Technological and Research Challenges in Data Engineering for Sustainable Agriculture

Sandro Bimonte, Gianni Bellocchi, Francois Pinet, Guillaume Charrier, Dimitris Sacharidis, Mahmoud Sakr, Ronan Tournier, Gentian Jakllari, Gerard Chalhoub, Tahar Kechadi, et al.

Check for updates

Technological and Research Challenges in Data Engineering for Sustainable Agriculture

Sandro Bimonte
Gianni Bellocchi
François Pinet
Guillaume Charrier
INRAE
Clermont-Ferrand, France

Dimitris Sacharidis
Mahmoud Sakr
Université Libre de Bruxelles
Brussels, Belgium

Ronan Tournier
Gentian Jakllari
IRIT, Université Toulouse Capitole
Toulouse, France

Gerard Chalhoub
Université Clermont Auvergne
Clermont-Ferrand, France

Tahar Kechadi
University College Dublin
Dublin, Ireland

Boualem Benatallah
Dublin City University
Dublin, Ireland

Francesco Marinello
University of Padova
Padova, Italy

Roberto Oberti
Università degli Studi di Milano
Milan, Italy

Jérôme Bindelle
University of Liège
Liège, Belgium

Ginta Majore
Vidzeme University of Applied Sciences
Valmiera, Latvia

Piotr Skrzypczyński
Poznań University of Technology
Poznań, Poland
piotr.skrzypczynski@put.poznan.pl

ABSTRACT

This paper presents a concise exploration of the evolving landscape of sustainable agriculture through the lens of data engineering. The aim is to highlight the most important challenges in the field and sketch ideas for how to address them. This position paper delves into key research challenges such as multi-modal data integration, data quality assurance, network optimisation and edge versus fog data processing strategies. Additionally, it emphasises the significance of performance enhancement in driving innovation within sustainable agriculture. By addressing these challenges and following the proposed visionary approaches for future research endeavours, we claim that data engineering will serve as a catalyst for advancing sustainable agriculture practices.

CCS CONCEPTS

• **Information systems** → **Mediators and data integration**; **Information integration**;

ACM Reference Format:
Sandro Bimonte, Gianni Bellocchi, François Pinet, Guillaume Charrier, Dimitris Sacharidis, Mahmoud Sakr, Ronan Tournier, Gentian Jakllari, Gerard Chalhoub, Tahar Kechadi, Boualem Benatallah, Francesco Marinello, Roberto Oberti, Jérôme Bindelle, Ginta Majore, and Piotr Skrzypczyński. 2024. Technological and Research Challenges in Data Engineering for Sustainable Agriculture. In *International Workshop on Big Data in Emergent Distributed Environments (BiDEDE '24), June 09–15, 2024, Santiago, AA, Chile.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3663741.3664786

BiDEDE '24, June 09–15, 2024, Santiago, AA, Chile

ACM ISBN 979-8-4007-0679-0/24/06
https://doi.org/10.1145/3663741.3664786

1 INTRODUCTION

According to the definition provided by the United Nations Food and Agriculture Organization (FAO) [14], *sustainable agriculture* entails managing and conserving natural resources and directing technological change towards ensuring the continued satisfaction of human needs for the present and future generations. This concept leverages principles of environmental non-degradation, technical appropriateness, economic viability and social acceptability.

Disruptive digital technologies such as the Internet of Things (IoT), robotics, Big Data analytics, and Artificial Intelligence (AI) has fostered transformations across various sectors, including agriculture. This convergence has given rise to the concept of smart farming or Agriculture 4.0 [29], which leverages IoT devices and data-driven approaches to optimise operational efficiency while balancing economic, environmental and social imperatives. Robots play a pivotal role in smart farming. Capable of executing repetitive and precise farming tasks over extended time periods, they offer the advantage of low environmental impact and can operate effectively within a fleet. Equipped with specialised tools and coupled with advanced data acquisition and processing technologies, robots demonstrate efficiency in several agricultural tasks such as spot weeding, variable rate seeding or harvesting [56].

Motivated by the transformative potential of AI, IoT, robotics and data-driven approaches in agriculture [37], we propose the concept of the Internet of Robotics Things (IoRT) Data for Sustainable Agriculture (Fig. 1), which embarks on a mission to explore flexible

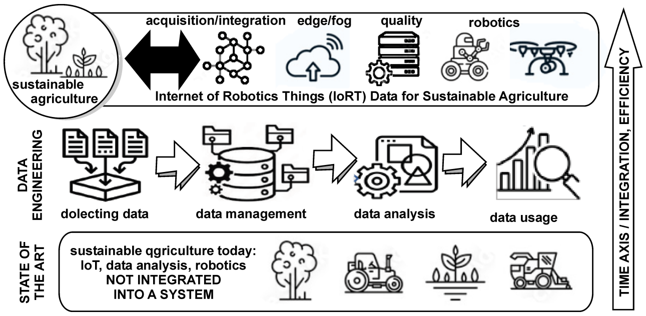

Figure 1: Concept of using data engineering to transform the field of sustainable agriculture into an efficient, data-driven, distributed system integrating sensors and robots.

data-driven solutions in the sustainable agriculture domain. However, the transition towards smart, sustainable agriculture presents unique hurdles, particularly within data engineering [54], like the integration of diverse data modalities, assurance of data quality, optimisation of network infrastructure, delineation between edge and fog data processing paradigms, and enhancement of overall system performance. While some of these challenges can be addressed by a data management reference architecture for sustainable agriculture proposed in [15], our work focuses on a vision for the transition that can fully explore the AI, IoT and robotics potential.

Therefore, we briefly survey the key IoRT-based methods and technologies, aiming to establish a coherent, effective and flexible framework rooted in IoRT data. The methodological and technological research endeavours are structured around three primary dimensions: (i) the Internet of Robotic Things, focusing on the integration and optimisation of data yielded by Iot and robotic systems for agricultural tasks; (ii) data engineering, aimed at developing robust data management, integration and analysis techniques tailored to the unique challenges of agricultural data, and (iii) the sustainability dimensions of agricultural systems encompassing agronomic, environmental and ecological aspects. Recognising the interrelated nature of the data value chain steps, our approach considers innovations across data collection, management, analysis and interaction. Moreover, by acknowledging the diverse array of agricultural practices, the research on IoRT data extends to various domains, fostering synergy and knowledge exchange through generic agricultural research methodologies and innovations.

The IoRT data for sustainable agriculture concept has the potential to yield immediate impacts on agricultural productivity, resource conservation, and global food security. By leveraging real-time data from IoT devices and robotics, farmers can make more informed decisions, leading to optimised crop yields and enhanced productivity. Using IoRT data can help in the efficient management of resources such as water and energy, contributing to their conservation and sustainability. In the long term, the IoRT data for sustainable agriculture concept holds promise in promoting resilience in agricultural systems. By providing insights into ecosystem dynamics and climate variability, IoRT data can enable farmers to adapt to changing environmental conditions and mitigate the impact of extreme weather events and environmental degradation.

2 DATA ENGINEERING FOR SUSTAINABLE AGRICULTURE

Data engineering is commonly applied in sustainable agriculture, facilitating the management, analysis, and interpretation of a diverse array of data generated by field devices and sensors [44]. Agricultural IoRT applications require the use of intricate spatiotemporal data, encompassing robot trajectories and time-stamped meteorological data; the use of streaming data, including sensor data from field deployments and multimedia data such as videos and images, and the integration of historical data covering all facets of an IoRT system [10]. Data engineering for sustainable agriculture introduces unique requirements.

- It relies on autonomous robots and vehicles operating within dynamic and uncontrolled environments. Handling fast-arriving data streams from field devices, and particularly from robots [59], requires real-time analysis capabilities to facilitate timely decision-making and event handling. Techniques must be devised to address unexpected events such as obstacles, weather changes, or equipment failures effectively.
- It involves tightly constrained computational and communication resources, often deployed in rural areas. Efficiently storing large volumes of multi-modal data necessitates the development of robust data ingestion and storage architectures, e.g. in the form of a data lake architecture [39]. Moreover, data processing pipelines have to strike the right balance in the edge-fog continuum.
- It deals with terrain and crop Variability. Research on moving objects has primarily focused on two types of movement: constrained movement in networks, such as vehicles on roads or trains on tracks, and free-space movement, such as animals. However, agricultural robots operate within constantly changing crop layouts and diverse weather conditions. Planning routes must account for these factors to ensure smooth navigation without causing harm to crops or encountering obstacles. Although the movement is constrained within a network, this network has a dynamic nature, which presents a unique challenge, for instance to trajectory similarity search.
- It involves stakeholders with diverse profiles, including equipment providers and farmers lacking proficiency in information technologies.

The existing data engineering paradigms for meeting the challenges of data-intensive agricultural applications encompass various approaches such as (i) distributed relational databases RDBMS [4, 5, 63], (ii) NoSQL systems [23, 24, 48], and (iii) data flow processing systems like those using Spark [3, 62], (iv) Data lakes [45, 49] and lakehouses [60]. These approaches compare to each other in their advantages and disadvantages [32, 33, 50]. While RDBMS offer a diverse range of features and methods, they encounter challenges in scaling out effectively. NoSQL systems exhibit excellent scalability but are still evolving and may lack sophistication in handling mobility data, often featuring limited query capabilities and basic partitioning and indexing. Systems based on the Spark, while scalable and offering a rich array of queries, may rely on ad-hoc solutions for indexing and querying. Data lakes and lakehouses emerge as promising solutions, offering metadata management, data governance, and scalable storage capabilities. However, challenges persist in representing metadata effectively and providing efficient exploration of the data catalogue. Modelling a data lake to store and retrieve heterogeneous field-level data efficiently presents a significant research challenge. In the context of performance, benchmarks proposed for distributed query processing in dynamic smart city scenarios [57] might be applied for sustainable agriculture scenarios as well. Addressing the multifaceted challenges in data engineering for sustainable agriculture demands innovative approaches to data ingestion, storage, analysis, and retrieval. Implementing the IoRT-based concept requires identifying and addressing research challenges in areas such as multi-modal data integration, ensuring data quality, optimising wireless communication protocols, and harnessing the power of edge computing for real-time decision-making.

3 RESEARCH CHALLENGES

Based on the characteristics and requirements that data engineering for sustainable agriculture introduces, we next detail specific research challenges and sketch proposed solutions or directions worth exploring.

3.1 Data Integration

Multi-modal data refers to information gathered from various sources or sensors deployed in the considered scenario. Unlike other distributed systems where data integration takes place, IoRT systems in agriculture are characterised by large diversity of sensing modalities and data sources, encompassing a range of technologies, such as drones, satellite imaging, ground sensors, weather stations, and robotic agents. Each of the modalities provides distinct types of data, including soil moisture levels, crop health indicators, weather patterns, pest infestations, and more. By integrating data from these diverse sources, farmers and agricultural experts can gain a more comprehensive and nuanced understanding of crop conditions, environmental factors, and overall farm management.

In contrast, heterogeneous data in the context of sustainable agriculture involves the diverse nature of agricultural data itself. This diversity can manifest in terms of data formats, structures, scales, and characteristics. For instance, data may vary in formats such as text, images, geospatial information, and time series data. Additionally, data may differ in resolution, accuracy, frequency of collection, and spatial coverage. This heterogeneity poses challenges in aggregating, processing, and analyzing the data effectively for decision-making in agricultural practices. Integrating multi-modal and heterogeneous data in the domain of IoRT for sustainable agriculture presents several challenges:

- Data fusion: combining data streams from various sensors and sources while ensuring consistency and preserving the unique characteristics of each data type.
- Data alignment: resolving disparities in data formats, coordinate systems, temporal resolutions, and spatial resolutions to facilitate meaningful integration and analysis.
- Semantic interoperability: establishing common vocabularies, ontologies, and metadata standards to enable seamless communication and interoperability between different agricultural data types and systems.

Integrating multi-modal data in sustainable agriculture involves employing advanced data fusion techniques to merge information from various sensors and sources while preserving the unique characteristics of each data type [22]. Sensor data fusion, for instance, facilitates the unified representation of agricultural phenomena by combining data streams from drones, satellites, ground sensors, and robotic agents [6]. Machine learning techniques can further enhance this process by autonomously extracting meaningful patterns and relationships from multi-modal data [30], enabling more accurate and efficient integration while incorporating the insights and expertise of domain experts and stakeholders.

Addressing the heterogeneity of agricultural data requires the implementation of data alignment strategies to resolve disparities in formats, coordinate systems, temporal resolutions, and spatial resolutions. Techniques such as data normalisation, interpolation, and georeferencing are employed to align disparate datasets, enabling coherent analysis and decision-making in agricultural practices. Additionally, integrating data from state-of-the-art robotic platforms and sensors, modern localisation techniques [34], advanced task and path planning approaches [9], and the potential of AI for autonomous mapping in agriculture [11] further enhances data alignment efforts. According to the findings of a recent survey [55], many examples of agricultural mobile robots provide full navigation and autonomous mapping capabilities, showcasing significant advancements in this field. By ensuring data consistency and compatibility through these strategies, the effectiveness of data integration efforts is heightened, particularly in managing the heterogeneous nature of agricultural data.

Achieving semantic interoperability is crucial for integrating heterogeneous data in the domain of sustainable agriculture. Establishing common vocabularies, ontologies [16], and metadata standards [12, 53] facilitates seamless communication and interoperability between different agricultural data types and systems [20]. By enabling efficient data exchange, integration, and interoperability across diverse agricultural applications and stakeholders, semantic interoperability solutions streamline data management processes [27] and support informed decision-making in agricultural operations, as well as in handling the various data modalities involved.

3.2 Data Quality

Data quality is defined as the adequacy of the data with certain criteria [18]. It is the conformity of the data with the reality that people want to model. A quality level is a set of requirements allowing a correct and safe use of the data. Quality is not an absolute concept. Data may be of quality for one application but not for another.

In the agriculture domain, the available data from IoT sensors, connected autonomous equipment or external digital services (such as weather) has drastically increased in availability and quantity. Integrating or harmonising such large volumes of data inevitably leads to quality issues, whether it be, for example, a badly calibrated sensor or a service provider that updates its data format. Having poor quality data can lead to a bad decision, for example, over-irrigating crops or applying unnecessary treatments, which produce environmental impacts.

Data quality is a multifaceted problem, and each issue has to be tackled specifically. Based on previous work, the authors of [46] identify different components of data quality:

- Genealogy and traceability of data, which provide information on their origin and the methods to produce them.
- Accuracy, which assesses how closely the data conforms to the real world in terms of precision.
- Completeness, which describes whether the objects in the dataset represent all objects we want to model.
- Logical consistency, which requires the absence of contradiction in the data.
- The consistency of the data with their domain of values (ranges or types).

Data quality issues have to be detected. There are usually several techniques to manage data quality, such as:

- Reject data that does not meet the required level of quality. For example, integrity constraints can be modelled and controlled in order to prohibit the insertion of new data which is not consistent with a database [19].
- Correct data values when possible, for example, by cross-checking.
- Interpolate or extrapolate missing data.
- Model data uncertainty, and take it into consideration in processing, for example, using probability theory or fuzzy sets [31].

At the conceptual level, uncertainty can also be modelled thanks to specific data model methods [40].

Detecting low quality data is an important task that requires complex processing. Modern techniques, such as machine learning algorithms, may detect outliers or anomalies much faster on big data sets, but the probabilistic output of such algorithms will have to be taken into account. In agriculture, training datasets must also be further developed. To ensure machine learning models are reliable, special attention has to be paid to ensure they are trained on good quality data, they remain explainable [21], and do not cause unwarranted harm [47]. Moreover, farm data are spatiotemporal by nature, which leads to specific problems, such as the difficulty of managing changing farm plots in the event of their splitting or merging [2]. The aggregation of farm data from plots to territories can also lead to the fusion of contradictory data, which must be addressed. The confidentiality of certain farm data may also make it necessary to use effective anonymization or aggregation techniques. These methods must guarantee confidentiality even when the data is cross-referenced with external information (for example, found on the Web). The use of such techniques will increase the confidence of data producers [25].

With the ever-growing volume of data in agriculture, upon a certain volume, statistical tools will help mitigate the impact of data quality issues. However, not all systems will be able to reach the threshold. The latter will require numerous specific mitigation approaches, robust algorithms, and more resilient approaches.

3.3 IoT and Wireless Communications

The use of IoRT has enabled the agriculture domain to switch to precision agriculture and develop new and sustainable techniques [58]. This transition is not without challenges from the technological point of view [42]. On one hand, there are many wireless communication technologies available on the market, such as WiFi, ZigBee, LoRa, Bluetooth, and Cellular. Each of these technologies answers specific performance and deployment requirements in terms of coverage, data rate, reliability and latency. The choice of one technology over the other involves a trade-off in some aspects. For example, using LoRa technology can solve the issue of coverage and lifetime of the network, but at the same time, LoRa can support very limited data rates and can only be used for transmitting a few bytes per second, depending on the deployment strategy [17]. On the other hand, once a specific technology is chosen, many aspects should be taken into consideration in order to optimise its performance. For instance, technologies such as Wi-Fi or ZigBee use a probabilistic medium access control method, which cannot guarantee a specific level of performance in terms of throughput or latency. Such technologies are easy to use and deploy but suffer severe performance degradation under high congestion [35].

One main challenge when optimising the performance of wireless communications is the ability to adapt to changes in the deployment environment [52]. These changes can arise due to mobility of the communicating nodes, such is the case for connected robots. When a robot is close to the network infrastructure, the access point is Wi-Fi, the base station is cellular networks, and the wireless link is able to support high data rates. But when the robot starts moving away from the network infrastructure, the quality of the wireless link deteriorates, and performances decrease. In such dynamic scenarios, AI algorithms come in handy [38]. Using machine learning techniques can help predict the behaviour of the wireless link. This prediction can be very useful for adapting the application behaviour in order to avoid performance degradation [51].

In order to learn efficiently, the collected data should describe the state of the network. These data are then stored and analysed using learning methods to optimise the decision-making process. Decisions can be made at different levels of the network protocol stack, from the physical layer to the application layer [61]. Each learning model integrated into the protocol stack will have its own set of input data to analyse and its own actions to take.

Training these models, mainly applying reinforcement learning models, can be very challenging. This is where network simulators such OMNeT++, NS3, or OPTNET can offer the possibility of defining a wide variety of simulated training scenarios without taking the risk of making false decisions in real scenarios [13]. One of the main challenges of this training approach is to make sure that the simulation tool is realistic, i.e., it emulates wireless link behaviour in a realistic manner. Extending the simulated training phase to online real training scenarios by the means offered by Transfer Learning techniques is a promising method in this domain [43].

3.4 Edge vs. Fog Data Processing

The debate between edge and fog data processing is significant within the domain of the IoRT and data management for sustainable agriculture. Edge computing involves processing data locally on the devices themselves, such as robots or sensors, minimising latency and reducing the need for constant data transmission to centralised servers. This approach enhances real-time decision-making capabilities crucial for tasks like precision farming [36].

On the other hand, fog computing extends the edge paradigm by incorporating intermediate nodes between the edge devices and the cloud, allowing for more complex data processing closer to the source while still leveraging cloud resources when necessary [26]. In sustainable agriculture, where efficient resource management is paramount, edge and fog computing play pivotal roles. Edge processing ensures rapid responses to immediate environmental changes [53], optimising tasks like irrigation and pest control, while fog computing facilitates more sophisticated analytics and predictive modelling, enabling long-term planning and optimisation of agricultural practices for sustainability and productivity [41].

From the viewpoint of the IoRT concept, there exist notable research gaps and challenges, particularly concerning the integration of robots, which, unlike typical IoT devices, can be considered autonomous agents [8]. Modern robots often work with the Robot Operating System (ROS), with the current standard being ROS2 [28]. While edge computing enables ROS2 robots to swiftly analyse sensory inputs and execute tasks autonomously with minimal latency, there remain challenges in optimising the coordination and communication between these robots and other edge devices. Additionally, fog computing, which extends computational capabilities across intermediate nodes, including ROS2-enabled robots, faces challenges in achieving seamless collaboration and data exchange among heterogeneous devices in dynamic agricultural environments. Furthermore, ensuring the security and privacy of sensitive agricultural data processed at the edge and fog layers remains a critical concern. Despite these challenges, the potential of integrating ROS-based robots with edge and fog computing holds promise for advancing sustainable agriculture practices. Further research is needed to address these gaps and overcome challenges, ultimately enabling the seamless integration of edge and fog data processing methodologies with robots.

4 EXPECTED IMPACT

The concept of IoRT data for sustainable agriculture can address various agro-ecological challenges. For instance, in mitigating drought impacts on grassland ecosystems, IoT-enabled visual sensors facilitate high-resolution spatial and temporal data collection, enhancing monitoring for effective ecosystem management. Similarly, in combating plant diseases like Esca and Powdery mildew in vineyards, decision support systems using sensor data assist winegrowers in disease control strategies, reducing reliance on plant protection products and minimising environmental impact. Additionally, IoRT enables precision agriculture by integrating data from weather conditions, soil type, and crop management practices, empowering farmers to make informed decisions and mitigate yield-limiting factors such as extreme weather events and soil limitations.

The IoRT data concept anticipates delivering significant advancements in sustainable agriculture by establishing a comprehensive reference architecture for managing IoRT data derived from agroecological processes. This framework streamlines data collection, integration, and analysis [1, 7], offering a versatile toolkit tailored for researchers, farmers, and policymakers. Through IoRT utilisation, stakeholders gain valuable insights into the agronomic, environmental, and ecological benefits of IoRT-based agro-ecological practices, fostering a deeper understanding of their impact on agricultural systems.

Moreover, the IoRT system is poised to generate actionable recommendations for scaling up technological solutions, enabling widespread adoption across diverse agricultural practices and fostering innovation in the sector. Ultimately, IoRT holds promise in revolutionising agro-ecological practices by leveraging advanced technologies to optimise resource utilisation, minimise environmental impact, and enhance agricultural resilience.

5 FINAL REMARKS

This paper underscores the visionary nature of the IoRT concept and its profound relevance in addressing the pressing ecological and economic challenges confronting agricultural systems. It offers insights into pathways towards sustainable and resilient agricultural practices by elucidating the potential applications of IoRT in sustainable agriculture contexts, particularly through data engineering, big data analytics, and edge/fog computing solutions. Embracing the IoRT paradigm, with its emphasis on harnessing data-driven approaches and advanced computing technologies, has the potential to catalyse transformative changes in agricultural management, ultimately leading to a more prosperous and sustainable future.

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency project IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25).

REFERENCES

[1] G André, B Bachelet, P Battistoni, A Belhassena, S Bimonte, C Cariou, F Chabot, G Chalhoub, A Couvent, G Garani, J Laneurit, R Moussa, K Oikonomou, I Sammour, M Sebillo, M Vilela Souza, N Tricot, and R Wrembel. 2023. LambdAgrIoT: a new architecture for agricultural autonomous robots' scheduling: from design to experiments. *Clust. Comput.* 26, 5 (2023), 2993–3015.

[2] Florence Le Ber Aurélie Leborgne, Ezriel Steinberg and Stella Marc-Zwecker. 2024. Representation and Analysis of the Evolution of Agricultural Territories by a Spatio-temporal Graph. In *Geographic Data Imperfection 2: Use Cases*. Wiley.

[3] Mohamed Bakli, Mahmoud Sakr, and Taysir Hassan A. Soliman. 2019. HadoopTrajectory: a Hadoop spatiotemporal data processing extension. *Journal of Geographical Systems* (2019), 1–25.

[4] Mohamed Bakli, Mahmoud Sakr, and Esteban Zimanyi. 2019. Distributed moving object data management in MobilityDB. In *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*. 1–10.

[5] Mohamed Bakli, Mahmoud Sakr, and Esteban Zimányi. 2020. Distributed spatiotemporal trajectory query processing in SQL. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*. 87–98.

[6] V Barrile, S Simonetti, R Citroni, A Fotia, and G Bilotta. 2022. Experimenting Agriculture 4.0 with Sensors: A Data Fusion Approach between Remote Sensing, UAVs and Self-Driving Tractors. *Sensors* 22, 20 (2022).

[7] S Bimonte, H Badir, P Battistoni, H Bazza, A Belhassena, C Cariou, G Chalhoub, J-C Corrales, A Couvent, J Laneurit, R Moussa, J E Plazas, M Sebillo, and N Tricot. 2023. Data-centric UML profile for agroecology applications: Agricultural autonomous robots monitoring case study. *Comput. Sci. Inf. Syst.* 20, 1 (2023), 459–489.

[8] Grazyna Brzykcy, Jacek Martinek, Adam Meissner, and Piotr Skrzypczynski. 2001. Multi-agent blackboard architecture for a mobile robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2369–2374.

[9] Ruth Cordova-Cardenas, Luis Emmi, and Pablo Gonzalez-de Santos. 2023. Enabling Autonomous Navigation on the Farm: A Mission Planner for Agricultural Tasks. *Agriculture* 13, 12 (2023).

[10] Ania Cravero, Sebastián Pardo, Patricio Galeas, Julio López Fenner, and Mónica Caniupán. 2022. Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability* 14, 23 (2022).

[11] Haizhou Ding, Baohua Zhang, Jun Zhou, Yaxuan Yan, Guangzhao Tian, and Baoxing Gu. 2022. Recent developments and applications of simultaneous localization and mapping in agriculture. *J. Field Robotics* 39, 6 (2022), 956–983.

[12] Miguel-Angel Sicilia Elena García-Barriocanal and Salvador Sánchez-Alonso. 2013. Providing semantic metadata to online learning resources on sustainable agriculture and farming: combining values and technical knowledge. *Interactive Learning Environments* 21, 3 (2013), 301–318.

[13] Serap Ergun, Ibrahim Sammour, and Gerard Chalhoub. 2023. A survey on how network simulators serve reinforcement learning in wireless networks. *Computer Networks* 234 (2023), 109934.

[14] Food and Agriculture Organization of the United Nations. 2014. Building a common vision for sustainable food and agriculture. https://www.fao.org/cofi/46037-0799fded181eabdcf681755783a3601b.pdf

[15] Görkem Giray and Cagatay Catal. 2021. Design of a Data Management Reference Architecture for Sustainable Agriculture. *Sustainability* 13, 13 (2021).

[16] Anat Goldstein, Lior Fink, and Gilad Ravid. 2021. A Framework for Evaluating Agricultural Ontologies. *Sustainability* 13, 11 (2021).

[17] Gunjan Gupta, Robert Van Zyl, and Vipin Balyan. 2022. Evaluation of LoRa nodes for long-range communication. *Nonlinear Engineering* 11, 1 (2022), 615–619.

[18] Clément Iphar, Benjamin Coste, Aldo Napoli, Cyril Ray, and Rodolphe Devillers. 2019. Integrity and trust of geographic information. In *Geographic Data Imperfection 1: From Theory to Applications.*, Eric Desjardin et François Pinet Mireille Batton-Hubert (Ed.). Wiley/ISTE Editions, Chapter 4, 28 p.

[19] Sandro Bimonte Kamal Boulil and Francois Pinet. 2014. Spatial OLAP integrity constraints: From UML-based specification to automatic implementation: Application to energetic data in agriculture. *Journal of Decision Systems* 23, 4 (2014).

[20] Andreas Kamilaris, Feng Gao, Francesc X. Prenafeta-Boldu, and Muhammad Intizar Ali. 2016. Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications. In *IEEE 3rd World Forum on Internet of Things (WF-IoT)*. 442–447.

[21] Loukas Kavouras, Konstantinos Tsopelas, Giorgos Giannopoulos, Dimitris Sacharidis, Eleni Psaroudaki, Nikolaos Theologitis, Dimitrios Rontogiannis, Dimitris Fotakis, and Ioannis Z. Emiris. 2023. Fairness Aware Counterfactuals for Subgroups. In *NeurIPS*.

[22] Timo Korthals, Mikkel Kragh, Peter Christiansen, Henrik Karstoft, Rasmus N. Jørgensen, and Ulrich Rückert. 2018. Multi-Modal Detection and Mapping of Static and Dynamic Obstacles in Agriculture for Process Evaluation. *Frontiers in Robotics and AI* 5 (2018).

[23] Nikolaos Koutroumanis and Christos Doulkeridis. 2021. Scalable Spatio-temporal Indexing and Querying over a Document-oriented NoSQL Store. In *International Conference on Extending Database Technology*.

[24] Nikolaos Koutroumanis, Panagiotis Nikitopoulos, Akrivi Vlachou, and Christos Doulkeridis. 2019. NoDA: Unified NoSQL Data Access Operators for Mobility Data. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases* (Vienna, Austria) *(SSTD '19)*. 174–177.

[25] B. Lauga, B. Balvay, L. Topart, J. Leclaire, A. Clenet, F. Brun, F. Pinet, C. Roussey, and M. Sine. 2019. MULTIPASS: Managing the consents of access to farm data in a chain of trust to make new services emerge for farmers. In *Conference of the European Federation for Information Technology in Agriculture, Food and the Environment*.

[26] Yi-Bing Lin, Whai-En Chen, and Ted C.-Y. Chang. 2023. Moving from Cloud to Fog/Edge: The Smart Agriculture Experience. *IEEE Communications Magazine* 61, 12 (2023), 86–92.

[27] Djakhdjakha Lynda, Farou Brahim, Seridi Hamid, and Cissé Hamadoun. 2023. Towards a semantic structure for classifying IoT agriculture sensor datasets : An approach based on machine learning and web semantic technologies. *Journal of King Saud University - Computer and Information Sciences* 35, 8 (2023), 101700.

[28] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. 2022. Robot Operating System 2: Design, architecture, and uses in the wild. *Science Robotics* 7, 66 (May 2022).

[29] Federico Maffezzoli, Marco Ardolino, Andrea Bacchetti, Marco Perona, and Filippo Renga. 2022. Agriculture 4.0: A systematic literature review on the paradigm, technologies and benefits. *Futures* 142 (2022), 102998.

[30] Md. Suruj Mia, Ryoya Tanabe, Luthfan Nur Habibi, Naoyuki Hashimoto, Koki Homma, Masayasu Maki, Tsutomu Matsui, and Takashi S. T. Tanaka. 2023. Multimodal Deep Learning for Rice Yield Prediction Using UAV-Based Multispectral Imagery and Weather Data. *Remote Sensing* 15, 10 (2023).

[31] Francois Pinet Mireille Batton-Hubert and Andre Miralles. 2024. Implementation and computation of fuzzy geographic objects in agriculture. In *Geographic Data Imperfection 2: Use Cases*. Wiley.

[32] Mohamed Mokbel, Mahmoud Sakr, Li Xiong, Andreas Züfle, Jussara Almeida, Taylor Anderson, Walid Aref, Gennady Andrienko, Natalia Andrienko, Yang Cao, et al. 2022. Mobility data science (dagstuhl seminar 22021). In *Dagstuhl reports*, Vol. 12. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[33] Mohamed Mokbel, Mahmoud Sakr, Li Xiong, Andreas Züfle, Jussara Almeida, Walid Aref, Gennady Andrienko, Natalia Andrienko, Yang Cao, Sanjay Chawla, et al. 2023. Towards mobility data science (vision paper). *arXiv preprint arXiv:2307.05717* (2023).

[34] Mateusz Nijak, Piotr Skrzypczyński, Krzysztof Ćwian, Michał Zawada, Sebastian Szymczyk, and Jacek Wojciechowski. 2024. On the Importance of Precise Positioning in Robotised Agriculture. *Remote Sensing* 16, 6 (2024).

[35] Kvitoslava Obelovska, Olga Panova, and Vincent Karovič. 2021. Performance Analysis of Wireless Local Area Network for a High-/Low-Priority Traffic Ratio at Different Numbers of Access Categories. *Symmetry* 13, 4 (2021).

[36] M.J. O'Grady, D. Langton, and G.M.P. O'Hare. 2019. Edge computing: A tractable model for smart agriculture? *Artificial Intelligence in Agriculture* 3 (2019), 42–51.

[37] ISO/TMB Strategic Advisory Group on Smart Farming. 2023. Final report with Recommendations. https://www.iso.org/smart-farming/smart-farming-data-driven

[38] Koji Oshima, Daisuke Yamamoto, Atsuhiro Yumoto, Song-Ju Kim, Yusuke Ito, and Mikio Hasegawa. 2022. Online machine learning algorithms to optimize performances of complex wireless communication systems. *Mathematical Biosciences and Engineering* 19, 2 (2022), 2056–2094.

[39] El Mehdi Ouafiq, R. Saadane, A. Chehri, and M. Wahbi. 2022. Data Lake Conception for Smart Farming: A Data Migration Strategy for Big Data Analytics. In *Human Centred Intelligent Systems*, Alfred Zimmermann, Robert J. Howlett, and Lakhmi C. Jain (Eds.). Springer Nature Singapore, Singapore, 191–201.

[40] F Pinet and C De Runz. 2019. Representing Diagrams of Imperfect Geographic Objects. In *Geographic Data Imperfection 1: From Theory to Applications*. Wiley.

[41] Rehan Qureshi, Syed Haris Mehboob, and Muhammad Aamir. 2021. Sustainable Green Fog Computing for Smart Agriculture. *Wirel. Pers. Commun.* 121, 2 (2021).

[42] Prem Rajak, Abhratanu Ganguly, Satadal Adhikary, and Suchandra Bhattacharya. 2023. Internet of Things and smart sensors in agriculture: Scopes and challenges. *Journal of Agriculture and Food Research* 14 (2023), 100776.

[43] Mahesh Ranaweera and Qusay H. Mahmoud. 2021. Virtual to Real-World Transfer Learning: A Systematic Review. *Electronics* 10, 12 (2021).

[44] O Rozenstein, Y. Cohen, V. Alchanatis, K. Behrendt, J Bonfil, G. Eshel, A. Harari, W. Harris, I. Klapp, Y. Laor, R. Linker, K. Paz, S. Peets, Y. Rutter, Y. Salzer, and J. Lowenberg-DeBoer. 2023. Data-driven agriculture and sustainable farming: friends or foes? *Precision Agriculture* 25 (08 2023), 520–531.

[45] Philip Russom. 2017. Data Lakes: Purposes, Practices, Patterns, and Platforms. TDWI white paper.

[46] A Puricelli S Servigne and R Laurini. 2000. A Methodology for Spatial Consistency Improvement of Geographic Databases. *GeoInformatica* 4, 1 (2000), 7–34.

[47] Dimitris Sacharidis, Giorgos Giannopoulos, George Papastefanatos, and Kostas Stefanidis. 2023. Auditing for Spatial Fairness. In *EDBT*. OpenProceedings.org.

[48] Mahmoud Sakr. 2018. A data model and algorithms for a spatial data marketplace. *International Journal of Geographical Information Science* 32, 11 (2018), 2140–2168.

[49] M Sakr and G Merten. 2023. Brussels Mobility Twin. In *Proceedings of the 31st ACM Int'l. Conf. on Advances in Geographic Information Systems*. 1–4.

[50] Mahmoud Sakr, Cyril Ray, and Chiara Renso. 2022. Big mobility data analytics: recent advances and open problems. *GeoInformatica* 26, 4 (2022), 541–549.

[51] Ibrahim Sammour and Gerard Chalhoub. 2022. Application-Level Data Rate Adaptation in Wi-Fi Networks Using Deep Reinforcement Learning. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. 1–7.

[52] Ibrahim Sammour, Gerard Chalhoub, and Gil De Sousa. 2023. Capacity aware Wi-Fi networks deployment. *Annals of Telecommunications* (2023).

[53] D. Sathya, R. Thangamani, and B. Saravana Balaji. 2024. *The Revolution of Edge Computing in Smart Farming*. Springer Nature Switzerland, Cham, 351–389.

[54] Norman Siebrecht. 2020. Sustainable Agriculture and Its Implementation Gap—Overcoming Obstacles to Implementation. *Sustainability* 12, 9 (2020).

[55] D Tiozzo Fasiolo, L Scalera, E Maset, and A Gasparetto. 2023. Towards autonomous mapping in agriculture: A review of supportive technologies for ground robotics. *Robotics and Autonomous Systems* 169 (2023), 104514.

[56] Manas Wakchaure, B.K. Patle, and A.K. Mahindrakar. 2023. Application of AI techniques and robotics in agriculture: A review. *Artificial Intelligence in the Life Sciences* 3 (2023), 100057.

[57] Benjamin Warnke, Johann Mantler, Sven Groppe, Yuri Cotrado Sehgelmeble, and Stefan Fischer. 2022. A SPARQL benchmark for distributed databases in IoT environments. In *Proceedings of The International Workshop on Big Data in Emergent Distributed Environments* (Philadelphia) *(BiDEDE'22)*. Article 5, 6 pages.

[58] Jinyuan Xu, Baoxing Gu, and Guangzhao Tian. 2022. Review of agricultural IoT technology. *Artificial Intelligence in Agriculture* 6 (2022), 10–22.

[59] Darío Fernando Yépez-Ponce, José Vicente Salcedo, Paúl D. Rosero-Montalvo, and Javier Sanchis. 2023. Mobile robotics in smart farming: current trends and applications. *Frontiers in Artificial Intelligence* 6 (2023).

[60] Matei Zaharia, Ali Ghodsi, Reynold Xin, and Michael Armbrust. 2021. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. In *11th Conference on Innovative Data Systems Research, CIDR*.

[61] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. 2019. Deep Learning in Mobile and Wireless Networking: A Survey. *IEEE Communications Surveys & Tutorials* 21, 3 (2019), 2224–2287.

[62] Zhigang Zhang, Cheqing Jin, Jiali Mao, Xiaolin Yang, and Aoying Zhou. 2017. TrajSpark: A Scalable and Efficient In-Memory Management System for Big Trajectory Data. In *Web and Big Data*, Lei Chen, Christian S. Jensen, Cyrus Shahabi, Xiaochun Yang, and Xiang Lian (Eds.). Springer, 11–26.

[63] Esteban Zimányi, Mahmoud Sakr, and Arthur Lesuisse. 2020. MobilityDB: A mobility database based on PostgreSQL and PostGIS. *ACM Transactions on Database Systems (TODS)* 45, 4 (2020), 1–42.