Artificial Intelligence Research and Development
K. Trejo et al. (Eds.)

doi:10.3233/FAIA250589

Agrixel. AI-Based Service Standard for Agricultural Data in Agri-Food Use Cases

Joan VAN EECKHOUT [a,1], Miona DIMIC [a] and Cecilio ANGULO [a,b]
[a] *Intelligent Data Science and Artificial Intelligence Research Centre (IDEAI), Universitat Politècnica de Catalunya, Barcelona, Spain*
[b] *Institut de Robòtica i Informàtica Industrial (IRI), Barcelona, Spain*
ORCiD ID: Joan van Eeckhout https://orcid.org/0000-0002-2640-0454, Miona Dimic https://orcid.org/0009-0008-1869-3585, Cecilio Angulo https://orcid.org/0000-0001-9589-8199

Abstract. We introduce a new interoperable geographic standard called *agrixel*, which employs a unique and innovative data processing methodology—similar to pixels in computer vision or tokens in language processing—leveraging computational power and machine learning in shared data spaces. Focused on AI-driven services for the agri-food sector and founded in a defined ontology, it offers a service portfolio for the sector itself, its externalities, and the circularity of its products.

The *agrixel* standard ensures findable, accessible, interoperable, and reusable (FAIR) data, compatible with existing solutions. It supports multiple data layers, enabling extensions to externalities, such as carbon footprint, energy use, water management, logistics, and transport—enhancing efficiency in agri-food production and consumption.

Keywords. Ontology, Semantic Interoperability, Agricultural Information Model, Adaptive Context-Aware, Decision Support System, Precision Agriculture

1. Introduction

In agriculture, effective decision-making requires understanding the complex interplay between environmental factors-such as weather, soil conditions, and topography-and human interventions, including irrigation strategies, crop selection, and land management practices [1]. While many of these variables, like soil type and climate variability, lie beyond human control, others can be optimized to improve productivity and resource efficiency. Optimization efforts are increasingly challenged by the effects of climate change, which introduces greater variability in temperature and precipitation patterns. Simultaneously, consumer demand for transparency, sustainability, and accountability is growing, particularly with regard to the carbon and water footprints of agricultural production.

To address these complexities, the integration of diverse data sources - ranging from satellite imagery and weather forecasts to in-situ sensor readings and field observations

[1]Corresponding Author: Joan van Eeckhout, Intelligent Data Science and Artificial Intelligence Research Center, Universitat Politècnica de Catalunya, Nexus II Building, Jordi Girona 29, 08034 Barcelona, Spain. Universitat Politècnica de Catalunya · UPC BarcelonaTECH. joan.van.eeckhout@upc.edu

- has become essential for environmental monitoring, risk prediction, and precision agriculture [2]. However, these data sources are often heterogeneous in format, spatial resolution, and temporal frequency, posing significant challenges for unified analysis and scalable AI applications.

Over the past decade, research on smart farming and the application of AI techniques in agriculture has grown exponentially [3]. A large proportion of this work addresses tasks such as weed detection, disease diagnosis, water resource management, and crop classification. However, several persistent challenges continue to limit the scalability and robustness of these approaches. One of the main obstacles is the lack of standardization in data collection and representation. Agricultural systems are inherently diverse, encompassing different farming practices, crop varieties, and environmental conditions. This variability complicates the development of consistent data and hinders the transferability of AI models across regions. In addition, the availability of high-quality labelled training datasets remains limited. Supervised learning methods require substantial volumes of annotated data to achieve reliable performance, yet such datasets are often insufficient, especially for niche crops and farming practices.

To overcome these challenges, we propose the *agrixel* standard—an interoperable geographic framework designed to integrate high-resolution, multi-layered data across the agri-food sector and its externalities. Its foundation is the agrixel: a $1\text{m} \times 1\text{m}$ geospatial semantic unit that serves as the anchor for organizing multiple data layers at precise geographic coordinates. These layers include environmental conditions, agricultural practices, resource management, crop health, production outputs, and market dynamics. The layers are semantically linked via the agrixel ontology, which structures and harmonizes data around each geospatial concept. Building upon the DEMETER Agriculture Information Model (AIM) [4] and adhering to Open Geospatial Consortium (OGC) guidelines, this ontology ensures interoperability with existing standards and systems.

Aligned with FAIR data principles, the agrixel standard enables scalable, cross-sector data integration. Like pixels in computer vision or tokens in natural language processing (NLP), the agrixel acts as a fundamental analytical unit. Its high-dimensional structure allows efficient processing of agricultural data by machine learning algorithms, providing a robust foundation for AI-driven agricultural services.

The rest of the manuscript is organized as follows: Firstly, the agrixel concept is introduced as a standard for data interoperability. Next, the core ontology of the agrixel standard is outlined. Then, the prototype implementation using software tools is explained. Conclusions and future work conclude the manuscript.

2. Agrixel Design

Georeferenced agricultural data are heterogeneous and fragmented [5] across domains, sources, formats, and spatio-temporal resolutions. Raster imagery (e.g., from satellites and drones), vector layers (e.g., administrative and cadastral boundaries), tabular datasets (e.g., crop yields and treatments), IoT sensor time series, soil data (from lab or probes), and unstructured documents (e.g., reports and declarations) coexist—often lacking a shared structure, resolution, or semantics. This fragmentation is a drawback for automated data integration, semantic alignment, and the effective use of advanced tools like AI or decision support systems. The core challenge, therefore, lies in reconciling three key dimensions.

- Spatial resolution. Datasets range from sub-meter drone imagery to administrative units defined at the hectare or plot scale.
- Temporal resolution. Observations range from hourly sensor readings to crop related annual declarations or static soil profile probes.
- Semantic alignment. Terminologies, classifications, and data structures differ between sources and domains.

Existing semantic standards—such as W3C Time[2], OGC GeoSPARQL[3], QUDT[4], and SSN/SOSA[5]—provide robust foundations for modeling time, space, and quantitative observations. Domain-specific vocabularies, meanwhile, are offered by agricultural ontologies like FOODIE[6] [6], SAREF4Agri[7], and FIWARE-Agrifood[8]. The AIM [4], developed in the DEMETER project, unifies many of these into a single interoperable framework. However, while these models enable semantic alignment, they lack mechanisms for fine-resolution spatio-temporal data integration and the specialized vocabulary required to represent precision field-level concepts.

To address these limitations, we propose the agrixel—a standardized geospatial semantic $1\,\text{m} \times 1\,\text{m}$ unit. Each agrixel serves as a fixed-resolution spatial anchor, comparable to a pixel in computer vision or a token in NLP, enabling that all agricultural information can be projected, timestamped, and semantically contextualized. This atomic unit facilitates:

- Fine-grained spatial anchoring. Each agrixel is a square polygon of $1\,\text{m} \times 1\,\text{m}$ in a fixed Coordinate Reference System (Web Mercator EPSG:3857 WGS 84/Pseudo-Mercator), aligned with the spatial resolution of in-field sensors, drone imagery, and precision machinery.
- Temporal versioning. Data linked to agrixels can be organized into daily snapshots, supporting temporal aggregation or disaggregation as needed.
- Semantic enrichment. Agrixels are modeled with an ontology that reuses and extends AIM, SSN/SOSA, O&M[9], QUDT, GeoSPARQL, and agricultural vocabularies to ensure semantic consistency and interoperability.
- Metadata-rich structure compliant with ISO 19115-1 and interoperable via SpatioTemporal Asset Catalog (STAC) specifications.

The 1 m resolution is deliberately chosen to support both bottom-up precision and top-down integration:

- Allows direct association with high-resolution data sources (e.g. drone NDVI, soil sensors, localized treatments).
- Enables the integration of medium- or low-resolution datasets (e.g. $10\,\text{m} \times 10\,\text{m}$ Sentinel-2, $1\,\text{km}^2$ ERA5) via interpolation or zonal statistics.
- Supports downscaling of coarse data with statistical or ML-based models to infer fine-grained distributions.

[2] https://www.w3.org/TR/owl-time/
[3] https://www.ogc.org/standards/geosparql/
[4] https://www.qudt.org/
[5] https://www.w3.org/TR/vocab-ssn/
[6] https://foodie-cloud.github.io/model/FOODIE.html
[7] https://saref.etsi.org/saref4agri/v2.1.1/
[8] https://github.com/smart-data-models/dataModel.Agrifood
[9] https://www.ogc.org/standards/om/

- Preserves spatial structure, enabling spatially aware AI workflows that rely on proximity, gradients, or field heterogeneity.

Agrixels act as semantic bridges: every piece of agricultural information—whether a crop type, irrigation event, NDVI value, or harvest record—is contextualized by its agrixel. This facilitates consistent querying, explainability, and cross-modal AI integration. The result is a high-resolution, interoperable, and explainable geospatial data fabric ready to power agricultural reasoning, machine learning pipelines, and decision support systems.

The agrixel semantic standard ensures interoperability by formalizing how diverse datasets link to this geospatial framework. It provides:

- A shared data structure for cross-platform agricultural tools and services.
- Flexible integration for diverse data sources and vocabularies.
- Semantic relationships enabling consistent data interpretation and reasoning.

While the agrixel unit itself is our core contribution, we also propose a set of extensions to existing semantic models to better represent field-level agricultural concepts that are often missing from generic ontologies—such as cultivar-rootstock combinations, soil compaction, or irrigation modes. These refinements further enhance the semantic resolution of agrixel-based systems, making them suitable for the high specificity required in AI-driven precision agriculture.

The next section presents the agrixel core ontology, formalizing relationships between key concepts to enable consistent data interpretation and integration.

3. Agrixel Ontology

The Agrixel ontology serves as the foundational conceptual model for the agrixel standard, positioning it as the central component for agricultural data integration. It defines and links diverse entities representing concepts derived from heterogeneous data sources. The ontology enables spatial disaggregation of data from coarser resolutions down to the standardized $1\,\text{m} \times 1\,\text{m}$ agrixel level. This granularity preserves traceable aggregation relationships with broader data sources while enabling detailed analysis.

The ontology organizes agrixel units through three interconnected dimensions, together enabling their comprehensive characterization.

- Spatial. This dimension captures both the agrixel's geometric representation and its position within a spatial hierarchy (e.g., parcels, regions), enabling interoperability across scales and facilitating multi-level data integration.
- Observational. The agrixel standard is defined by multiple attributes derived from location-specific data. This dimension encompasses all system observations, both qualitative and quantitative, that affect the agrixel units.
- Temporal. The temporal dimension tracks the agrixel's lifecycle, recording key events such as its creation, updates, and changes in its conditions over time. With a fixed minimum resolution, it ensures consistent updating and precise tracing of the agrixel units evolution.

The Agrixel ontology is built upon the AIM developed by the DEMETER project. It incorporates both cross-domain and domain-specific ontologies to ensure semantic alignment and interoperability with established standards. The AIM scope is expanded by integrating supplementary ontologies that support diverse agricultural domains, leveraging existing concepts to represent domain-specific knowledge. Furthermore, the ontology formally defines the agrixel entity along with other ontology-specific classes and properties, to link domain-specific information to agrixel units.

The Agrixel ontology integrates data from Spain's agricultural administration, specifically aligning with SIGPAC[10] (GIS for Agricultural Lands). Hence, SIGPAC is essential for accurately linking agricultural activity with land use. The ontology also incorporates administrative data from SEC[11](Electronic Cadastre Office) and DUN[12] (Farm Declaration), ensuring compatibility with official cadastral records and annual farm declarations.

It reuses the SAREF4AGRI class `Farm` and defines a cadastre-specific class `Parcel` to represent administrative data in a spatial context. The SAREF4AGRI `Parcel` class was not adopted because its definition does not strictly align with that of a cadastral `Parcel`. Moreover, its equivalence to more general concepts like `Plot` introduces ambiguity, as a `Plot` does not necessarily have any relation to the cadastre [7]. By reusing the relationships `contains` and `containedIn` between GeoSPARQL spatial objects, a `Farm` can contain multiple `Parcels`, and each `Parcel` is contained in a single `Farm`.

The ontology also introduces the `Agrixel` class to represent $1\,\mathrm{m} \times 1\,\mathrm{m}$ units derived from dividing cadastral-referenced `Parcels`: each `Agrixel` is contained within a `Parcel`, and each `Parcel` contains its associated `Agrixel` units. This structure enables detailed spatial representation while supporting aggregation to higher administrative units, such as `Parcels` and `Farms`.

To further align with SIGPAC, two additional classes are introduced: `ParcelUnit` and `CropArea`. Both are GeoSPARQL spatial objects linked to their containing `Parcel` (see Figure 1). A `ParcelUnit` refers to a stable, continuous area within a `Parcel`, characterized by a unique identifier and a single land use. A `CropArea` represents the declared crop surface for a specific agricultural campaign; it is always located within a `Parcel`, but may not align with `ParcelUnit` boundaries. To accommodate this, the ontology introduces the `intersects` relationship between `CropArea` and `ParcelUnit`, enabling the identification of crop declarations within specific units without enforcing strict hierarchical containment.

The same modeling approach, based on spatial intersections, is applied for `Agrixel` and `CropArea` classes to integrate crop-related properties into the finer geospatial units represented by `Agrixel`. As illustrated in Figure 2, the `CropArea` class uses the object property `hasAgriCrop`, defined in the FIWARE standard, to link a spatial object to its associated crop type, represented by the `AgriCrop` class. This class reuses object properties from AIM and FIWARE to associate crop types with reference vocabularies such as AGROVOC [8] and EPPO[13].

The Agrixel ontology introduces two additional properties—`upovConcept` and `siexConcept`—which enable linkage to crop type definitions in external catalogs main-

[10]https://www.mapa.gob.es/es/agricultura/temas/sistema-de-informacion-geografica-de-parcelas-agricolas-sigpac-
[11]https://www.sedecatastro.gob.es/
[12]https://agricultura.gencat.cat/ca/ambits/desenvolupament-rural/declaracio-unica-agraria/
[13]https://www.eppo.int/

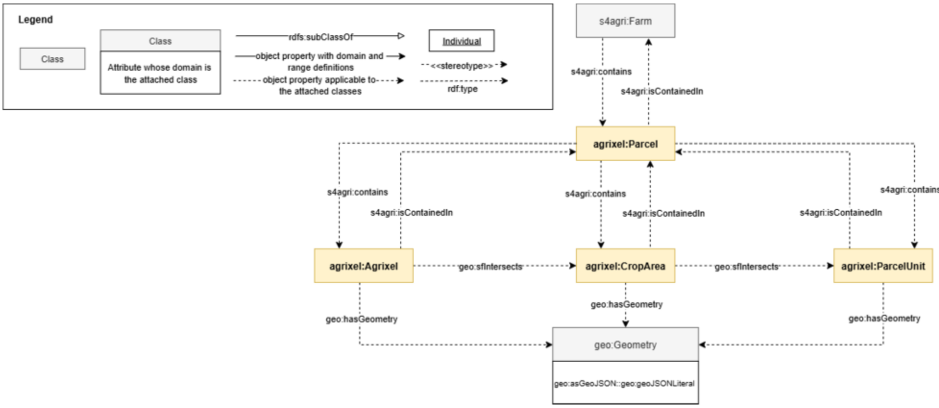

Figure 1. Illustration of the classes and object properties, showing the relationships between Farm, Parcel, Agrixel, ParcelUnit, and CropArea.

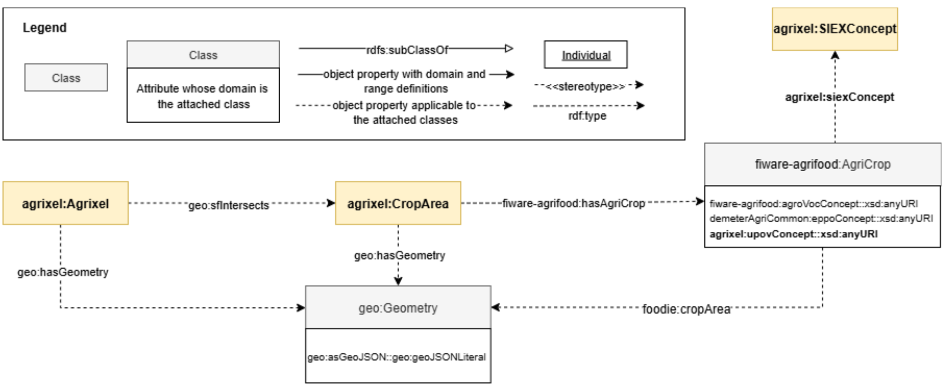

Figure 2. Illustration of the classes and object properties, showing the relationships between Agrixel, CropArea, and AgriCrop.

tained by UPOV[14] and SIEX[15] (Agricultural Holdings Information System). By incorporating class `SIEXConcept`, the ontology establishes a direct link to the relevant SIEX catalog record, commonly used in Spanish agricultural information systems. As a result, each `Agrixel` is associated with a single crop type, consistently described across multiple vocabularies through its connection to a `CropArea` and the corresponding `AgriCrop` instance. Additionally, the `cropArea` object property, defined in the FOODIE ontology, allows to group all geometries that share the same crop type, providing a structure to organize spatial units by crop classification.

In the SOSA ontology, an `Observation` represents the act of carrying out a procedure to estimate or determine the value of a property for a specific `FeatureOfInterest`. To enable observations at the agrixel level, the Agrixel class is defined (see Figure 3) as `sosa:FeatureOfInterest`. This allows associating each agrixel instance with measured variables like soil texture or temperature. By modeling `Agrixel` in this way, the

[14]https://www.upov.int/portal/index.html.en
[15]https://www.fega.gob.es/es/siex1

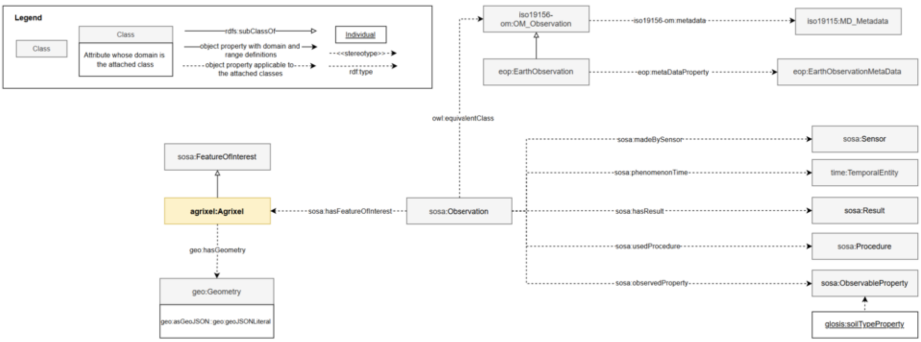

Figure 3. Illustration of Observation made on Agrixel as the Feature of Interest.

ontology enables direct semantic and spatial linking of sensor data to fine-grained agricultural units.

The Agrixel ontology extends AIM [4] by integrating additional ontologies to enhance agrixel's description with a broader range of observable properties. Observable properties are drawn from existing standards that define individuals, which can be treated as `sosa:ObservableProperty`.

Hence, from SAREF4AGRI, observable properties like `IrrigationWater`, `Yield`, and `Precipitation` are agricultural attributes that can be linked to Agrixel. To support a broader range of quantitative observations, the Agrixel ontology imports the QUDT Quantity Kinds vocabulary[16], enabling the use of standardized quantity kinds for properties such as temperature, humidity, and pressure. Furthermore, it incorporates the GLOSIS ontology [9], which provides domain-specific descriptors for soil characteristics, including `textureFieldClassProperty`, `physioChemicalPropertyCode-pH`, and `moistureContentProperty`. This modular integration facilitates extensibility, allowing the ontology to accommodate additional observable properties and domain semantics as required.

To support the integration of Earth Observation (EO) and other raster-based data sources, the Agrixel ontology also models gridded observations derived from satellite platforms (e.g., ERA5, Sentinel-2) and high-resolution campaigns using UAVs. All raster datasets are converted into a set of individual `sosa:Observation` instances, where each value of a cell in the raster corresponds to a distinct observation, geolocated at the centroid of the respective cell. To manage these large collections of observations, we adopt the ssn-ext[17] ontology, which extends SOSA/SSN with support for `ssn-ext:ObservationCollection`. Each raster file is represented as an observation collection that groups all observations it contains. These collections may also be subdivided according to common characteristics as `sosa:observedProperty`, `sosa:madeBySensor`, or `sosa:phenomenonTime`. This hierarchical structure facilitates efficient querying, storage, and reasoning over EO data.

Furthermore, the metadata of each observation can be explicitly modeled using the `om:metadata` property from the O&M standard. Observations may refer to in-

[16] http://www.qudt.org/doc/2025/06/DOC_VOCAB-QUANTITY-KINDS-ALL.html
[17] https://www.w3.org/TR/vocab-ssn-ext/

stances of `iso19115:MD_Metadata`, which provide standardized geographic metadata. In the case of `eop:EarthObservation`, metadata can also be represented using `eop:EarthObservationMetaData`, linked via `eop:metaDataProperty`, in accordance with the OGC Earth Observation Metadata profile[18].

Although the STAC specification is widely adopted to describe EO datasets, it is not yet a formal OGC or ISO standard[19]. Therefore, while we adopt the practical and flexible STAC model to expose and query data assets, we align its content with the ISO19115 and the OGC Earth Observation Metadata profile to ensure compatibility with existing geospatial standards. This hybrid approach leverages the strengths of both models—STAC's modern ecosystem and O&M's rigor—for robust metadata management.

4. Prototype Implementation

The Agrixel ontology is implemented in Protégé[20], where the necessary reference ontologies are imported, and domain-specific classes, object properties, and their logical relationships are defined. Each data source is individually processed to identify concepts that match those defined in the ontology. Once the relevant mappings are established, tools such as YARRRML[21] and RDFLib[22] are used to map the data to the ontology. These tools instantiate individuals of the defined classes and establish relationships between them, generating RDF triples based on the ontology. This process enables the transformation of heterogeneous real-world data into a semantically structured knowledge graph.

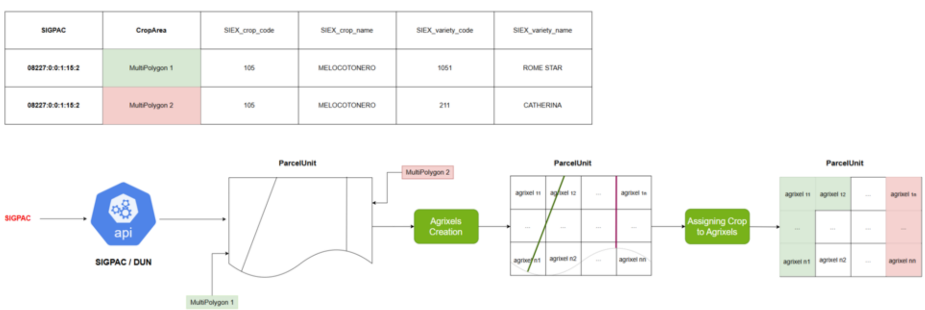

Figure 4. Illustration of mapping public administrative data from DUN to *agrixel* standard objects.

Figure 4 and Figure 5 illustrate the transformation of public datasets into semantically aligned instances within the ontology. Figure 4 shows the mapping of administrative parcel units and crop data from the DUN dataset to *agrixel* classes such as `agrixel:CropArea` and `agrixel:ParcelUnit`, each linked to its respective geometry. Instances of `agrixel:Agrixel` are created and, by means of the `geo:sfIntersects` property, the agrixels are linked with the `agrixel:CropArea` and

[18] http://docs.opengeospatial.org/is/10-157r4/10-157r4.html
[19] https://www.ogc.org/requests/ogc-looks-to-adopt-stac-as-community-standard-seeking-public-comment-before-moving-to-vote/
[20] https://protege.stanford.edu/
[21] https://rml.io/yarrrml/
[22] https://rdflib.readthedocs.io/en/stable/

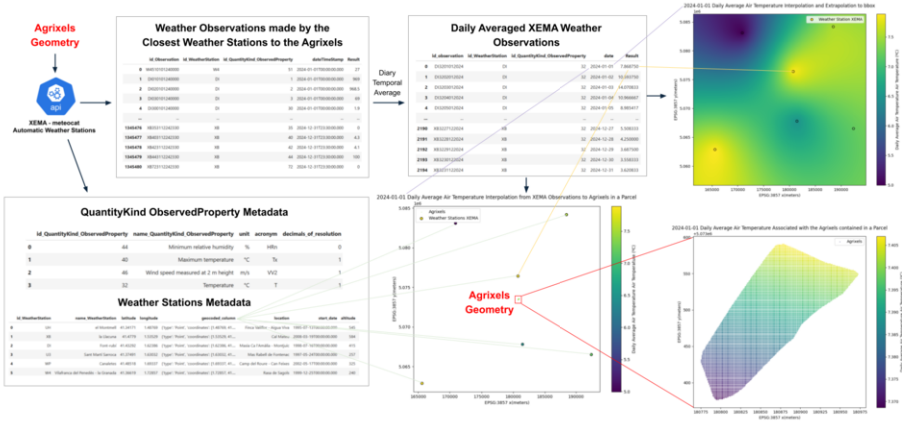

Figure 5. Illustration of mapping public weather data from XEMA-METEOCAT to *agrixel* standard objects.

`agrixel:ParcelUnit` objects. This spatial relationship infers a semantic link between the administrative attributes and crop properties.

Figure 5 demonstrates the integration of weather station data from the automatic weather stations network (XEMA) of the Catalan meteorological service (METEOCAT). The XEMA observations are recorded at 30-minute intervals and are initially mapped as `sosa:Observation` instances linked to the corresponding meteorological stations as `sosa:madeBySensor`. A `sosa:Procedure` representing a daily averaging operation is then applied to derive new observations with aggregated values. Subsequently, a spatial interpolation procedure (e.g. kriging or inverse distance weighting) is applied to generate `sosa:Observation` instances centered on each agrixel. Interpolated observations are linked to `agrixel:Agrixel` instances via `sosa:hasFeatureOfInterest`, thus enabling spatiotemporal alignment between meteorological variables and agrixel-level entities defined in the ontology.

5. Conclusions and Future Work

The agrixel standard presents a unified spatial unit for agricultural data, enabling the integration of diverse data sources across various agricultural domains. By aligning with the FAIR principles, the proposed standard ensures that agricultural data is easily discoverable, accessible, and reusable, thereby promoting collaboration and facilitating more informed decision-making. This standardization provides a cohesive structure for agricultural data, which supports enhanced analysis and more effective management of agricultural systems. Moreover, the standard and its associated ontology are being co-developed with private and public sector partners.

Future work should focus on three priorities: (i) refining and standardizing the agrixel ontology to enable broad adoption and interoperability; (ii) developing large-scale benchmark datasets for deeper agricultural insights; and (iii) implementing the standard in real-world AI systems to demonstrate its optimization potential. Moreover, integrating the *agrixel* standard into international governance frameworks, such as GAIA-

X[23] and the International Data Spaces Association[24] (IDSA), will ensure its alignment with global data-sharing standards and facilitate seamless data exchange on an international scale.

Acknowledgements

Joan van Eeckhout and Miona Dimic acknowledge the support of the Spanish Ministry for Digital Transformation and Public Service through the Secretariat of State for Digitalisation and Artificial Intelligence (project Agrixel_ES, TSI-100121-2024-019) as part of the "Plan to Promote Sectoral Data Spaces" under the Recovery, Transformation and Resilience Plan, financed by the European Union – NextGenerationEU. Prof. Cecilio Angulo is also acknowledging co-funds from the European Union – NextGenerationEU, through the I+D+i PID2021-122835OBC21 research project ARTIFACTS (MCIN/AEI/10.13039/501000011033) and FEDER "Una manera de hacer Europa".

References

[1] Zhai Z, Martínez JF, Beltran V, Martínez NL. Decision support systems for agriculture 4.0: Survey and challenges. Computers and Electronics in Agriculture. 2020;170:105256. Available from: https://www.sciencedirect.com/science/article/pii/S0168169919316497.

[2] Sundmaeker H, Verdouw C, Wolfert S, Pérez-Freire L. Internet of Food and Farm 2020. In: Digitising the Industry Internet of Things Connecting the Physical, Digital and VirtualWorlds. River Publishers; 2016. p. 129-51.

[3] Attri I, Awasthi LK, Sharma TP. Machine learning in agriculture: a review of crop management applications. Multimedia Tools and Applications. 2024;83(5):12875-915.

[4] Palma R, Roussaki I, Döhmen T, Atkinson R, Brahma S, Lange C, et al. Agricultural Information Model. In: Information and Communication Technologies for Agriculture—Theme III: Decision. Springer; 2022. p. 3-36. Available from: https://EconPapers.repec.org/RePEc:spr:spochp:978-3-030-84152-2_1.

[5] Gardeazabal A, Lunt T, Jahn MM, Verhulst N, Hellin J, and BG. Knowledge management for innovation in agri-food systems: a conceptual framework. Knowledge Management Research & Practice. 2023;21(2):303-15. Available from: https://doi.org/10.1080/14778238.2021.1884010.

[6] Palma R, Reznik T, Esbrí M, Charvat K, Mazurek C. An INSPIRE-based vocabulary for the publication of Agricultural Linked Data. In: Ontology Engineering: 12th International Experiences and Directions Workshop on OWL, OWLED 2015, co-located with ISWC 2015, Bethlehem, PA, USA, October 9-10, 2015, Revised Selected Papers 12. Springer; 2016. p. 124-33.

[7] Řezník T, Lukas V, Charvat K, Horakova S, Kepka M. Foodie data models for precision agriculture. In: Proceedings of the 13th International Conference on Precision Agriculture; 2016. p. 1-17.

[8] Caracciolo C, Stellato A, Morshed A, Johannsen G, Rajbhandari S, Jaques Y, et al. The AGROVOC linked dataset. Semantic Web. 2013;4(3):341-8.

[9] Palma R, Janiak B, de Sousa LM, Schleidt K, Řezník T, van Egmond F, et al.. GloSIS: The Global Soil Information System Web Ontology. arXiv; 2024. Available from: https://arxiv.org/abs/2403.16778.

[23] https://gaia-x.eu/

[24] https://internationaldataspaces.org/