Conference on Advances in Databases and Information Systems*, pages 136–151. Springer, 2021.