

An AI-Driven Framework for Agricultural Data Interoperability Informed by Literature and Stakeholder Insights

Fjolla Berisha
Computer Science
Maynooth University
Maynooth, Ireland

fjolla.berisha.2024@mumail.ie

Leidiane Da Silva
School of Business
Maynooth University
Maynooth, Ireland

leidiane.goretedasilva.2021@mumail.ie

Erica Tegolo
School of Business
Maynooth University
Maynooth, Ireland

erica.tegolo.2024@mumail.ie

Peter Mooney
Computer Science
Maynooth University
Maynooth, Ireland

peter.mooney@mu.ie

Zohreh Pourzolfaghar
School of Business
Maynooth University
Maynooth, Ireland

zohreh.pourzolfaghar@mu.ie

Markus Helfert
School of Business
Maynooth University
Maynooth, Ireland

markus.helfert@mu.ie

Abstract— This study examines AI-driven solutions for agricultural data interoperability through a combined empirical-theoretical approach, integrating stakeholder input with peer-reviewed research comprising two stakeholder workshops, one involving 28 participants and another involving 50 participants (encompassing agri-tech, farming professionals, policymakers, researchers, professors and industry experts), alongside a literature review of 50 peer-reviewed publications. The workshop revealed that structural incompatibility, semantic mismatches, and reliance on manual processes are major operational bottlenecks. Using thematic analysis of workshop discussions and literature synthesis, we identify five critical interoperability stages: Data Ingestion, Standardization, Ontology Alignment, Integration, and Access/Sharing. By systematically reviewing literature, we identified AI techniques, such as anomaly detection, semantic mapping, and rule-based automation, that directly address these stakeholder-validated challenges, forming the foundation of a sequential AI-driven framework. Findings highlight the need for adaptive AI tools to automate processes like schema alignment and ontology mapping, to reduce preprocessing time. This work bridges theoretical frameworks with practical needs, offering actionable insights for scalable smart farming solutions.

Keywords - Artificial Intelligence, Data Interoperability, Agricultural Data

I. INTRODUCTION

The agricultural sector is undergoing a rapid transition from traditional methods to smart agriculture, in which vast amounts of data are generated through modern technologies [6, 18, 28]. Sensors, IoT devices, satellite imagery, drones, and farm management information systems (FMIS) [24, 61] all contribute to this data influx, such as soil health metrics and weather patterns, supporting crop monitoring and yield predictions [6, 24, 28]. If used effectively, these data sources support improved productivity, sustainability, and profitability in agriculture [15, 22]. However, integrating such a variety of data remains one of the most challenging hurdles to fully unlocking the benefits of smart agriculture [15, 18, 28,

29]. Data interoperability, the ability of different systems and organizations to exchange and make use of information without interference under established processes and protocols [6, 10, 21], is often limited. Because agricultural data is generated from multiple devices, it tends to be stored in isolated silos, using proprietary formats and governed by varied standards [15, 26]. For example, one manufacturer's IoT sensor outputs JSON data, while satellite data may arrive in GeoTIFF, and the FMIS might rely on proprietary XML schemas [6]. This heterogeneity leads to significant barriers for seamless data exchange and reuse, since applications cannot easily leverage data from multiple IoT systems to enhance overall data quality and end-user benefits [29].

Existing frameworks such as the Agricultural Data Application Programming Toolkit (ADAPT) and standards from the Open Geospatial Consortium (OGC) aim to enable automated interoperability [1, 43]. However, these approaches generally do not incorporate artificial intelligence (AI), meaning that integrated data cannot “leverage prior integration mappings for improved harmonization” [50]. Meanwhile, the FAIR (Findable, Accessible, Interoperable, Reusable) principles [60] and design guidelines for Agricultural Data Spaces [42, 54] provide a valuable conceptual foundation for data exchange, yet robust AI-enabled solutions are still needed to operationalize these principles across diverse agricultural contexts [12, 13, 42].

One established route for improving data interoperability is to define and adopt standardized data formats and frameworks, as promoted by organizations like AgGateway and OGC [1, 6]. Recent studies, such as [57], have reviewed well-known interoperability initiatives (e.g., European Interoperability Framework, Data Connector Report, DEMETER) [51] that aim to unify agricultural data. However, these initiatives commonly focus on structural standardisation rather than on intelligent, adaptive integration. Hence, semantic alignment across heterogeneous datasets remains largely unsolved, especially in scenarios where data from multiple sensors, platforms, and software systems must be reconciled in real time [5]. In addition to standardisation

efforts, several studies have explored the potential of AI and machine learning (ML) to automate data processing tasks, particularly in environments where data originates from multiple, heterogeneous sources. Reference [4], by first identifying an IoT/AI technical architecture capable of supporting the development of Smart Sustainable Agriculture (SSA) platform, is proposing a holistic IoT/AI platform to tackle interoperability. Reference [25] also states a lack of interconnection among different technologies. The role of AI in addressing these challenges is further supported by research in other industries, such as healthcare and finance, where machine learning (ML) models have been used to harmonize disparate datasets and enable more efficient data exchange [16, 20]. Ontology alignment has emerged as a critical component of achieving semantic interoperability in agricultural data. Different agricultural systems often use varying terminologies to describe the same concepts, leading to semantic mismatches that hinder data integration. Reference [35] examined the use of AI-based ontology alignment tools to automatically map concepts between different datasets, ensuring that data from different systems can be interpreted and used in a consistent manner. These tools leverage ML models trained on agricultural datasets to recognize relationships between different terminologies and align them with a unified ontology. According to the systematic review [8], AI is revolutionizing agriculture, significantly enhancing efficiency and sustainability. Random Forest (19.2%) and Support Vector Machine (15.9%) are the most frequently used ML algorithms, alongside Artificial Neural Networks (24.9%) and Ensemble Learning (35.6%), demonstrating their versatility across various agricultural challenges. To support interoperability and effective data sharing, organizations like the Dublin Core Metadata Initiative (DCMI) and the International Standards Organization (ISO) develop and maintain standardized metadata. Notable agricultural initiatives include the Agricultural Information Management Standards (AIMS) and the Agricultural Metadata Element Set (AgMES) [10].

Complementing these standardization initiatives, AI has emerged as a critical enabler of interoperability in the public sector. According to [21], around 26% of analyzed AI use cases focus specifically on supporting interoperability, underscoring its importance. AI contributes by structuring, curating, standardizing, and linking public administration data, with common actions including 'Detect' (42%), 'Structure' (22%), and 'Classify' (16%) [21], aligning with and enhancing the utility of metadata standards.

This paper argues that AI can both complement and serve as an alternative to current interoperability frameworks. Its findings emerge from an 18-month ongoing discussion on agricultural data interoperability, incorporating insights from both: two stakeholder workshops with 78 participants in total, including farmers, policymakers, agri-tech companies, and researchers and a review of 50 peer-reviewed studies (2009-2024). The workshops highlighted three major challenges: structural incompatibility, semantic mismatches, and reliance on manual processes, that impede effective data interoperability. Our findings combined identify five key stages essential for data interoperability: Data Ingestion, Standardisation, Ontology Alignment, Integration, and Access/Sharing. These stages leverage techniques such as machine learning, natural language processing, and semantic

mapping to automate the harmonization of agri-data. We synthesize an AI-driven framework that unifies these stages into a sequential process, transforming raw data into shared, actionable information, and a novel contribution to the field. Although interest in AI for agricultural challenges is growing, only a few studies directly address the intersection of AI and data interoperability in agriculture. To address this gap, our review also includes research from sectors from other sectors that use AI for IoT data interoperability [23, 2, 9, 14, 16].

II. METHODOLOGY

We adopt the Design Science Research Methodology (DSRM) [44] to structure the development of the framework. To gather empirical evidence, two stakeholder workshops¹ were conducted in Ireland: the first in June 2024, and the second in February 2025. Together, the workshops engaged 78 participants, including agri-tech developers, farmers, policymakers, researchers, and industry experts. These hybrid (online and in-person) workshops used structured brainstorming and scenario-based discussions to surface recurring interoperability challenges, such as incompatible IoT formats, semantic mismatches in crop terminology, and reliance on manual data handling. These stakeholder-validated insights informed the criteria and focus of the subsequent literature review, which identified AI-driven techniques capable of addressing the same operational gaps. DSRM ensured that the resulting framework was both empirically grounded and systematically refined, bridging practical needs with conceptual design to produce a viable artefact for addressing agricultural data interoperability.

Guided by frameworks on conducting and reporting on systematic reviews [56, 58], this study aims to answer the question: "Which AI-driven methods are being used to address data interoperability challenges in agriculture, and how effective are they in overcoming issues of structural incompatibility, semantic mismatches, and reliance on manual processes?". The studies were manually sourced from Scopus and Google Scholar, with contributions from IEEE (13 papers), MDPI (9 papers), Science Direct (15 papers), Springer (7 papers) and other databases (6 papers). The search focused on this search string: ("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning") AND ("data interoperability" OR "Interoperability" OR "structural incompatibility" OR "format mismatch" OR "vocabulary conflict" OR "manual data processing" OR "conflict resolution" OR "data sharing governance") AND ("agricultur*" OR "precision farming" OR "farm" OR "IoT" OR "Internet of things"). The results were exported to an Excel file. The literature review spans peer-reviewed studies from 2009 to 2024. This broad year range was chosen to capture both the early foundational work on AI methods for data interoperability and the latest innovative approaches. Our initial search yielded 372 publications. After removing duplicates (79) and non-English articles (35), we screened 114 records by title/abstract. We excluded 64 records for: not discussing AI/ML techniques (17), lacking agricultural focus (41), or not addressing stakeholder-identified interoperability challenges (6). The remaining 50 full-text papers were assessed for eligibility. We included papers that explicitly addressed AI/ML techniques in connection with agricultural data interoperability challenges (structural, semantic, or

¹ <https://www.entrustdn.eu>

manual processes) when integrating sources like IoT sensors, satellite imagery, or farm management systems [5]. The review was limited to English-language publications due to English search terms.

The authors did not apply a quality assessment of the papers' discussions or arguments since the goal was to capture how AI methods are being used rather than to evaluate their rigor in depth. Each included publication was read in its entirety, with a focus on the sections describing AI methods, models, techniques and their role in facilitating interoperability. The reviewed papers are [2], [4]-[8], [9]-[11], [12]-[20], [23]-[33], [34]-[36], [38]-[41], [45]-[55], [57], [59]-[60].

Their insights were systematically coded and cross-referenced with themes from the literature review (Structural incompatibility, semantic mismatches, and reliance on manual processes) to ensure alignment between theoretical frameworks and on-the-ground needs. By synthesizing approaches from 50 peer-reviewed papers and integrating them with workshop insights, we observed that several AI-driven methods for data interoperability consistently tackled specific challenges, starting with capturing and preprocessing raw data, then converting and harmonizing various formats, followed by resolving semantic differences, merging heterogeneous sources, and finally ensuring secure, regulated access to the integrated data. This sequential progression, evident across multiple studies, naturally gave us the idea to group the studies into five stages of interoperability: ingestion, standardisation, ontology alignment, integration, and access/sharing. Each stage represents a critical transformation point, reflecting both the technical methodologies and the inherent challenges noted in the literature. Due to their broad focus, not all papers could be classified into specific categories.

III. RESULTS

Workshop participants: farmers, agri-tech developers, policymakers, and researchers, consistently emphasized interoperability pain points through scenario-based discussions. Farmers reported semantic mismatches as a daily hurdle ('One platform calls it "soil moisture"; another says "groundwater", same data, different terms, impossible to combine'). Agri-tech developers highlighted structural incompatibility in IoT ecosystems ('JSON from sensors, XML from tractors, manual reformatting eats 30% of dev time'). Policymakers and researchers universally decried reliance on manual processes ('We have the data, but stitching spreadsheets for subsidy audits takes weeks'). Through iterative thematic coding of session transcripts, these specific frustrations merged into three cross-cutting challenges: (1) structural incompatibility across data formats, (2) semantic mismatches in agricultural terminology, and (3) workflow dependency on error-prone manual interventions. The recurrence of these themes across both workshops confirmed their centrality to agricultural interoperability.

Analysis of the resulting 50 studies revealed a consistent sequential pattern in how interoperability was achieved and they reflect the logical flow of agricultural data management, starting with the initial capture and preprocessing of datasets (Data Ingestion) [27], through harmonization of formats (Data Standardisation) [47] and semantic concepts (Ontology Alignment), culminating in unified datasets (Data Integration)

that can be securely accessed and shared among stakeholders (Data Access and Sharing) [57]. Each stage faces its own distinct challenges and employs specialized AI techniques to address them. Building on these insights, we subsequently merge these stages into a conceptual AI framework, illustrating how they can collectively address data interoperability challenges in agriculture. This framework serves as a synthesized representation of current AI-driven practices and their interrelationships, forming the backbone of our discussion on future directions and potential gaps in the field.

A. Data Ingestion

The Data Ingestion Stage is responsible for collecting and preprocessing agricultural data from different sources such as IoT sensors, satellite imagery, weather reports, and FMIS [24, 51, 55]. To handle this data heterogeneity effectively, several AI models are suggested in the literature for automating categorization, anomaly detection, and metadata extraction, which would make the stage more efficient compared to traditional rule-based methods [9]. For data categorization, Support Vector Machines (SVM) [16, 32] and Decision Trees [19, 47] are suggested due to their capability to handle high, dimensional data and classify structured data inputs effectively. Decision Trees are particularly useful for classifying data with clear hierarchies or decision rules [47, 36], which aligns well with the structured nature of agricultural data inputs (e.g., categorizing crop data based on season, soil type, or location) [33].

Anomaly detection could be achieved using models like Isolation Forest or One-Class SVM [2, 25, 52, 53], which are ideal for identifying irregularities within high, dimensional datasets [38]. Isolation Forest isolates anomalies by evaluating the distance of data points from expected patterns [52], while One-Class SVM constructs a decision boundary that encompasses most normal data, making it effective for detecting inconsistencies [19]. These models are particularly effective for detecting anomalies in IoT sensor data [55]. For instance, if a soil moisture sensor usually records values between 20-30%, but suddenly shows 80%, this anomaly can be flagged as a potential sensor malfunction or an irrigation system leak. Such anomaly detection helps in maintaining data quality and operational efficiency in smart farming systems [32]. Rule-based methods would handle basic entity extraction, while decision trees can classify and extract complex metadata attributes like timestamps and locations by leveraging from predefined rules and patterns [9, 46]. However, while high interpretable models such as decision trees and rule-based systems offer clear, human-readable decision paths, this simplicity often comes at the cost of lower predictive accuracy in complex or high-dimensional datasets [2]. On the other hand, SVMs work well with limited data and are better at minimizing overfitting compared to neural networks and decision trees [32].

B. Data Standardisation

The Data Standardisation Stage ensures that data from agricultural sources is compatible and can be integrated seamlessly [2, 34]. This process involves operations like anomaly detection, format conversion, and schema alignment, which traditionally require extensive manual effort [50]. By incorporating AI models, these tasks can be automated to reduce complexity and increase efficiency [20].

For anomaly detection within the standardisation process, basic statistical models such as Z-score analysis are suggested [59]. Z-score can effectively identify outliers based on deviations from the mean, making it suitable for detecting anomalies like missing values or extreme outliers [59]. In the context of agricultural data standardisation, it can detect irregularities in continuous variables, such as temperature or precipitation levels. To standardize diverse input formats (e.g., CSV, JSON, XML), decision tree-based models are recommended [15, 47]. Decision trees can be used to implement conversion rules by learning from predefined patterns, allowing for the automatic transformation of heterogeneous data structures into a unified format [47]. For schema alignment and normalization, a rule-based system could be applied [14]. Rule-based methods are effective for enforcing schema consistency and unit standardisation, such as converting temperature values to a common unit [14]. For example, one dataset might record yield in “tons per hectare,” while another uses “kilograms per square meter.” A rule-based system can automatically convert and align these units to ensure a unified schema, making the data comparable across sources [32]. Despite such capabilities, [25] emphasize that many decision support tools, including rule-based systems, see limited practical use in agriculture, primarily because they depend on highly detailed and often hard-to-obtain input data. Another major limitation of rule-based systems according to [32] is poor scalability, since creating rules for every subtype or new class becomes computationally expensive and impractical. As for the Z-score method, while effective for outlier recognition, it does not allow for the automatic detection of the optimal number of anomalies within a dataset [59].

C. Ontology Alignment

The Ontology Alignment Stage addresses the semantic inconsistencies that often arise when integrating agricultural data from multiple sources. This stage aligns incoming data with predefined ontologies to ensure that different terminologies and structures are interpreted consistently, thus enabling semantic interoperability across various systems [45]. To automate ontology alignment, lightweight AI techniques such as word embeddings are proposed in literature [2, 39]. Word embeddings like Word2Vec are effective for capturing semantic relationships between terms, making them suitable for matching and aligning data fields with ontology concepts [23]. For example, in one dataset, the term “irrigation” may be referred to as “water management” in another. Word2Vec can identify the semantic similarity between these terms and align them to a common ontology, ensuring that different datasets can be integrated seamlessly [45]. Additionally, rule-based methods can be combined with word embeddings to refine the alignment process [39]. In cases where terminologies vary significantly (e.g., “herbicide application” vs. “weed control measures”), word embeddings can identify the semantic overlap, while rule-based methods validate these alignments using predefined rules [2]. For example, if “herbicide” is always associated with “chemical application” in the ontology, the rule-based system can confirm that “weed control measures” are indeed a subset of herbicide applications (Deppa et al., 2024). Once aligned, the

data can be enriched with metadata indicating the alignment status and ontology mappings. This ensures that the data is semantically coherent and ready for further integration and analysis [31]. The Semantic Web is presented in [9] as a foundational technology for achieving knowledge management, sharing, and reuse through formal ontologies. Protégé, an ontology editor developed by Stanford in 1999, is widely used by ontology engineers to design and manage ontologies, such as the initial Common Learning Object Ontology. Complementing this, WordNet, a lexical database introduced by Miller in 1995, is employed to calculate semantic similarity between concept and property terms [9]. This facilitates the detection of naming conflicts during ontology alignment, helping ensure consistency and coherence across integrated data models. However, Natural Language Processing (NLP) techniques, such as Word2Vec-based word embeddings, still face limitations in fully capturing nuanced meanings, leading to persistent challenges in achieving truly context-aware understanding [39]. A significant challenge arises when a word has multiple meanings, as static embeddings like Word2Vec assign a single vector to each word, ignoring context-specific variations in meaning [45]. While rule-based reasoning offers powerful deductive capabilities, its practical application in ontology mapping is constrained by the need for significant manual effort in rule and mapping creation [9], and its current inability to fully address the vast spectrum and complexity of existing conflicts. To address these limitations, there is a pressing need for well-defined and widely accepted technologies for representing data semantics. Technologies such as the Resource Description Framework (RDF), Web Ontology Language (OWL), and Linked Data [30] provide a structured foundation for semantic interoperability [30]. When combined with domain-specific ontologies like AGROVOC and AgOnt, these tools enable the use of shared vocabularies and structured relationships [30], thereby reducing semantic ambiguities and enhancing data interoperability across agricultural systems.

D. Data Integration

The Data Integration Stage merges data from various external sources with preferably the aligned data produced in the ontology alignment phase, ensuring that all information is consolidated into a unified dataset [9, 47]. This process traditionally involves manual tasks like schema matching and conflict resolution, which can be time-consuming and error prone [39]. To automate these tasks, several AI techniques are used in the literature to facilitate more efficient integration. For schema matching, simple similarity, based methods such as Jaccard Similarity are proposed [38]. Jaccard Similarity measures the similarity between datasets based on shared attributes, making it effective for matching schemas that use different terminologies to describe the same concept [38]. In agricultural data integration, one dataset may use “harvest date,” while another uses “crop maturity date.” Jaccard Similarity can identify these attributes as related, enabling their integration into a unified schema.

In cases where data conflicts arise, such as differing values or formats for the same attribute, rule-based conflict resolution methods can be employed. These methods use predefined rules to decide which data source should be prioritized based on criteria like source reliability or timestamp, ensuring consistency in the integrated dataset [46]. For more complex integration scenarios, probabilistic reasoning models, such as Bayesian networks, could be applied to handle uncertainties and dependencies between data fields [15, 16]. However, to maintain simplicity and efficiency, these models should be reserved for situations where rule-based methods are insufficient.

E. Data Access and Sharing

Most interoperability frameworks stop at data integration. The Access and Sharing module in this framework extend beyond technical integration to embed governance and control mechanisms aligned with data sovereignty [41, 53, 54]. This module ensures that integrated data is accessible and reusable by various stakeholders through APIs and other data sharing mechanisms [41, 49, 10]. The stage is responsible for managing data access based on user roles and enforcing privacy policies to protect sensitive information [34]. Traditionally, implementing these controls requires manual configuration of role, based access control (RBAC), and privacy enforcement rules, making it labour-intensive and prone to errors [52]. To automate these tasks, a rule-based AI approach is suggested. This approach dynamically assigns access permissions through operations like granting access, which determines user access rights based on predefined roles (e.g., user, admin) and privacy policies [55]. For monitoring, basic anomaly detection models are used to track and flag any unauthorized access attempts in real time [59]. When sharing data and enforcing privacy rules to ensure compliance with established privacy levels, such as anonymization requirements or consent, based sharing restrictions [5] are used in literature. For instance, a farm manager might have full access to all data, while a third, party agronomist might only have access to aggregated crop health data. The rule-based system can enforce these access rights dynamically as user roles change or as new data is added [25, 32].

F. AI-driven Framework for Data Interoperability

In Figure 1 we propose a conceptual AI-driven framework for achieving data interoperability in agriculture, developed through an iterative design process grounded in stakeholder input and supported by a review of 50 peer-reviewed studies spanning 2009 to 2024. Applying Design Science Research Methodology (DSRM), we first identified three recurring challenges from stakeholder workshops: structural incompatibility across data formats, semantic mismatches between agricultural terminologies, and a heavy reliance on manual integration processes. These challenges shaped the

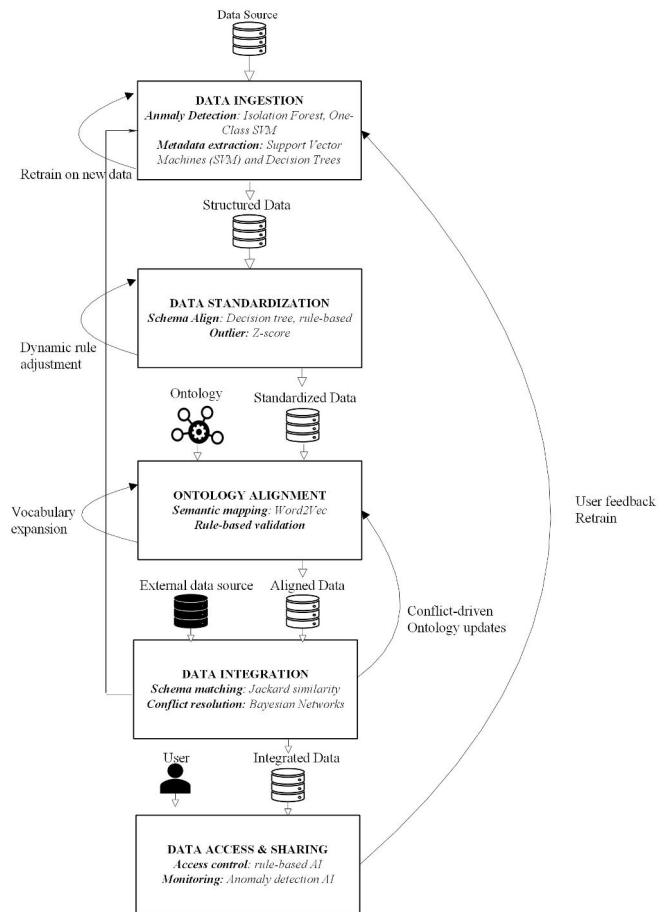


Fig 1. AI-driven Framework for Data Interoperability

structure and focus of the literature review, which sought AI methods specifically capable of addressing each identified gap. The resulting five-stage framework, comprising Data Ingestion, Standardization, Ontology Alignment, Integration, and Access/Sharing, maps these AI techniques into a sequential pipeline. Raw data from IoT sensors, satellite imagery, and FMIS enters the Ingestion stage, where anomaly detection and metadata extraction are applied to clean and classify inputs. In the Standardization stage, format conversion and unit normalization harmonize heterogeneous data [14]. The Ontology Alignment stage applies semantic mapping and rule-based validation to reconcile terminological inconsistencies [17]. At the Integration stage, aligned data is merged using schema matching and conflict resolution. Finally, in the Access and Sharing stage, AI-enforced access controls and real-time monitoring mechanisms regulate data distribution in compliance with privacy and usage rules. This sequential design directly addresses the three core interoperability challenges identified by stakeholders: structural incompatibility is mitigated in Ingestion (anomaly detection for format errors) and Standardization (automated format conversion) stages.

Semantic mismatches are resolved in the Ontology Alignment stage (semantic mapping of agricultural terminologies). Reliance on manual processes is reduced through AI automation across all stages, most critically in Integration (automated conflict resolution) and Access/Sharing (rule-based access controls).

IV. DISCUSSION AND FUTURE DIRECTIONS

The transition to smart agriculture hinges on overcoming data interoperability challenges exacerbated by structural incompatibility, semantic mismatches, and reliance on manual processes, as identified from the workshops participants. Our paper builds on foundational efforts like the FAIR principles [60] and structural standards such as ADAPT [1] but diverges by prioritizing AI-enabled automation to address scalability and adaptability gaps. For instance, while ADAPT focuses on format standardization, our Ontology Alignment stage leverages semantic AI tools (e.g., Word2Vec) to resolve terminology conflicts, a challenge persistently noted in cross-platform agricultural data integration [6, 45]. Similarly, rule-based systems for schema alignment [14] were extended in our framework to incorporate probabilistic reasoning (e.g., Bayesian networks) for conflict resolution, addressing critiques of rigidity in traditional standardization frameworks [15, 32].

The sequential structure of the framework, progressing from ingestion to sharing, mirrors dependencies observed in real-world data pipelines. For example, [28] emphasized that standardizing units (e.g., converting “tons/hectare” to “kg/m²” using decision trees) [14] must precede schema matching to avoid cascading errors. [47] further validated Jaccard Similarity for schema alignment, achieving higher accuracy in heterogeneous systems. Crucially, the framework design embeds unresolved challenges identified in the literature. For instance, federated learning, proposed for privacy, preserving integration [5], remains untested in large, scale agricultural systems [2], while edge computing for real-time irrigation management is limited to theoretical models [25]. By mapping these gaps alongside validated solutions, the framework serves as both a synthesis of current capabilities and a roadmap for future research.

While AI literature often highlights the strengths of models like SVMs, NLP and Decision Trees, their limitations, especially in adapting to various data types, are less thoroughly examined. A key oversight is the tendency to evaluate these techniques in isolation, without accounting for how errors may propagate across different stages of the AI pipeline. This is especially problematic given that the effectiveness of AI systems depends heavily on the quality, quantity, diversity, and privacy of the underlying data [2, 25]. Poor data quality can lead to inaccurate or biased models, with potentially harmful outcomes in critical domains such as healthcare, finance, and agriculture. Insufficient or non-representative datasets may result in oversimplified models that fail to generalize to complex, real-world scenarios, while lack of diversity can perpetuate systemic bias [2,25]. Many AI models face high computational overhead and latency [48]. This challenge is closely linked to a broader dilemma identified by [48]: whether to prioritize complex machine learning algorithms that demand large datasets, such as deep learning models, or to rely on smaller but higher-quality datasets. There is no clear consensus among ML experts on which approach is more effective. Based on interviews with agriculture and food system researchers, [5] highlights that developing trustworthy AI in agriculture faces significant hurdles, primarily due to farmers' reluctance to share data, ethical conflicts surrounding data privacy and transparency, and regulations that are often inadequate or misaligned with practical implementation [11]. Furthermore, Machine

Learning-based systems often require real-time processing and decision-making capabilities which remote regions or resource-constrained enterprises may lack [8]. Advancements in Application Programming Interfaces (APIs) can further enhance interoperability between different precision agriculture tools, allowing farmers to choose solutions that best fit their specific needs [40].

Future work must prioritize practical implementation and validation of the framework, such as harmonizing weather sensors, soil data, and satellite imagery for crop yield prediction, to test the framework's robustness against real-world noise and legacy system incompatibility [15, 28]. Integration with platforms like AgGateway and European Data Spaces could operationalize the framework as a modular API, while co-designed governance models [42] would address farmer-centric ethical concerns [37]. As a next step, we propose developing a small-scale prototype using mock agricultural data sets. This use case would simulate the integration of soil moisture sensor readings (CSV), satellite imagery (GeoTIFF), and weather forecast data (JSON) across the five framework stages. AI techniques such as Isolation Forest for anomaly detection, decision trees for format conversion, Word2Vec for semantic alignment, and Jaccard Similarity for schema matching would be applied at their respective stages. Role-based access control would demonstrate dynamic sharing restrictions. While based on synthetic data, the prototype would serve as a proof-of-concept to demonstrate the framework's technical feasibility and logical coherence, laying the groundwork for future field-scale implementations.

V. CONCLUSION

Agricultural data interoperability faces three core challenges: structural incompatibility, semantic mismatches and reliance on manual processes, identified through a workshop with 78 agricultural stakeholders. Complementing this, a review of 50 studies (2009–2024) revealed five critical stages in data interoperability: Data Ingestion, Data Standardisation, Ontology Alignment, Data Integration, and Data Access & Sharing. The five stages, each aligned with the most suitable AI techniques, were integrated into an AI-driven framework designed to transform raw agricultural data into shared, actionable information, addressing key challenges such as structural incompatibility, semantic mismatches, and reliance on manual processes.

Data Ingestion is responsible for collecting and preprocessing raw agricultural data from sources such as IoT sensors, satellite imagery, weather reports, and Farm Management Information Systems (FMIS). Anomaly Detection leverages techniques like Isolation Forest and One-Class SVM, to identify irregularities in high-dimensional datasets. Metadata Extraction leverages Support Vector Machines (SVM) and Decision Trees to automate categorization and extract metadata.

Data Standardization ensures that agricultural data are compatible and can be seamlessly integrated. Anomaly Detection employs basic statistical models, such as Z-score analysis, to identify outliers, like missing values or extreme deviations in continuous variables such as temperature or precipitation levels. Format Conversion can be facilitated using decision tree-based models. Schema Alignment and Normalization rely on rule-based systems to enforce consistency in schema design and unit standardization.

Ontology Alignment addresses the semantic inconsistencies that often arise when integrating agricultural data from multiple sources. One effective approach involves the use of word embeddings, such as Word2Vec, which can capture semantic relationships between terms. In addition, rule-based validation methods can be integrated with word embeddings to further refine the alignment process.

Data Integration involves merging data from various external sources with the semantically aligned data generated during the ontology alignment phase. To streamline schema matching, similarity-based methods like Jaccard Similarity are commonly proposed. For conflict resolution, rule-based methods are typically used to address inconsistencies such as differing values or formats for the same attribute. Bayesian networks can be employed to infer the most likely correct values by considering conditional dependencies and uncertainties across sources.

Data Access and Sharing represents the final stage of the interoperability framework, extending beyond technical integration to incorporate governance and control mechanisms rooted in data sovereignty principles. This module ensures that the integrated agricultural data is not only accessible but also reusable by various stakeholders, such as farmers, policymakers, and agri-tech providers, via APIs and other data-sharing channels. To automate and scale these functions, a rule-based AI approach is recommended by assigning permissions based on predefined user roles and associated privacy rules. Additionally, monitoring mechanisms are embedded to ensure ongoing compliance and security. Basic anomaly detection models are applied to track and flag unauthorized access attempts in real time, supporting immediate responses to potential breaches and reinforcing trust in the data-sharing ecosystem.

The framework provides a structured approach to agricultural data interoperability, moving from raw data ingestion to secure access and sharing. It highlights how AI can automate and enhance each step, addressing the inherent challenges of data heterogeneity and semantic inconsistencies in smart farming.

This study is constrained by its focus on conceptual design since the proposed framework awaits empirical validation. Additionally, the literature review excluded non-English studies, potentially omitting region-specific innovations, and the relatively low number of workshop participants limits the breadth of stakeholder perspectives. Future work should expand the scope to include a larger, more distinct sample of participants, global case studies, and sector-specific adaptations to enhance the framework's generalizability and practical applicability. We call for research on pilot implementations of the framework we presented, to test real-world viability and expand stakeholder engagement for global adaptability. To the best of our knowledge, this is the first study to propose a comprehensive AI-driven framework for agricultural data interoperability that integrates theoretical insights from literature and empirical stakeholder input, addressing structural, semantic, and manual integration challenges across five sequential stages from ingestion to secure sharing.

By grounding future efforts in DSR's iterative cycles of build-evaluate-refine, this framework can evolve from a theoretical construct into a robust, stakeholder-validated solution. Collaboration across academia, industry, and policy

will be critical to translate interoperability theory into equitable, scalable practice, ultimately empowering agriculture to harness AI's full potential.

ACKNOWLEDGMENT

This research was supported by the EnTrust project: Next Generation of Trustworthy Agri-Data Management, funded under the Marie Skłodowska-Curie Actions (MSCA) Doctoral Network as part of Horizon Europe (Grant Agreement No. 101073381). The authors would like to thank the industrial supervisors Dr Aoibhin Moore Heslin, Dr Paul O'Keefee from Airfield Estate and Dr Stuart Green from Teagasc.

REFERENCES

- [1] AgGateway, "AgGateway Standards for Agricultural Data Exchange," 2019. [Online]. Available: <https://www.aggateway.org>.
- [2] A. Aldoseri, K. N. A. - Khalifa, and A. M. Hamouda, "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges," *Applied Sciences*, vol. 13, no. 12, p. 7082, 2023.
- [3] C. S. Alexander, M. A. Yarborough, and A. Smith, "Who is responsible for 'responsible AI'? Navigating challenges to build trust in AI agriculture and food system technology," *Precision Agriculture*, vol. 25, Aug. 2023.
- [4] E. Alreshidi, "Smart Sustainable Agriculture (SSA) Solution Underpinned by Internet of Things (IoT) and Artificial Intelligence (AI)," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019.
- [5] A. A. AlZubi and K. Galya, "Artificial Intelligence and Internet of Things for Sustainable Farming and Smart Agriculture," in *IEEE Access*, vol. 11, pp. 78686-78692, 2023.
- [6] A. Anidu and R. Dara, "A review of data governance challenges in smart farming and potential solutions," *2021 IEEE International Symposium on Technology and Society (ISTAS)*, Waterloo, ON, Canada, 2021, pp. 1-8.
- [7] J. M. Antle, J. W. Jones, and C. E. Rosenzweig, "Next generation agricultural system data, models and knowledge products: Introduction," *Agricultural Systems*, vol. 155, pp. 186-190, Jul. 2017.
- [8] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives," *Agronomy*, vol. 13, no. 12, p. 2976, Dec. 2023.
- [9] N. Arch-int and S. Arch-int, "Semantic Ontology Mapping for Interoperability of Learning Resource Systems using a rule-based reasoning approach," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7428-7443, Dec. 2013.
- [10] C. Bahlo, P. Dahlhaus, H. Thompson, and M. Trotter, "The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review," *Computers and Electronics in Agriculture*, vol. 156, pp. 459-466, Jan. 2019.
- [11] A. Z. Bayih, J. Morales, Y. Assabie, and R. A. de By, "Utilization of Internet of Things and Wireless Sensor Networks for Sustainable Smallholder Agriculture," *Sensors*, vol. 22, no. 9, p. 3273, Apr. 2022.
- [12] I. CHONG and S. Ali, "Schema Ontology Model to Support Semantic Interoperability in Healthcare Applications: Use Case of Depressive Disorder," *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*, Jeju Island, Korea, Republic of, 2021, pp. 409-412.
- [13] A. C. Cob-Parro, Y. Lalangui, and R. Lazcano, "Fostering Agricultural Transformation through AI: An Open-Source AI Architecture Exploiting the MLOps Paradigm," *Agronomy*, vol. 14, no. 2, p. 259, Feb. 2024.
- [14] M. R. Costa-jussà and R. E. Banchs, "Automatic normalization of short texts by combining statistical and rule-based techniques," *Language Resources and Evaluation*, vol. 47, no. 1, pp. 179-193, May 2012.
- [15] A. Cravero, S. Pardo, P. Galeas, J. López Fenner, and M. Caniupán, "Data Type and Data Sources for Agricultural Big Data and Machine Learning," *Sustainability*, vol. 14, no. 23, p. 16131, Jan. 2022.

- [16] S. Cui et al., "Advances and applications of machine learning and deep learning in environmental ecology and health," *Environmental Pollution*, vol. 335, p. 122358, Oct. 2023.
- [17] R. Deepa, V. Jayalakshmi, P. Thilakavathy, A. Bennet Prabhu and R. Surendran, "Quantum Machine Learning Techniques based on Nurturing Agri-Ontology Framework in Agricultural Science," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAIC), Salem, India, 2024, pp. 413-420.
- [18] T. Dibbern, L. Alvim, and S. Maria, "Main drivers and barriers to the adoption of Digital Agriculture technologies," *Smart agricultural technology*, vol. 8, pp. 100459–100459, Apr. 2024.
- [19] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, 'High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,' *Pattern Recognition*, vol. 58, pp. 121–134, 2016.
- [20] A. Esteva et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019.
- [21] L. Tangi, M. Combetto, J. Martin Bosch, and A. P. Rodriguez Müller, 'Artificial Intelligence for Interoperability in the European Public Sector: An exploratory study', Publications Office of the European Union, Luxembourg, 2023, EUR 31675 EN, doi: 10.2760/633646.
- [22] FAO, "Home," Food and Agriculture Organization of the United Nations, 2019. <http://www.fao.org>.
- [23] P. Fatima, S. K. Fathima, A. M. A. Khatatneh and M. K. A. Qudah, "Machine Learning Methods for Extracting Newspaper Articles from PDF Files," 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2024, pp. 1-6.
- [24] A. Gabriel and M. Gandorfer, "Adoption of digital technologies in agriculture—an inventory in a european small-scale farming region," *Precision Agriculture*, Jul. 2022.
- [25] D. Gackstetter et al., "Autonomous field management – An enabler of sustainable future in agriculture," *Agricultural Systems*, vol. 206, p. 103607, Mar. 2023.
- [26] Maya Gopal P.S. and B. R. Chintala, "Big Data Challenges and Opportunities in Agriculture," *International Journal of Agricultural and Environmental Information Systems*, vol. 11, no. 1, pp. 48–66, Jan. 2020.
- [27] T. Hlupić and J. Puniš, "An Overview of Current Trends in Data Ingestion and Integration," 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2021, pp. 1265-1270.
- [28] J. W. Jones et al., "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science," *Agricultural Systems*, vol. 155, pp. 269–288, Jul. 2017.
- [29] N. Kalatzis, N. Marianos and F. Chatzipapadopoulos, "IoT and data interoperability in agriculture: A case study on the gaiasense™ smart farming solution," *IEEE 2019 Global IoT Summit (GloTS)*, pp. 1-6, June 2019.
- [30] A. Kamilaris, A. Kartakoullis and F.X. Prenafeta-Boldú, (2017), "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, pp. 23-37, December, 2017.
- [31] E. E. Kim et al., "Water economy with smart water system in the City of Carouge," 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), Barcelona, Spain, 2022, pp. 1-6.
- [32] B. Koo, S. La, N.-W. Cho, and Y. Yu, "Using support vector machines to classify building elements for checking the semantic integrity of building information models," *Automation in Construction*, vol. 98, pp. 183–194, Feb. 2019.
- [33] P. Kumar et al., "Role of artificial intelligence, sensor technology, big data in agriculture: next-generation farming," *Elsevier eBooks*, pp. 625–639, Jan. 2022.
- [34] E. Lee, Y. -D. Seo, S. -R. Oh and Y. -G. Kim, "A Survey on Standards for Interoperability and Security in the Internet of Things," in *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1020-1047, Secondquarter 2021.
- [35] J. Lehmann et al., "DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, Jan. 2015.
- [36] P. H. Luzolo, Z. Elrawashdeh, I. Tchappi, S. Galland, and F. Outay, "Combining Multi-Agent Systems and Artificial Intelligence of Things: Technical challenges and gains," *Internet of Things*, vol. 28, p. 101364, Sep. 2024.
- [37] R. Iphofen, Ed., *Handbook of Research Ethics and Scientific Integrity*. Cham: Springer International Publishing, 2020.
- [38] A. Makwana and A. Ganatra, "A Better Approach to Ontology Integration using Clustering Through Global Similarity Measure," *Journal of Computer Science*, vol. 14, no. 6, pp. 854–867, Jun. 2018.
- [39] N. Melluso, I. Grangel-González, and G. Fantoni, "Enhancing Industry 4.0 standards interoperability via knowledge graphs with natural language processing," *Computers in Industry*, vol. 140, p. 103676, Sep. 2022.
- [40] G. Mgendi, "Unlocking the potential of precision agriculture for sustainable farming," *Discover Agriculture*, vol. 2, no. 1, Nov. 2024.
- [41] S. Mohr and R. Kühl, "Acceptance of artificial intelligence in German agriculture: an application of the technology acceptance model and the theory of planned behavior," *Precision Agriculture*, vol. 22, Jun. 2021.
- [42] L. Nagel and Douwe Lycklama, "Design Principles for Data Spaces - Position Paper," Jul. 2021.
- [43] OGC. "Open Geospatial Consortium: Standards for Geospatial Data". 2019. [Online]. Available: <http://www.opengeospatial.org>
- [44] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007.
- [45] I. Pembeci, "Using Word Embeddings for Ontology Enrichment," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. 3, p. 49, Jul. 2016.
- [46] B. Qin, Y. Xia, S. Prabhakar and Y. Tu, "A Rule-Based Classification Algorithm for Uncertain Data," 2009 IEEE 25th International Conference on Data Engineering, Shanghai, China, 2009, pp. 1633-1640.
- [47] M. Raoui, L. Rassam, M. H. El Yazidi and A. Zellou, "Automated Interoperability based on Decision Tree for Schema Matching," 2022 International Conference on Computational Modelling, Simulation and Optimization (ICCMO), Pathum Thani, Thailand, 2022, pp. 48-52.
- [48] Z. Rehman, N. Tariq, Syed Atif Moqurrab, J. Yoo, and G. Srivastava, "Machine learning and internet of things applications in enterprise architectures: Solutions, challenges, and open issues," *Expert Systems*, Oct. 2023.
- [49] I. Roussaki et al., "Building an interoperable space for smart agriculture," *Digital Communications and Networks*, Feb. 2022.
- [50] V. Saiz-Rubio and F. Rovira-Más, "From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management," *Agronomy*, vol. 10, no. 2, p. 207, Feb. 2020.
- [51] M. San Emeterio de la Parte, J.-F. Martínez-Ortega, V. Hernández Díaz, and N. L. Martínez, "Big Data and precision agriculture: a novel spatio-temporal semantic IoT data management framework for improved interoperability," *Journal of Big Data*, vol. 10, no. 1, Apr. 2023.
- [52] L. Sanchez et al., "Data Enrichment Toolchain: A Data Linking and Enrichment Platform for Heterogeneous Data," *IEEE Access*, vol. 11, pp. 103079–103091, Jan. 2023.
- [53] E. Elikem et al., "IoT Solutions with Artificial Intelligence Technologies for Precision Agriculture: Definitions, Applications, Challenges, and Opportunities," *Electronics*, vol. 13, no. 10, pp. 1894–1894, May 2024.
- [54] M. Šestak and D. Copot, "Towards Trusted Data Sharing and Exchange in Agro-Food Supply Chains: Design Principles for Agricultural Data Spaces," *Sustainability*, vol. 15, no. 18, p. 13746, Jan. 2023.
- [55] A. Shoomal, M. Jahanbakht, P. J. Componation, and D. Ozay, "Enhancing Supply Chain Resilience and Efficiency through Internet of Things Integration: Challenges and Opportunities," *Internet of Things*, vol. 27, pp. 101324–101324, Oct. 2024.
- [56] D. Strech and N. Sofaer, "How to write a systematic review of reasons," *Journal of Medical Ethics*, vol. 38, no. 2, pp. 121–126, Nov. 2011.
- [57] K. Sullivan, J. McLaughlin, C. O'Meara, K. McDonnell and C. Kehoe, "Open-Source Tools and Supports to Advance Data Interoperability in the Agriculture Domain," 2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS), London, United Kingdom, 2024, pp. 1-6.
- [58] D. Tranfield, D. Denyer, and P. Smart, "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review," *British Journal of Management*, vol. 14, no. 3, pp. 207–222, Sep. 2003.

- [59] P. Venkatanusha, Ch. Anuradha, Dr. P. S. R. C. Murty, and Dr. S. K. Chebrolu, "Detecting Outliers in High Dimensional Data Sets using Z-Score Methodology," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 48–53, Nov. 2019.
- [60] M. D. Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data*, vol. 3, no. 1, Mar. 2016.
- [61] Q. Zaman, "Precision Agriculture", Academic Press 2023. Available at: <https://www.sciencedirect.com/book/9780443189531/precision-agriculture..>