Interoperable Agritech Data Pipelines with NGSI-LD and Smart Data Models

Rustem Dautov and Simeon Tverdal
SINTEF
Oslo, Norway *rustem.dautov@sintef.no, simeon.tverdal@sintef.no*

André Skoog Bondevik and Svein Arild Frøshaug
Agdir Drift
Arendal, Norway *asb@agdir.no, saf@agdir.no*

Vera Szabo and Jan Robert Fiksdal
Aersea
Kristiansand, Norway *vera.szabo@aersea.com,jan.robert.fiksdal@aersea.com*

Abstract

The increasing use of drones, robotic platforms, and IoT sensors in agriculture has resulted in a growing volume of heterogeneous data that is difficult to integrate due to lack of interoperability. This paper presents three data pipelines designed within the Norwegian research project SMARAGD, targeting the transformation of siloed agritech data into interoperable NGSI-LD-compliant entities using Smart Data Models and the FIWARE framework. The pipelines cover aerial imagery, robotic imagery from ROS-based systems, and IoT sensor measurements, enriching the data with temporal and geospatial context and integrating it into a shared FIWARE-powered ecosystem. This architecture provides a foundation for decision-support tools and interoperability in land-based food production systems.

Keywords: Data Interoperability, Smart Agriculture, FIWARE, Smart Data Models, NGSI-LD

1. Introduction and Motivation

The digital transformation of the agrifood sector presents opportunities for more sustainable, efficient, and data-driven agricultural practices. Aerial imagery from drones, close-up observations from robots, and continuous measurements from IoT sensors are becoming standard tools in modern farming. Yet, these technologies often operate in isolation, generating data in incompatible formats, stored in disconnected systems. This fragmentation severely limits the ability to combine and analyse data across sources – an essential capability for delivering actionable insights and supporting precision agriculture.

This paper addresses these challenges by designing and implementing three data pipelines that transform raw agritech data into interoperable information assets. Developed within the Norwegian research project SMARAGD (Smart Agriculture Data Fusion for Decision Support) [3], the pipelines target three key data domains: (1) UAV-based aerial imagery, (2) robotic imagery collected via the Robot Operating System (ROS), and (3) environmental measurements from IoT sensors. All pipelines convert their respective data into NGSI-LD-compliant entities aligned with Smart Data Models, converging into a shared FIWARE-based data ecosystem.

The contributions of this work are twofold. First, we propose and implement three semantically enriched pipelines that align heterogeneous agritech data with recognised interoperability standards. Second, we introduce two domain-specific Smart Data Models for aerial and robotic imagery to support integration into the NGSI-LD framework.

A common motivation across all three pipelines is the urgent need to overcome siloed data systems in current agritech practice. For instance, to optimise fertiliser application, a farmer may need to combine data from drone imagery, robotic cameras, and soil sensors. However, each of these systems stores data differently, requiring manual harmonisation, typically involving spreadsheet exports, ad hoc processing, and spatial or temporal alignment by hand. This approach is time-consuming, error-prone, and ultimately limits the scalability and utility of digital farming solutions.

Fragmented Entry Points Across Data Domains

Aerial imagery: Drone-based imagery is often a by-product of other services such as spraying or mapping. Due to constraints like bandwidth, battery life, or terrain, the captured images are typically offloaded after flight, either manually via SD cards or automatically at docking stations. While drone APIs tend to be well-documented, the variety of onboard camera types (RGB, multispectral, LIDAR, or thermal) introduces complexity. These cameras often embed metadata using proprietary formats (e.g., EXIF or XMP), creating barriers for integration with external tools or platforms.
Robotic imagery: Agricultural robots are increasingly used for harvesting, weeding, and spraying. These platforms generate large volumes of image and sensor data stored in ROS bag files, timestamped and synchronised across multiple sensor streams. Although ROS is widely adopted within the robotics community, its internal formats (e.g., message types and topic structures) are not natively understood by external data systems. This limits the reuse of robotic data unless pre-processed and reformatted to align with general-purpose standards.
IoT measurements: Environmental monitoring with IoT devices is critical for soil and crop health assessment. However, no single manufacturer covers the full spectrum of use cases from tunnels to open fields, and each vendor typically applies their own proprietary data models and APIs. As a result, farmers often install devices from multiple suppliers, leading to fragmented datasets and incompatible dashboards. Metadata about location, units, sensor types, and measurement semantics is rarely standardised, impeding automated data exchange and fusion.

Need for a Unified and Interoperable Framework

The lack of semantic alignment and machine-readable formats across these domains hampers the development of scalable decision support systems. While several modelling approaches and domain-specific ontologies have been proposed in research [8, 7], many remain narrowly scoped or lack industry traction. Similarly, pipeline implementations are often custom-built, with limited reusability beyond the original use case [6].

Our work addresses these gaps by committing to (1) widely adopted open technologies, particularly the FIWARE stack and NGSI-LD standard, (2) structured data representation using Smart Data Models, and (3) modular, Dockerised pipeline components that can be deployed at the edge or in the cloud, depending on operational needs. By aligning with these principles, the resulting architecture facilitates semantic interoperability, spatial and temporal fusion [1, 2], and integration with third-party datasets (e.g., satellite imagery or weather forecasts).

The motivation for this work lies in both practical needs, such as reducing manual labour and enhancing data reuse, and strategic goals of aligning with EU-level initiatives like Agri-DataSpace and the EU Data Act, which promote federated, standards-based data ecosystems for the agricultural sector.

2. Proposed Approach and Proof of Concept

This work proposes a unified, standards-based approach for transforming heterogeneous agritech data into interoperable information assets. It builds on FIWARE technologies, NGSI-LD data modelling principles, and Docker-based microservices to integrate aerial imagery, robotic sensing, and IoT measurements into a shared data space. This section outlines the key enabling technologies, the pipeline architecture, and practical implementation details.

Enabling Technologies

FIWARE Ecosystem. FIWARE[1] is an open-source platform that enables the development of interoperable, data-driven smart applications. Its core component, the Orion-LD Context Broker, manages context information in real time using standard APIs and NGSI-LD data structures. This modular architecture supports scalable ingestion and querying across heterogeneous data providers and consumers.
Smart Data Models. Smart Data Models[2] are semantic templates built on NGSI-LD, defining structured entities, properties, and relationships. They promote semantic interoperability across domains. Existing models are reused where applicable (e.g., `DeviceMeasurement`), while new domain-specific models were developed for aerial and robotic imagery (`DroneImage`, `RoboticFrame`) to handle use-case-specific metadata.
NGSI-LD Format. NGSI-LD [4] is a standard for representing context-aware data, enabling linked data structures with shared semantics. Key advantages include support for GeoJSON-based geolocation, ISO 8601 timestamps, and extensible relationships among entities. This enables time and spatial alignment across otherwise incompatible datasets.

Unified Pipeline Architecture

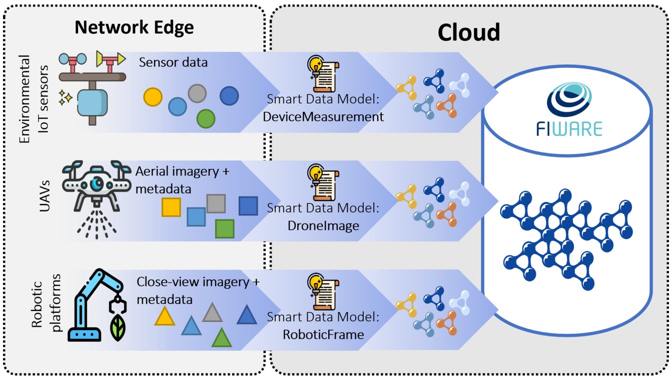

Fig. 1. Conceptual architecture of the three pipelines.

Our approach enables semantic data integration across three core domains: drone-collected aerial imagery, robotic imagery from ROS platforms, and IoT-based sensor streams. Each domain is handled by a dedicated transformation pipeline. These pipelines differ in entry point, format complexity, and deployment requirements but converge on a shared NGSI-LD model and integration into the FIWARE context broker (Figure 1).
Pipeline 1 – Aerial imagery: DJI drones capture high-resolution RGB imagery stored in DNG format with embedded EXIF metadata. Although EXIF includes essential context (GPS, timestamps, focal length), it is fragmented across namespaces (EXIF, XMP, Composite), lacks semantic clarity, and includes proprietary tags. **Implementation:** A Python script extracts and transforms this metadata using a custom Smart Data Model (`DroneImage`) into an NGSI-LD entity. The script is Dockerised for portability and executed on-premise after flight, due to the high image size (~100MB). Metadata includes capture date, geo-location, camera settings, and a URL to the image.
Pipeline 2 – Robotic imagery (ROS): Robotic platforms log multi-sensor data (e.g., RGB images, IMU, GPS) in ROS bag files. These contain fragmented topics with varying frequencies and message types (e.g., `sensor_msgs/Image`, `geometry_msgs/Vector3Stamped`). ROS-native structures are not interoperable with external systems. **Implementation:** A Dockerised Python script extracts and synchronises relevant topics, then consolidates them into an NGSI-LD entity using the `RoboticFrame` Smart Data Model. The pipeline runs locally after field

[1] https://www.fiware.org/
[2] https://smartdatamodels.org/

missions due to file size (hundreds of MBs). The resulting metadata captures image, position, orientation, and timestamp, with a URL to the original image.
Pipeline 3 – IoT measurements: IoT sensors transmit periodic measurements via NB-IoT to vendor-specific cloud platforms. Metadata is often disjointed (timestamps, geolocation, units), proprietary, and lacks semantic links to context (e.g., device type, provider). **Implementation:** A Python script, running continuously in the cloud or near the gateway, transforms sensor readings into NGSI-LD using `DeviceMeasurement` and other agrifood-related models. Each measurement is enriched with standard attributes (timestamp, geo-location, measurement type) and linked to devices and providers via NGSI-LD relationships.

Benefits of NGSI-LD Transformation

Despite format diversity, all three pipelines produce standardised NGSI-LD entities that are ingested into the Orion Context Broker. This common data space supports spatio-temporal fusion, uniform querying, and dashboard visualisation. This architecture provides a modular, future-proof foundation for building advanced decision support tools, integrating external datasets (e.g., weather, satellite imagery), and enabling new agritech services aligned with European interoperability directives.

Key challenges of the original formats	Key benefits of the transformed NGSI-LD format
Scattered/disjoint metadata: Contextual data (e.g., timestamps, geolocation, sensor or device metadata) is spread across multiple namespaces, files, or topics: `EXIF`, `ROS bags`, and separate sensor metadata files.	**Unified entity representation:** Each dataset is transformed into a single NGSI-LD entity (e.g., `DroneImage`, `RoboticFrame`, `DeviceMeasurement`) representing all relevant context in one structure.
Proprietary & inconsistent structures: Each data source uses its own format: EXIF for UAV imagery, ROS message types for robots, and custom enumerations for IoT sensors, requiring domain-specific logic to interpret or align.	**Semantic interoperability and enrichment:** All properties and relationships are semantically defined using Smart Data Models (e.g., `measurementType`, `deviceDetails`, `captureDate`, `geoLocation`).
Limited interoperability: Raw data structures are tightly coupled to specific vendors/platforms (e.g., DJI cameras, ROS message types, vendor-specific IoT APIs), blocking cross-system integration.	**Spatial and temporal alignment:** Geo-coordinates are consistently represented using GeoJSON; timestamps are standardised to ISO 8601 across all data types.
No semantic clarity or linked context: Field names lack semantic meaning and are not explicitly linked (e.g., sensor measurements not linked to geolocation or device metadata in machine-readable form).	**Cross-platform integration:** NGSI-LD enables seamless ingestion into context-aware systems (e.g., FIWARE's Orion-LD), supporting integration with third-party data sources like satellite imagery or weather forecasts. **Machine-readable relationships:** Explicit links are established between entities (e.g., sensor→ device→ location→ company), enabling scalable querying and analytics.

Table 1. Key challenges in original data formats vs. benefits of NGSI-LD transformation.

3. Discussion and Conclusion

The data pipelines presented in this work offer a robust foundation for building decision-support systems in the agrifood sector. By transforming raw data from drones, robotic platforms, and IoT sensors into NGSI-LD-compliant entities, they enable the creation of a unified, interoperable data space where diverse data streams can be geospatially and temporally fused. This harmonised representation improves data quality and usability, allowing end users to derive actionable insights for tasks such as crop monitoring, irrigation planning, and disease or pest control. While the approach introduces trade-offs, such as increased data verbosity and greater initial modelling effort, it ensures semantic clarity, structural consistency, and long-term maintainability. These trade-offs serve as a safeguard against technical debt, allowing systems to evolve as new data sources and services are integrated. The architecture also facilitates the incorporation of third-party datasets (e.g., satellite imagery, weather forecasts, fertiliser schedules), significantly broadening its utility.

The benefits extend beyond end users. Data providers gain monetisation opportunities by contributing standardised datasets to the common space, which can be bundled into commercial information products. Shared standards such as NGSI-LD create synergy effects across datasets and enable automated integration with third-party platforms, unlocking new business models including real-time analytics services, advisory platforms, and autonomous farming tools. The architecture is well aligned with current EU strategies for agricultural data interoperability. Initiatives such as AgriDataSpace[3] and the EU Data Act [5] call for standardised, reusable formats and shared principles for secure and efficient data exchange. By adopting NGSI-LD and Smart Data Models, this work contributes to that vision, offering a reusable, scalable, and standards-compliant solution that promotes sustainability and innovation in agriculture. As future steps, we will continue refining the custom Smart Data Models for aerial and robotic imagery through broader community collaboration under the Smart Data Models initiative.

Acknowledgements

This work is funded by the Research Council of Norway under grant agreement no. 337012 (SMARAGD) and the European Union's Horizon Europe Research and Innovation Programme under grant agreement No. 101135576 (INTEND).

References

[1] Dautov, R. and Distefano, S.: Distributed data fusion for the Internet of Things. In: *Proceedings of 14th International Conference on Parallel Computing Technologies (PaCT 2017)*. Springer. 2017, pp. 427–432.

[2] Dautov, R. and Distefano, S.: Three-level hierarchical data fusion through the IoT, edge, and cloud computing. In: *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. 2017, pp. 1–5.

[3] Dautov, R., Tverdal, S., Bondevik, A. S., Frøshaug, S. A., Szabo, V., Fiksdal, J. R., and Stølen, M. F.: SMARAGD: Data Interoperability for Decision Support in the Norwegian Agrifood Sector. In: *Companion Proceedings of the 8th International Joint Conference on Rules and Reasoning (RuleML+RR-Companion 2024)*. 2024.

[4] ETSI Industry Specification Group (ISG): *Context Information Management (CIM); NGSI-LD API*. Tech. rep. RGS/CIM-009v161. ETSI, 2019.

[5] European Commission and Directorate-General for Communication: Data Act – The path to the digital decade. Publications Office of the European Union, 2022.

[6] Firmani, D., Leotta, F., Mathew, J. G., Rossi, J., Balzotti, L., Song, H., Roman, D., Dautov, R., Husom, E. J., Sen, S., Balionyte-Merle, V., Morichetta, A., Dustdar, S., Metsch, T., Frascolla, V., Khalid, A., Landi, G., Brenes, J., Toma, I., Szabó, R., Schaefer, C., Udroiu, C., Ulisses, A., Pietsch, V., Akselsen, S., Munch-Ellingsen, A., Pavlova, I., Kim, H.-G., Kim, C., Allen, B., Kim, S., and Paulson, E.: INTEND: Intent-Based Data Operation in the Computing Continuum. In: *CEUR Workshop Proceedings*. Vol. 3692. 2024, pp. 43–50.

[7] Malvar, S., Badam, A., and Chandra, R.: FarmBeats: Digital Water for Agriculture. In: *Resource Magazine* 29.4 (2022), pp. 40–42.

[8] Pesce, V., Kayumbi, G.-W., Tennison, J., Mey, L., and Zervas, P.: Agri-food data standards: a gap exploration report. In: *F1000Research* 7.176 (2018), p. 176.

[3] `https://agridataspace-csa.eu/`