Knowledge graph for Intelligent Farming and Crop Monitoring

Arjun Harish, Dhairya Rupesh Kamalia, Kavya Chetankumar Parekh,
Raj Mani Mitresh Mani, Samanvitha Patel

School of Computing and Augmented Intelligence
Arizona State University, AZ, USA

{aharish9, dkamali1, kparekh5, rmitresh, spate372}@asu.edu

***Abstract*—Precision agriculture generates soil, weather, crop, and pest data that often remain in silos, limiting cross-domain reasoning. This proposal leverages OWL 2 with SWRL to enable dynamic inference over a unified knowledge graph. It imports and harmonizes the Sensor, Observation, Sample, and Actuator (SOSA/SSN) ontology, the Environment Ontology (ENVO), AGROVOC, and the Crop Ontology (CO) to address interoperability challenges. Inferences such as optimal fertilizer application timing and crop rotation sequencing are derived based on soil fertility, historical and forecasted rainfall, crop type, and pest risk. The outcome is an integrated and semantically interoperable system that enables fertilizer optimization, crop rotation planning, and data-driven agricultural decision-making. The project culminates in a prototype web application that demonstrates reasoning-driven recommendations for sustainable agriculture.**

***Index Terms*—Semantic Web, OWL 2, SWRL, SOSA/SSN, ENVO, AGROVOC, Crop Ontology, Smart Farming, Crop Monitoring, Reasoning Systems, Precision Agriculture**

I. Introduction

Farming creates large-volume sensor, prediction, and farm activity data now. Such heterogeneous sources of data convey rarely interoperable information. **This project is motivated by the need to support sustainable practices such as crop rotation — determining which crop should follow another to improve soil fertility, minimize pest recurrence, and enhance yield.**

The Semantic Web technologies convey the fix to this issue by making ideas formalized and linking data. The project encodes knowledge on farms in OWL 2 and encodes decision rules in SWRL. The project reuses and realigns the available ontologies: SOSA/SSN (Sensor, Observation, Sample, and Actuator) for observation, ENVO (Environment Ontology) for environment, AGROVOC and CO (Crop Ontology) for crop wordings and attributes.

The combined graph enables dynamic thinking for moves such as delaying fertilization within rain-prone windows or proposing rotations that reduce pest cycles and recover nutrients. **Previous works have largely focused on isolated domains such as soil chemistry or irrigation; this project extends those ideas by integrating multiple datasets — agronomic yield, soil chemistry, and weather records — to provide holistic reasoning for intelligent crop rotation and resource optimization.**

Application Focus: The final outcome is a prototype web application that allows users to query and visualize integrated soil, weather, and crop data through an interactive dashboard. The app provides reasoning-based recommendations on optimal crop rotation sequences, fertilizer timing, and risk mitigation. Key features include knowledge graph exploration, SWRL-driven recommendations, and real-time visualization of environmental trends.

II. Problem Definition

Independent soil, weather, crop, and pest information currently preclude comprehensive, cross-domain decision-making. Most existing approaches focus on single domains, resulting in limited interoperability, low-quality cross-domain inference, and misaligned vocabularies.

The objective of this work is to develop an integrated, reasoning-ready model that:

- Integrates four key dimensions: soil properties, weather signals, crop characteristics, and pest hazards.
- Imports and aligns established ontologies such as SOSA/SSN, ENVO, AGROVOC, and CO to ensure semantic interoperability.
- Employs OWL 2 in combination with SWRL to automatically infer agronomic activities such as crop rotation initiation and fertilization delays based on forecasted rain and soil conditions.
- Provides fertilizer optimization and crop rotation recommendations from a unified knowledge graph to enhance reuse and transparency across systems.

A. Datasets Used

- **Agronomic Yield Dataset (1989–2024):** Annual records of crop type, treatment, and yield (kg/ha) from long-term KBS LTER field plots.
- **Soil Chemistry Dataset (2011–2025):** Laboratory-measured soil properties by plot—pH, nutrients, cation exchange capacity, and organic matter—collected post-harvest.
- **Weather Dataset (1989–2024):** Daily precipitation and temperature data from the Michigan station aggregated into yearly climate summaries.

These datasets together support reasoning over long-term soil fertility trends, yield outcomes, and weather conditions for intelligent agricultural decision-making.

B. Representative Use Cases

1) **Crop Rotation Recommendation:** Infer the next optimal crop to cultivate based on soil nutrient status and historical yield data (e.g., legumes following cereals to restore nitrogen).
2) **Fertilizer Optimization:** Recommend appropriate timing or postponement of fertilizer application using soil nutrient levels and rainfall forecasts.
3) **Pest Risk Mitigation:** Identify potential pest recurrence by linking crop type, soil, and climate data, then suggest rotations that break pest cycles.

III. Approach and High-Level System Design and Implementation

The project follows a structured, phase-based development approach integrating ontology engineering, data processing, reasoning, and visualization within a unified web framework.

A. Phase-wise Development Plan

- **Phase 1 – Ontology Design:** Develop the OWL 2 ontology defining classes like `Plot`, `Crop`, `SoilMeasurement`, and `WeatherSummary`. Align ontology components with SOSA/SSN for observations and ENVO for environmental context.
- **Phase 2 – Data Integration:** Collect and preprocess yield, soil, and weather datasets. Convert key attributes (e.g., year, plot, crop type) into RDF triples to populate the ontology.
- **Phase 3 – Reasoning Implementation:** Encode SWRL rules for crop rotation and fertilizer scheduling based on soil nutrients and rainfall trends.
- **Phase 4 – Application Interface:** Implement a lightweight Flask-based interface that queries the ontology using SPARQL and displays reasoning results.
- **Phase 5 – Testing and Validation:** Evaluate system recommendations for accuracy and ensure reasoning consistency across datasets.

B. System Architecture Overview

The system consists of four logical layers:

- **Data Layer:** Stores raw CSV data for soil, yield, and weather.
- **Ontology Layer:** Hosts the OWL 2 knowledge graph with aligned ontologies (SOSA/SSN, ENVO, AGROVOC and CO).
- **Reasoning Layer:** Executes SWRL rules to infer insights like next crop recommendations and fertilizer scheduling.
- **Application Layer:** Exposes a Flask-based dashboard for user interaction and visualization.

C. Ontology Design Summary

The ontology defines relationships between `Plot`, `Crop`, `Treatment`, and corresponding observation records. Object properties such as `aboutPlot`, `forCrop`, and `withTreatment` interconnect the domains, while datatype properties like `soil_pH`, `soil_N_mgkg`, and `yieldKgPerHa` store empirical data. This design enables inferencing over temporal and contextual agricultural data, improving interoperability and reasoning outcomes.

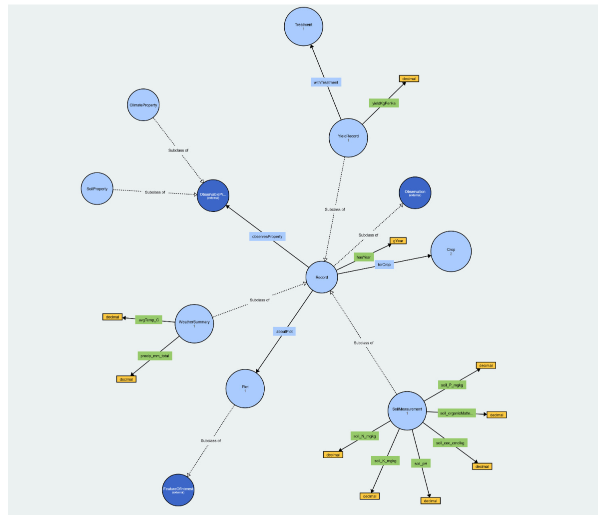

Fig. 1. OWL Ontology Design for Knowledge Graph–based Intelligent Farming System

D. Roles, Responsibilities and Future Scope

- **Arjun Harish:** Dataset Mining and Validation.
- **Dhairya Kamalia:** Consistency checks, finalizing ontology attributes, generating instances.
- **Kavya Parekh:** Coming up with object and datatype properties.
- **Raj Mani Mitresh Mani:** System architecture design, Sanity checks, Report writing.
- **Samanvitha Patel:** Designing OWL 2 Architecture.

IV. Data Collection and Processing

A. Data Sources

Three primary datasets are utilized:

1) **KBS LTER Agronomic Yield Dataset (1989–2024):** Contains yearly yield records for different crops and treatments. [13]
2) **KBS Soil Chemistry Dataset (2011–2025):** Provides laboratory measurements of soil pH, nutrients, and organic matter by plot. [14]
3) **Michigan Weather Dataset (1989–2024):** Daily and monthly summaries of precipitation and temperature for the KBS site area, derived from the Global Historical Climatology Network-Daily (GHCN-D) dataset [15]

B. Data Preparation and Processing

Data preprocessing involves cleaning and transforming the raw files:

- Handling missing or incomplete values through interpolation.
- Standardizing measurement units across datasets.
- Mapping crop and soil attribute names to AGROVOC and CO terms.
- Aggregating annual summaries by plot and year for uniform reasoning.

After cleaning, the data is converted into RDF format using Python's RDFLib and imported into the OWL ontology. Each data record becomes an individual instance (e.g., `YieldRecord_2024_Corn`), linked through object and datatype properties.

C. Scope and Exclusions

Pest-specific and irrigation datasets are excluded in this phase due to incomplete data availability. However, pest risks are indirectly modeled through soil and weather parameters. Future work can integrate direct pest observations.

V. Related Literature

Semantic Web technologies are now being used in agriculture to connect data from sensors, weather systems, and farm operations to support better decision-making. Agriculture generates vast amounts of data about soil, weather, crops, and pests, but this data often exists in separate systems that do not communicate with one another. Earlier research has developed smaller ontologies focused on individual areas, such as soil, irrigation, or pest management. This project aims to build on those foundations and combine them into a single, integrated reasoning framework that brings together multiple aspects of farming.

Haller *et al.* [1] and W3C/OGC [11] introduced the SOSA/SSN ontology, which provides a standard way to represent sensors, observations, and environmental features. This framework is used in this project to describe soil and weather data consistently, allowing observations from different devices to be compared and reasoned over. It forms the foundation for representing sensor-based data.

AGROVOC, developed by Subirats-Coll *et al.* [2] and Caracciolo *et al.* [3], is a multilingual vocabulary maintained by the Food and Agriculture Organization (FAO). It ensures consistent labeling of agricultural terms such as crops, fertilizers, and farming practices. In this work, AGROVOC terms are used to tag and interlink all the major agricultural concepts, which helps maintain consistency and promotes knowledge sharing across systems.

Arnaud *et al.* [4] and Matteis *et al.* [5] created the Crop Ontology (CO) to standardize crop traits and variables used in research and breeding programs. Incorporating CO allows the ontology to represent traits such as nutrient needs, growth cycles, and rotation patterns. This supports reasoning about relationships like how legumes help improve nitrogen levels or how cereals should be rotated to reduce pest risks.

The Environment Ontology (ENVO) [10] provides definitions for soil and land-use types that describe the environmental context of farms. Using ENVO allows this ontology to connect farming activities with environmental data, supporting spatial and ecological reasoning without the need to invent new soil classifications.

Kessler *et al.* [6] developed an ontology-based decision-support system for nitrogen fertilization using OWL and SWRL reasoning. Their work demonstrated that logical rules could turn soil test data into fertilizer recommendations. The proposed ontology follows this reasoning approach but extends it beyond nitrogen to include other nutrients such as phosphorus and potassium. It also combines soil, weather, and pest data to produce more adaptive recommendations.

Alharbi *et al.* [7] designed an ontology-based pest management system that uses environmental and biological indicators to assess pest risks. While their work focuses on detection and diagnosis, this project expands on it by connecting pest risk to soil conditions and crop types. This enables preventive decision-making, such as adjusting fertilizer schedules or changing crop rotation to minimize future infestations.

Jonquet *et al.* [8] presented AgroPortal, a repository for agricultural ontologies that promotes reuse and alignment between different models. Aligning the current ontology with AgroPortal ensures that it can interoperate with other ontologies and remain accessible to the wider research community.

Finally, de Vaulx [9] and ETSI [12] introduced SAREF4AGRI, which connects sensors, actuators, and decision systems for agriculture. Their work focuses on modularity and device integration. This ontology builds on that approach, integrating real-time observation data with reasoning-based recommendations for fertilizer and crop management.

Together, these studies establish a strong base for semantic modeling in agriculture, but each focuses on a specific area. This project takes a broader approach. It integrates soil properties, weather data, crop characteristics, and pest risks into one framework that can reason over real-world data. By combining established ontologies like SOSA/SSN, AGROVOC, CO and ENVO, the model enables both fertilizer optimization and crop rotation planning, offering a more complete view of smart farming.

References

[1] A. Haller, K. Janowicz, S. R. Krisnadhi, S. Cox, D. Le Phuoc, K. Taylor, and M. Hert, "The SOSA/SSN ontology: A joint W3C and OGC standard for sensor data on the Web," *Semantic Web Journal*, vol. 10, no. 1, pp. 9–32, 2019.

[2] . Subirats-Coll, C. Caracciolo, J. E. Keizer, and A. Stadler, "AGROVOC: The Linked Data concept hub for food and agriculture," *Computers and Electronics in Agriculture*, vol. 194, p. 106736, 2022.

[3] C. Caracciolo, A. Stadler, and J. Keizer, "The AGROVOC Linked Dataset," *Semantic Web Journal*, vol. 4, no. 3, pp. 341–348, 2013.

[4] E. Arnaud *et al.*, "The Crop Ontology: A source of standard traits and variables for crop data integration," *Int. J. Plant Genomics*, 2016.

[5] L. Matteis, E. Arnaud, and M. Bonaiuti, "Crop Ontology: Vocabulary for crop-related concepts," *Computers and Electronics in Agriculture*, vol. 96, pp. 27–34, 2013.

[6] . Kessler, A. Perzylo, and M. Rickert, "Ontology-based decision support for the nitrogen fertilization of winter wheat," in *Proc. Metadata and Semantic Research (MTSR 2020)*, CCIS 1355. Cham: Springer, 2021, pp. 259–270.

[7] A. F. Alharbi, A. A. Alanazi, and A. A. Almutairi, "An ontology-based agriculture decision-support system with an evidence-based explanation model," *Smart Agricultural Technology*, vol. 11, p. 100351, 2024.

[8] C. Jonquet, A. Toulet, C. Arnaud, N. Pesce, P. Larmande, and M. P. López, "AgroPortal: A vocabulary and ontology repository for agronomy," *Computers and Electronics in Agriculture*, vol. 144, pp. 126–143, 2018.

[9] F. de Vaulx, "Ontological requirement specification for smart irrigation systems," in *Proc. CEUR Workshop on Ontologies for Intelligent Systems*, 2018.

[10] R. G. Meech *et al.*, "Environment Ontology (ENVO): Classes for soil and land-use representation," *J. Biomed. Semantics*, vol. 12, no. 1, 2021.

[11] W3C/OGC, *Semantic Sensor Network Ontology (SSN/SOSA) Recommendation*, 2017. [Online]. Available: https://www.w3.org/TR/vocab-ssn/

[12] European Telecommunications Standards Institute (ETSI), *SAREF4AGRI: Extension of the Smart Applications REFerence ontology for agriculture and food*, ETSI TR 103 410-2, 2018.

[13] Kellogg Biological Station Long-Term Ecological Research (KBS LTER). (2024). *Agronomic yields of annual crops* [Data set]. Michigan State University. Available: https://lter.kbs.msu.edu/datatables/51

[14] Kellogg Biological Station Long-Term Ecological Research (KBS LTER). (2025). *Agronomic soil chemistry* [Data set]. Michigan State University. Available: https://lter.kbs.msu.edu/datatables/52

[15] Menne, M. J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R. S., Gleason, B. E., & Houston, T. (2012). *Global Historical Climatology Network – Daily (GHCN-D), Version 3* [Data set]. NOAA National Climatic Data Center. Available: https://doi.org/10.7289/V5D21VHZ