

RESEARCH

Open Access



Big Data and precision agriculture: a novel spatio-temporal semantic IoT data management framework for improved interoperability

Mario San Emeterio de la Parte^{1*}, José-Fernán Martínez-Ortega¹, Vicente Hernández Díaz¹ and Néstor Lucas Martínez¹

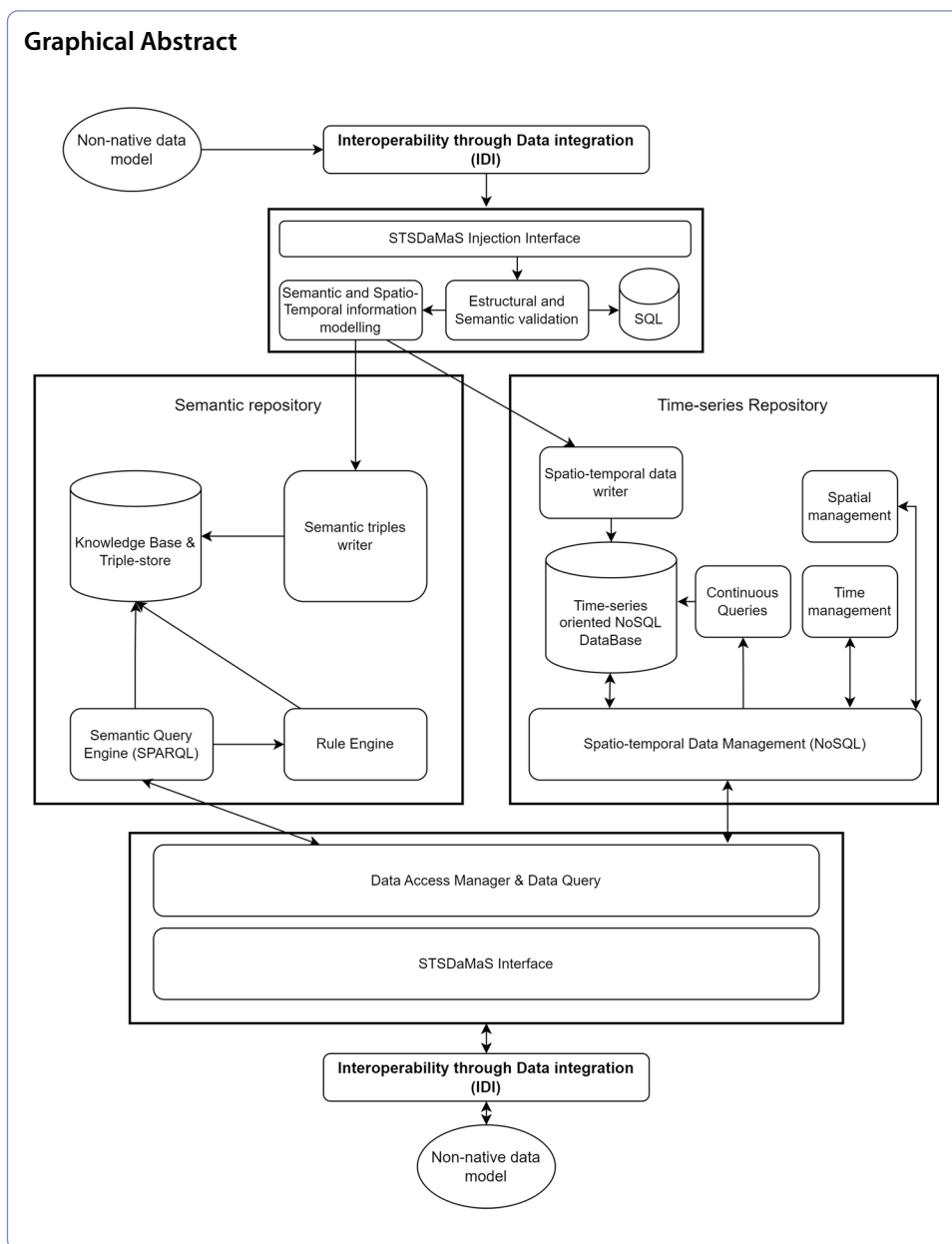
*Correspondence:
mario.sanemeterio@upm.es

¹ Group of Next Generation Networks and Services (GRyS), Departamento de Ingeniería Telemática y Electrónica (DTE), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Abstract

Precision agriculture in the realm of the Internet of Things is characterized by the collection of data from multiple sensors deployed on the farm. These data present a spatial, temporal, and semantic characterization, which further complicates the performance in the management and implementation of models and repositories. In turn, the lack of standards is reflected in insufficient interoperability between management solutions and other non-native services in the framework. In this paper, an innovative system for spatio-temporal semantic data management is proposed. It includes a data query system that allows farmers and users to solve queries daily, as well as feed decision-making, monitoring, and task automation solutions. In the proposal, a solution is provided to ensure service interoperability and is validated against two European smart farming platforms, namely AFarCloud and DEMETER. For the evaluation and validation of the proposed framework, a neural network is implemented, fed through STSDaMaS for training and validation, to provide accurate forecasts for the harvest and baling of forage legume crops for livestock feeding. As a result of the evaluation for the training and execution of neural networks, high performance on complex spatio-temporal semantic queries is exposed. The paper concludes with a distributed framework for managing complex spatio-temporal semantic data by offering service interoperability through data integration to external agricultural data models.

Keywords: Spatiotemporal Data Bases (STDB), Internet of Things (IoT), Big Data, Data auditing, Service interoperability, Precision agriculture, Neural networks, Machine Learning (ML), Decision Support Systems (DSS)



Introduction

The deployment of new technological solutions in the agricultural sector aims to achieve the Sustainable Development Goals (SDGs) [1] and the welfare of animals [2]. With this objective, the Food and Agricultural Organization [3] and the International Fund for Agricultural Development (IFAD) [4] arise to promote research [5], investment and sustainability in agriculture.

Since 2015, the number of people suffering from hunger has increased, reaching a percentage of almost 9% of the world’s population [6]. If the current trend continues, the Zero Hunger Goal [7] will not be achieved by 2030. Targets 2.3 and 2.a of the Zero Hunger goal aim to double agricultural productivity and increase agricultural research.

To achieve this, precision agriculture solutions can offer increased productivity and sustainable food production.

Precision agriculture provides a link between the farmer and technology. IoT systems are one of the main approaches to automate, monitor, and improve the performance of agricultural activities. However, despite the multitude of devices deployed in agricultural scenarios, bringing the understanding of technological solutions closer to farmers and agricultural specialists remains a challenge.

Technological solutions in agriculture must focus on understandable communication with the user, in this case, the farmer or the specialist. To do this, it is necessary to understand the way the professional or farm manager usually works and acts, trying to respond to the questions and tasks performed in a scenario that does not have high-technology or IoT solutions. Knowing the usual way of working will make the inclusion of intelligent and precision solutions more harmonious. The aim is to facilitate and improve routine tasks with as little impact as possible and without the need for technical training on the part of the user.

IoT solutions in agriculture are described by constant measurement of land, crop, or livestock characteristics depending on the use case. In turn, automation and monitoring solutions are fed by the data captured by the devices deployed in the field. Therefore, the representation or collection of data stored in repositories is one of the key elements in communication with the farmer.

Data management systems must be able to answer the questions the farmer asks. For example, When should I harvest the fodder to feed the livestock? When should I bale the harvested crop? In essence, such questions can be disaggregated according to three main dimensions: (i) temporal, (ii) spatial, and (iii) semantic. In turn, the data extracted through the decomposition of the queries posed by the farmer serve as input for Machine Learning models or Decision Support Systems (DSS), offering a much more useful and interpretable answer for the farmer.

In this article, an innovative spatio-temporal semantic data management framework is proposed for the resolution of the queries raised above. In addition, a solution for the preparation and transformation between agricultural data models is presented to ensure the interoperability of the system without dependence on the native data model. Finally, a use case is presented in which the system feeds a Machine Learning model to answer certain predictive questions in the field of harvesting or cultivation of fodder legumes for livestock feed.

The proposal presented in this paper has been developed in the framework of a European precision agriculture research project. The data set exposed for the system validation is made up of real data generated by static and mobile sensors and devices in an agricultural scenario. The system is designed for the management of large semantic spatio-temporal datasets, generated by sensors and devices deployed in agricultural scenarios. However, the system is easily adaptable to other IoT domains, with easy adaptability due to the nature of the agents involved.

Agricultural scenarios present certain limitations, such as low computational capabilities and limited resources of certain sensors or connectivity problems in some farms. Some services, solutions, or devices operate with native data models, which makes their incorporation into other architectures difficult. This article describes a proposal that

solves these problems through a distributed architecture and a component that guarantees service interoperability.

In [Related work](#), some of the most innovative solutions in precision agriculture are presented, exposing the importance of sensor networks, the Internet of Things, and data science. In addition, data management solutions for precision agriculture exposed in the literature are presented, and how the proposal offers a substantial and innovative improvement of current solutions. [Spatio-Temporal Semantic Data Management System \(STSDaMaS\)](#) presents the STSDaMaS proposal, detailing the architecture and functionalities offered by the framework. [Achieving interoperability, use case with the AFarCloud Data Model](#) presents the solution adopted to ensure the interoperability of the STSDaMaS framework with data models from other platforms, projects, or solutions. For interoperability validation, the native data model of the AFarCloud smart agriculture platform is presented, and data preparation, adaptation, and transformation are exposed. [System validation: Applying Neural Network for fodder legume harvesting and bailing prediction](#) describes the validation of the proposed spatio-temporal semantic data management framework through a use case to feed a machine learning model. [Conclusion](#) presents the conclusions of the proposal. Finally, [Future work](#) evaluates the limitations of the study and summarizes ongoing research to address them.

Related work

Automation of tasks in the application domain of precision agriculture is closely linked to the constant monitoring of the condition of crops and livestock. As such, IoT solutions offer a perfect symbiosis with the application domain. In the literature, several reviews of IoT solutions for process automation in precision agriculture are found [8].

The first and most characteristic phase of IoT implementations is the collection of data by devices deployed in the field. Therefore, data characterization is a key element for subsequent processing and implementation of solutions. Data captured by precision agriculture devices offer information related to three main dimensions (i) semantics [9], (ii) spatial, and (iii) temporal. This type of data description is provided by semantic models of linked data [10].

Field data management in precision agriculture has been linked to the use of Geographic Information Systems (GIS). This set of tools allows for the storage, analysis, manipulation, and mapping of any type of georeferenced information. For Precision Agriculture applications, a specific GIS system called Field-Level Geographic Information System (FIS) [11] was developed. However, this system was developed for older operating systems, such as Windows 3.1x, 95, and 98. The updated version is the Farm Management Information System (FMIS). According to Burlacu et al. [12] FMIS supports farmers in various areas, from operational planning, implementation, and documentation, to evaluation of the work done in the field. However, FMIS does not consider the temporal dimension and its application to scenarios that require real-time operation is limited.

Precision agriculture presents an environment rich in spatio-temporal data. The open-source GeoFIS software, designed to cover the entire process, from spatial data to spatial information and decision-making, is presented in [13]. GeoFIS has been evaluated through three case studies involving different types of crops, and the ability of its

embedded algorithms to address the needs of farmers, advisors, and spatial analysts in precision agriculture is examined. The main limitation of GeoFIS lies in the lack of characterization and management of the temporal and semantic information of the data.

With the increasing deployment of sensors and static and mobile devices for monitoring and automating tasks in all sectors, the most current literature gathers a prodigious amount of research on the efficient management of large amounts of spatio-temporal data. Ruiyuan Li et al. present JUST [14], a data management system that can efficiently store and manage large amounts of spatio-temporal data. It uses a combination of Apache HBase, GeoMesa, and Apache Spark and includes two indexing techniques and a compression mechanism to improve query efficiency. It also supports easy-to-use SQL-like queries and allows for both new data insertions and updates to existing data without requiring index reconstruction. Experimental results show that JUST has strong query performance and scalability compared to other distributed spatio-temporal data management systems. However, JUST has certain limitations regarding the semantic characterization of the data and the inference or treatment of relationships between entities. In addition, interoperability with other data models is a pending work that limits its application to other domains.

As the number of mobile devices and IoT terminals connected to the Internet increases, large amounts of spatio-temporal data are generated. To store and compute these data locally and share them effectively, data owners often store them on cloud servers. Yongjiun Ren et al. present BSMD [15], a blockchain-based secure storage mechanism that uses an on-chain and off-chain cooperative storage model to overcome the limited storage capacity of the blockchain. This novel mechanism is capable of storing large amounts of spatio-temporal data securely. However, the implementation of blockchain-based mechanisms to guarantee data integrity and security requires high computational capabilities and makes real-time operations impossible.

Real-time management of large data sets is becoming increasingly necessary in various sectors. In [16], Atsushi Isomura et al. present the *Axispot* technology, applied to obstacle detection and lane-specific congestion in the automotive domain. The article describes the need to store and query data captured by numerous moving objects. *Axispot* presents an improvement in the capability to search and aggregate spatial data by reducing the complexity of polygonal shape lines and the number of vertices to cope with the needs of a near real-time operating environment. However, the study does not consider the dimension of height or altitude, which is considered for future developments. The development of exposed technology is related to the automotive world and the high computational capabilities of the agents involved, so its adaptability to other sectors is reduced.

The application of spatio-temporal semantic data management systems in precision agriculture still has a long way to go in terms of research and improvement. However, in the literature, there are already some systems, such as SEMAP, presented by Henning Deeken et al. [17], mainly developed for spatio-semantic management for robot planning, which later on [18, 19], was successfully applied to the precision farming domain. The SEMAP framework is characterized by a powerful spatial description of the agents and the environment; offering 2D and 3D representations. However, this complex characterization is not necessary for most use cases in agriculture. Moreover, as the authors

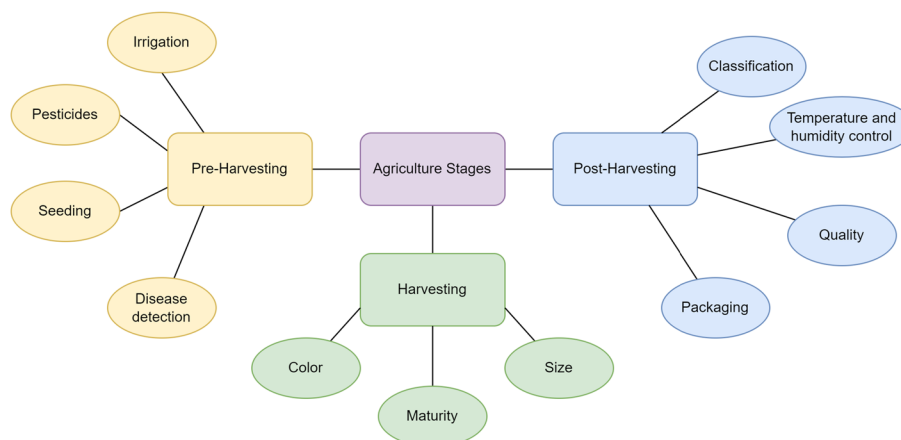


Fig. 1 Stages in crop management

argue in the article, the temporal dimension is a missing piece in the system. Thus, this article proposes a data management framework that prioritizes temporal-spatial and semantic characterization, decreasing the complexity of the spatial dimension; aiming to offer a novel and more complete solution.

In the last decade, advances in Web 2.0 technology have improved the accessibility of data across the cloud, with increasing amounts of Linked Open Data, Linked Open Statistical Data, and Open Government Data available for sharing. However, the availability and accessibility of data for production analytics in agriculture are currently limited and rarely include attributes for spatio-temporal analytics. Irya Wisnubhadra et al. [20] propose the development of a spatio-temporal data warehouse with an integration process using a service-oriented architecture and open data sources to address this problem and improve decision-making in agricultural production analytics. The storage approach provided in this article is geared towards agricultural production analysis, presenting a low temporal precision (aggregations or daily results). The query response time (even longer than 12 s) and the exposed data model complicate its adaptability for sensor networks and scenarios with real-time query requirements.

Due to information management systems and their post-processing and querying capabilities, raw data captured by various devices are injected as input for Machine Learning (ML) models, Decision Support Systems (DSS), or monitoring, among others. Due to the utility and success of machine learning model application, this paper presents the implementation of a neural network in precision agriculture, for the validation of the proposed data management framework. Several studies on the application of ML in IoT for precision agriculture are found in the literature [21–23].

Due to the number of tasks in the agricultural sector, this article focuses mainly on crop management. Crop management tasks can be divided into three main groups (i) pre-harvest, (ii) harvest, and (iii) post-harvest (see Fig. 1). In [24] a review of ML applications is presented in each of the areas.

One of the most interesting applications in crop management is the cultivation of fodder legumes for livestock feed. These crops form a link between two of the most

Table 1 A comparative between the most relevant related proposals

Proposal Ref #	Authors	Big ST data management	Distributed cloud-edge	Service Interoperability	Scalability	Agricultural application	Adaptable other domains
11	Zhang et al.	N	N	N	Y	Y	N
12	Burlacu et al.	N	N	N	Y	Y	N
13	Leroux et al.	N	N	N	N	Y	Y
14	Li et al.	Y	N	N	Y	N	Y
15	Ren et al.	Y	N	N	Y	N	Y
16	Isomura et al.	Y	N	N	Y	N	N
17	Deeken et al.	N	N	N	Y	N	Y
19	Deeken et al.	N	N	N	Y	Y	Y
20	Wisnubhadra et al.	Y	N	N/A	Y	Y	N/A
Our proposal		Y	Y	Y	Y	Y	Y

important areas of the agricultural sector, crops, and livestock management. The literature exposes the interest in the area, with studies such as [25] providing a review of the welfare of dairy cows in pasture or confinement. The importance of diet and therefore the quality of forage offered as feed to cattle for its yield and meat quality is defined in [26].

The determination of the quality of fodder is not immediate and depends on many factors [27]. Some of the most important factors in crop yield can be predicted with ML models thanks to solutions such as the one presented by Johann et al. for soil moisture in [28] or soil temperature by Behnaz Nahvi et al. in [29]. There are more modern solutions, such as the one presented by Vamseekrishna et al. in [30]. Once the soil characteristics and main factors affecting the fodder crop have been studied, the farmer's main concern will be to estimate the productivity or yield of the crop itself. Jeevan Nagendra Kumar et al. present a use of the ML application for the prediction of crop performance in [31].

This paper proposes a novel spatio-temporal semantic interoperable data management framework that allows the exploitation of big data, applied to an agricultural scenario. The system proposed in this paper allows the resolution of complex queries thanks to a distributed repository between time-series oriented databases and knowledge bases or triple stores. Due to a distributed architecture, the system offers real-time operation performance by deploying part of the architecture at the edge. Due to the proposed design, the system can adapt to environments with limited computational or network resources. Furthermore, the proposed framework includes a solution to ensure the interoperability of the service through integration with other data models available in the application domain. In Table 1, a comparison is presented between the most relevant proposals related to our work, based on the key features that our proposal presents: (i) Big spatiotemporal (ST) data management, (ii) architecture distributed between edge and cloud, (iii) service interoperability, (iv)

scalability, (v) specific application and design for agriculture, and (vi) adaptable to other domains.

Spatio-temporal semantic data management system (STSDaMaS)

Precision agriculture presents constantly evolving scenarios. Observations captured by devices deployed on the farm are characterized by three dimensions: (i) temporal evolution of conditions; observations are associated with a time stamp. (ii) Spatial geopositioning stamp; with static or dynamic nature, depending on the type of device that captures the measurement. (iii) Semantic characterization of the data and its relationships with other devices and the environment.

Precision agriculture data can be characterized by their spatio-temporal semantic nature. In this way, it is possible to answer queries such as: What has been the evolution of soil temperature and moisture in a given plot over the last week? Usual queries have an equivalent characterization.

The design of a system capable of managing spatio-temporal semantic data in precision agriculture faces several problems:

1. Implement a repository capable of managing a large volume of historical data with high performance.
2. Design a query system that allows data retrieval under spatial, temporal, semantic constraints, or combinations thereof.
3. Offer interoperability with other services through the transformation of input and output data.

This paper proposes a spatio-temporal semantic data management system that solves these problems, considering the importance of data retrieval for future monitoring, decision-making, and automation of agricultural processes.

Data management architecture

The STSDaMaS framework is designed for the collection and management of agricultural data from the spatio-temporal semantic environment. The agricultural domain model must be able to describe the temporal evolution of the observations captured by a device and provide its geopositioning and associated semantic information. For mobile devices, such as robots, unmanned aerial vehicles (UAVs), or autonomous or semi-autonomous vehicles, a spatial evolution linked to the temporal axis is presented. In addition, the semantic description presents a classification of devices, sensors, observations, and even the environment, as well as their relationships.

According to the characterization of data in the agricultural domain, it is essential to differentiate between temporal and spatial information and semantic information.

Regarding the type of database for the implementation of the repository, semantic databases or triple-stores (for RDF graphs) allow storing and managing individuals and their relationships, adding a powerful semantic description of data. However, their performance decreases with the management of a large-volume dataset. Non-relational databases allow the management of a large volume of heterogeneous data

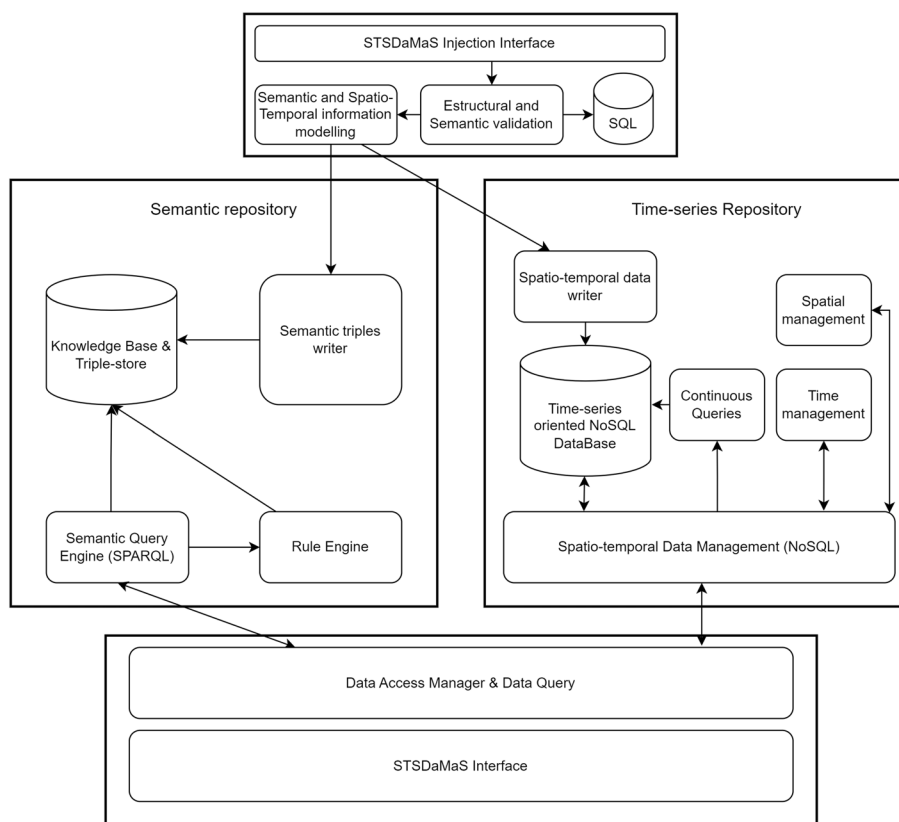


Fig. 2 Spatio-temporal semantic data management system (STSDaMaS)

with higher performance. However, they do not support inference operations or more complex semantic descriptions and relationships between objects and the environment.

It is proposed to distribute the repository internally between knowledge and non-relational databases. The different types of database engines prioritize the management of data characterized by a specific dimension (spatial, temporal, or semantic). In precision agriculture, the data have a strong characterization under the three defined dimensions, so it is reasonable to distribute the management framework to ensure maximum performance. The proposal consists of the use of a NoSQL Time-Series oriented Database (TSDB) to increase performance in the management of spatio-temporal data and the use of a Triple-Store to ensure the management of the semantic information. This distributed architecture allows increasing performance and speeding up queries in the Big Data environment, without losing the benefits of semantic description. To link the objects represented in both databases, unique element identifiers (UIDs) are used.

Figure 2 shows the architecture of the proposed spatio-temporal semantic data management system (STSDaMaS).

Semantic repository: triple store

The semantic repository implemented in the STSDaMaS framework is responsible for the management of semantic information associated with observations, devices, and the environment or terrain in which agents are deployed. Relationships between agents provide powerful information for decision-support solutions and even for identifying relevant information in the detection of anomalies or specific situations on the farm.

The semantic repository includes a rule engine, which enables the inference of new individuals and rule-based reasoning about the triples stored in the repository. An example of a use case favored by this type of implementation can be found in agricultural diversification. Thanks to information on the terrain and conditions of a given crop, specific reasoning can be performed to determine the viability and yield of a new crop.

For the testing and evaluation of the system proposed in this article, the Jena Rules [32] engine is used. The rules are defined in a file with the specific syntax of Jena Rules. However, other rule-based languages such as SWRL (Semantic Web Rule Language) [33], RIF (Rule Interchange Format) [34], N3 (Notation 3) [35], or RuleML [36], among others, could be used for the definition of rules.

To identify a given individual in the knowledge base, the property <http://purl.org/dc/terms/identifier> is used, allowing the association with its spatio-temporal information. To link the semantic container with the spatio-temporal information associated with a given sensor or device, the *sosa:madeBySensor* property is used. The content associated with the property *sosa:madeBySensor* reveals information about the type of device, the service offered, the provider, the name of the device, and the identifier of the TSDB in which the associated spatio-temporal information is stored. STSDaMaS internally uses this type of property to make associations between semantic information and the spatio-temporal evolution described by the primitive.

Space-time repository: time-series oriented database (TSDB)

Spatio-temporal characterization of data provides powerful information to observe the evolution of conditions associated with a given crop or terrain. In applications such as livestock farming, it makes it possible to track livestock and study their behavior under certain climatic or environmental conditions on the farm. Therefore, spatio-temporal information of data is a key element to increase agricultural performance and implement solutions for decision-making or automation.

As spatio-temporal data management is one of the main elements in the proposal and to ensure the most efficient performance, some of its weaknesses and strengths must be analyzed and understood.

There is a multitude of time-series oriented databases. Modern TSDBs include Prometheus, InfluxDB, TimescaleDB, PostgreSQL, and Cassandra. To maximize the performance of TSDBs, special attention must be paid to (i) data modeling, (ii) data structuring in tables, and (iii) Write-ahead logging (WAL).

In terms of data modeling, TSDBs use the timestamp of the data as an index for indexing, so there is high performance when planning and retrieving data through temporal

queries. However, the addition of filters as clauses in queries can increase planning and execution delays. To maintain high performance, careful modeling of the data is required. First of all, the types of attribute described in the data can be defined as “Tags” or “Fields”. “Tags” will be stored in the server memory until they disappear from the current shard.¹ In contrast, the information of attributes defined as “Fields” will be stored directly on disk. Therefore, queries whose filters include attributes defined as tags will perform better than those oriented to the search of attributes stored on disk.

Because STSDaMaS uses the TSDB engine for data gathering concerning their spatio-temporal characterization, it is understandable to think that spatial information attributes have to be defined as “Tags”. However, the spatial marking of a datum represents an attribute of large variability and, therefore, high cardinality. That is, it can present a multitude of values in the repository. This causes high usage of the host server’s RAM resources, even blocking the system.

The definition of tags should be limited to attributes with low-variability information and high frequency of use in the query predicate. In the proposal, only attributes with information related to the following are defined as tags: device identifier, service, scenario, device type, and provider.

Three main sources of delay can be identified in data collection: (1) the planning of data gathering; (2) the execution of the plan and concurrent attack of processing threads to the different tables in which data reside; (3) serialization of data for delivery as a response to the query. These three sources of delay can be reduced by correct structuring of tables for data storage according to the nature of the data. Table 2 proposes the high-level structure of the measurement tables for an agricultural scenario.

Write-Ahead Logging is an important element in a time-series oriented database architecture, but it can limit performance in a near real-time scenario for querying the latest data injected into the TSDB. The WAL is a temporary cache for recently injected points (writes). To reduce the frequency with which permanent disk storage files are accessed, the TSDBs engine stores new points in the WAL and groups them in batches until their total size or age triggers a flush into permanent storage. WAL is also used as protection against the loss of recently added data on power loss. However, the WAL storage format is not easily queryable. For this reason, the proposed system includes a small relational database (SQL), which allows to speed up the collection of the last injected measurements. The SQL database will only store the last measurement obtained by each sensor, and its use can be of great utility because it can be deployed directly on the edge to increase system performance in certain scenarios.

Space-time repository: relational database (SQL)

The implementation of a small relational database on top of the STSDaMaS repository reduces the delays introduced when querying the latest data injected into the system using the WAL, and querying repositories with large volumes of data. But its

¹ A shard is a horizontal partition of data in a database or search engine. Each shard remains in a separate instance of the database server, to spread the load.

Table 2 Repository structure in tables

Repository structure in tables					
Collars	Static devices	Mobile devices	Machinery	State vectors	Environmental devices
Contains the measurements captured by sensors embedded in collar-shaped devices. These devices are fitted to the neck of the cattle	Contains observations captured by sensors or static devices deployed in the field	Composed of observations from sensors installed in mobile devices, robots or vehicles, whose geoposition will be updated	Information on the condition of agricultural machinery. Battery, Fuel, Engine, etc	Information about the state vector of a vehicle, especially useful for solutions that integrate unmanned aerial vehicles. Orientation, linear velocity, etc	Observations captured by environmental information devices such as weather stations

```

PREFIX sosa: <http://www.w3.org/ns/sosa/>
PREFIX ns: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?urn

WHERE {
  ?sensor sosa:madeBySensor ?urn .
  ?sensor ns:type <https://w3id.org/demeter/default-
    context/SoilSensor> .
  FILTER (REGEX(?urn, "^urn:.*:fct:.*")) .
}

```

Listing 1 SPARQL query

introduction is even more interesting because of the distribution capacity it gives to the architecture.

Precision agriculture presents several scenarios in which data gathering with strong semantic characterization is not necessary but prioritizes the updating of values or measurements in real-time. Solutions, such as the inclusion of actuators in response to anomalous responses captured by sensors (such as in fire cases) or livestock grazing using drones, require real-time interaction. In other cases, the farm scenario has low networking capabilities or resources to deploy a system like STSDaMaS directly at the edge or in the fog. However, it can perform a data update on the system deployed in the cloud, still losing real-time operation.

However, the introduction of a relational database in the STSDaMaS framework allows this part of the architecture to be deployed directly at the edge, enabling real-time response to queries of less semantic complexity, but requiring real-time performance. In this way, batching of measurements can be performed at the edge and automatically sent to the cloud where the rest of the system is located when connection and conditions permit.

Equations 1, 2, and 3 describe the response times for a given query to the component deployed on the edge (Eq. 1) or in the cloud (Eq. 2). Note that the query processing times at the edge will be lower since the query target is a relational database with a low amount of data, achieving real-time performance. However, if the query is more complex or its predicate aims at extracting historical information, it will be necessary to redirect the query to the STSDaMaS query engine located in the cloud.

$$t_{gathering-edge} = t_{con-edge} + t_{querySQL} + t_{response'} \quad (1)$$

$$t_{gathering-cloud} = t_{con-cloud} + t_{querySTSDaMaS} + t_{response''} \quad (2)$$

where $t_{querySTSDaMaS} = t_{query-TripleStore} + t_{query-TSDB} + t_{serialization}$

$$t_{con-edge} < t_{con-cloud} \quad \text{and} \quad t_{response'} < t_{response''} \quad (3)$$

Due to this versatility, the proposed system can be adapted to numerous farming scenarios even with different requirements or needs.

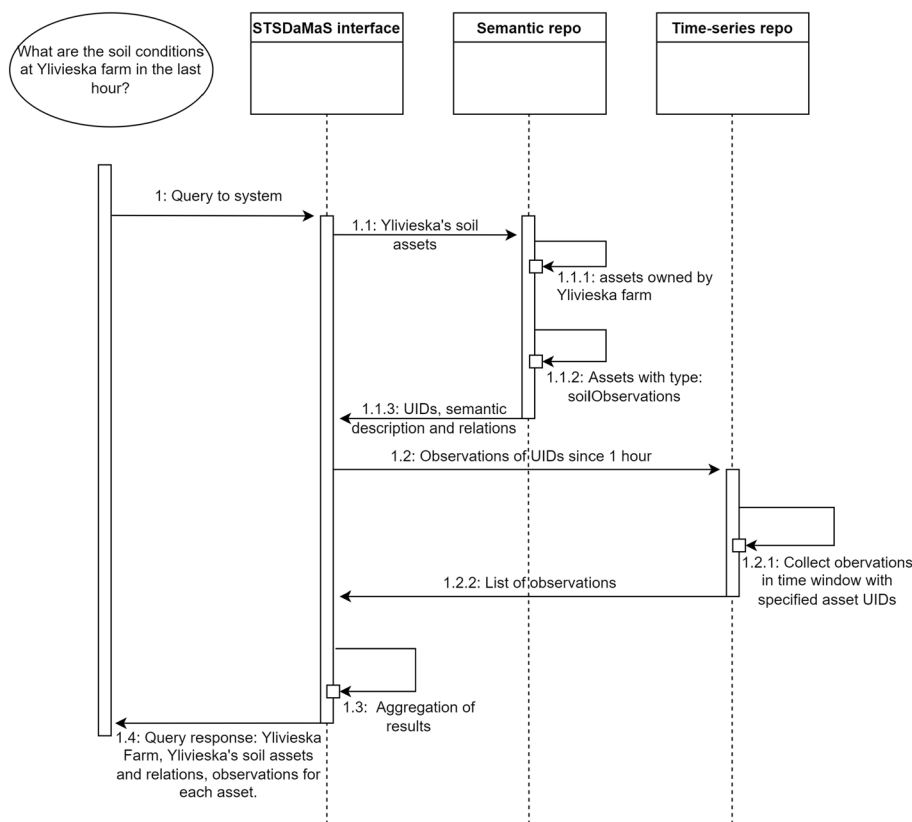


Fig. 3 High-level query processing by STSDaMaS

Query system and data retrieval

STSDaMaS is designed for the resolution of semantic queries whose subject varies in temporal and spatial character. Due to the distribution of the query system in knowledge and oriented to time series repositories, it offers high-performance retribution, whose response provides a powerful description of the environment, devices, and evolution present in the indicated context.

STSDaMaS processes the data query operation according to the following sequence of high-level actions: (1) Through a data query interface, the semantic predicate of the query is extracted from the request, attacking a triple-store knowledge database (see a simple example of an SPARQL query in Listing 1). (2) Through a unique identifier (UID) of the observations or devices involved, spatio-temporal information is extracted from the non-relational time-series-oriented database. This allows for a fast and lightweight collection of the desired historical or current data. (3) Finally, the data are aggregated for delivery as a response to the query.

To illustrate the high-level operation of querying data through the system, a practical example is presented. For the query “What are the soil conditions at the Ylivieska farm in the last hour?”: (i) The system extracts the semantic predicate, collecting from the knowledge database all those sensors that perform measurements on soil conditions and that are associated with the farm with the identifier “Ylivieska”. (ii) Through the UIDs of the sensors or observations associated with the farm and providing soil measurements,

the NoSQL database is accessed. Filtering by UIDs, the observations captured in the established time window are extracted, together with the ge positioning in the case of static devices or the different spatial marks in the case of mobile devices. (iii) The system will carry out the data aggregation, offering, as a response to the request, the farm object with all the associated soil-type devices and the observations captured in the indicated time window, together with their spatial mark (latitude and longitude or geohash). The process is depicted in Fig. 3.

STSDaMaS framework not only extracts a complete semantic description of the data and their relationships; it also avoids querying a time-series oriented repository with a large volume of data directly. Instead, it is pre-filtered by selecting the identifiers of those observations or devices to be queried. This approach helps to collect a complete, understandable, and semantically rich response, avoiding the long delays produced when querying directly on historical databases. The greater the volume of data to be extracted, the greater the efficiency shown by the framework with respect to conventional solutions consisting only of a homogeneous repository.

The framework includes a system for managing continuous queries against the historical repository (scheduled queries). Such queries can be configured to perform operations such as the calculation of average values at set time intervals or the detection of anomalous observations (e.g. out-of-range measurements, or ge positioning modification of a static device).

For monitoring, as input for training Machine Learning (ML) models, for statistical analysis, and other types of solution, it will be necessary to extract data over a multitude of time points. However, as input for graphical analysis applications or as feedback to the farmer, in most cases, it will be sufficient to obtain certain aggregations of the data. For example, average, minimum, and maximum values, or other simple aggregations, over the observations queried. The continuous query management system allows the establishment of rules or aggregation functions within the historical repository of the system. In this way, operations such as the calculation of average, maximum, or minimum values are carried out on a scheduled basis, reducing delays in data collection due to the steady availability of these values.

STSDaMaS data model—Agricultural Information Model (AIM)

The STSDaMaS framework is proposed in the context of the H2020-DEMETER project (Building an Interoperable, Data-Driven, Innovative, and Sustainable European Agri-Food Sector) [37]. The DEMETER goal is to lead the digital transformation of Europe's agri-food sector through the rapid adoption of advanced IoT technologies, data science, and smart agriculture, contributing to long-term viability and sustainability. It aims to empower farmers and agricultural cooperatives to (i) use their existing machinery and platforms by enabling knowledge extraction for improved decision-making and (ii) facilitate the evolution of their devices and platforms, targeting their investments where needed.

The STSDaMaS input and output data model is modeled according to the Agricultural Information Model (AIM) [38], developed in the DEMETER project.

```

{
  "@id": "urn:afc:fct:environmentalObservations:UPM:soil:
    KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653",
  "type": "Observation",
  "hasFeatureOfInterest": "http://www.w3id.org/afarcloud/poi?lat
    =64.05029&long=24.72468",
  "hasResult": "urn:afc:fct:environmentalObservations:UPM:soil:
    KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653/q1",
  "madeBySensor": "urn:afc:fct:environmentalObservations:UPM:soil:
    KotipeltoFarmWeatherStation",
  "observedProperty": "http://www.w3id.org/afarcloud/
    obs_air_humidity",
  "resultTime": "2018-01-01T12:36:12Z"
}, {
  "@id": "urn:afc:fct:environmentalObservations:UPM:soil:
    KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653/q1",
  "type": "QuantityValue",
  "identifier": "KotipeltoFarmWeatherStation/obs_air_humidity
    -1579311537653/q1",
  "numericValue": "3.5",
  "unit": "http://qudt.org/vocab/unit/DEG.C"
}

```

Listing 2 AIM JsonLD representation of an observation

The AIM model is based on the SOSA (Sensor, Observation, Sample, and Actuator) and SSN (Semantic Sensor Network) ontologies, which aim to model sensors, actuators, devices, and observations (measurements captured by sensors). AIM focuses on the semantic characterization of the data, providing detailed information about the observations captured by the sensors and their relationships with other agents in the field.

The input and output data of the platform are described by the AIM model using the JSON Linked Data (JSON-LD) syntax for the Resource Description Framework (RDF). An example of the modeling of observations captured by environmental sensors is presented in Listing 2.

Achieving interoperability, use case with the AFarCloud Data Model

An IoT scenario is characterized by the collection of information from devices deployed in the field. However, in most use cases, devices have low resource capabilities and low computational power, which limits the management and production of data under complex and extensive syntax. Therefore, the semantic characterization offered by a data model such as AIM can be a limitation for some of the devices or sensors. However, the STSDaMaS proposal includes a solution to ensure interoperability with services, thanks to a component called *Interoperability through Data Integration* (IDI).

The ISO/IEC 21823-1:2019 standard [39] outlines five levels of interoperability:

- *Transport interoperability* involves establishing a communication infrastructure to exchange data over different transport protocols or networks without loss or corruption of data. It also manages data quality of service (QoS) requirements, such as delivery, timeliness, ordering, durability, lifespan, and fault tolerance.
- *Syntactic Interoperability* is the ability of different IoT systems to exchange data using a common data format or syntax. This ensures that the data are correctly interpreted and processed by the receiving system, without ensuring that the meaning of the information is correctly understood.

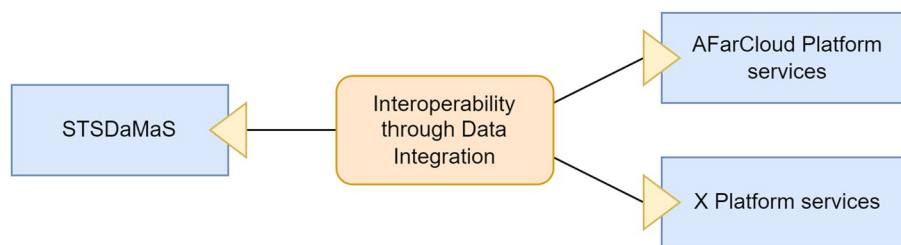


Fig. 4 Interoperability with different platform services through IDI

- *Semantic Interoperability* aims to ensure that all entities attribute the same meaning to the information exchanged by using metadata and shared information models (ontologies).
- *Behavioral Interoperability* aligns and integrates business processes to achieve the expected result of information exchange.
- *Legal or Policy Interoperability* ensures that organizations operating under different legal frameworks, policies, and strategies can exchange information and collaborate effectively. This includes issues such as data privacy, security, and ownership, as well as compliance with industry standards and regulations.

Due to the lack of standards, one of the most common problems in data management systems for precision agriculture lies in the interoperability between models or semantic Interoperability. Agricultural data management systems, as presented in the literature, are designed to ingest and retrieve data against the native model of the system. Such proposals force changes to the core of the system to adapt it to work with other data models.

Some proposals focus on acquiring interoperability between heterogeneous agents or devices [40] or even propose environments for semantic interoperability through specific trait ontologies [41].

The IDI component aims to facilitate access to the functionalities offered by STSDaMaS, ensuring interoperability with different services through its data models. Thanks to this approach, the system remains stable, without the need for core variations to ensure data management and spatio-temporal semantic query resolution. To illustrate the proposal, the native data model on the AFarCloud platform and how STSDaMaS can offer data query management for AFarCloud services through the IDI component are presented (Fig. 4).

The AFarCloud (Aggregate FARMing in the CLOUD) ECSEL JU project [42] presents a distributed platform capable of offering the integration and cooperation of cyber-physical agriculture systems to increase efficiency, productivity, food quality, and animal welfare and reduce costs in agricultural labor. Focused on near real-time data exchange, it offers services for mission management, task automation, decision support, or even data processing and analysis, among others.

Due to this approach, AFarCloud data offer strong temporal and spatial characterization, working with a JSON syntax for data exchange and communication between components.

Table 3 Data model for datasets 1, 2, and 3

		AFarCloud Datasets		
		Observations	Collars	State-Vectors
Dimensions	Time	Timestamp (Epochtime)		
	Geoposition	1 - Latitude and Longitude (Float) 2 - Geohash (String)		
	Semantic	Semantic information regarding data nature, identifiers, and service provide		
Attributes structure	Tags	6 attributes: regarding semantics	5 attributes: regarding semantics	1 attribute: regarding semantics
	Fields	6 attributes: Semantics and geopositioning	15 attributes: Semantics and geopositioning	10 attributes: Semantics and geopositioning
		Float: 5	Float: 8	Float: 9
		String: 1	Boolean: 6 String: 1	String: 1

The AFarCloud repositories comprise three main data sets:

1. Description of observations captured by devices or sensors in the field.
2. Measurements collected by collars installed on the neck of livestock.
3. State vectors of UAVs, autonomous and semi-autonomous vehicles.

Table 3 presents the information provided by the three data sets defined in the repositories of the AFarCloud platform. In addition, the structuring of attributes in the aforementioned Tags and Fields, for the storage of data in the TSDB is included. The set of JSON schemas that define the AFarCloud model is collected in [43].

Thanks to the IDI component, the query service of the proposed framework can be integrated to feed the services or algorithms of the AFarCloud platform. Figure 5 shows the diagram of the IDI component for data translation from an STSDaMaS query response to the AFarCloud data model.

To guarantee interoperability with the services offered by the AFarCloud platform, data preparation and transformation must be carried out from the output of the query interface of the STSDaMaS framework to the input of AFarCloud services. This process guarantees the interoperability of the STSDaMaS service for use in solutions offered by other frameworks.

In IDI, the mapping between the equivalent classes of each ontology is established through an RDF graph. Table 4 shows the mapping between the properties of an observation captured by a sensor between both models. Certain additional rules are established through a properties file, such as the conversion between timestamps in date format to epoch format or the reformulation of certain fields according to certain patterns set by the target model for translation.

Once the mapping rules have been established, the response to the query delivered by the STSDaMas platform is connected to the IDI enabler for the preparation and translation of the data into JSON syntax of the AFarCloud model. In this way, the AIM output (Listing 3) is translated into the AFarCloud model (Listing 4) as input for the services offered by the AFarCloud platform.

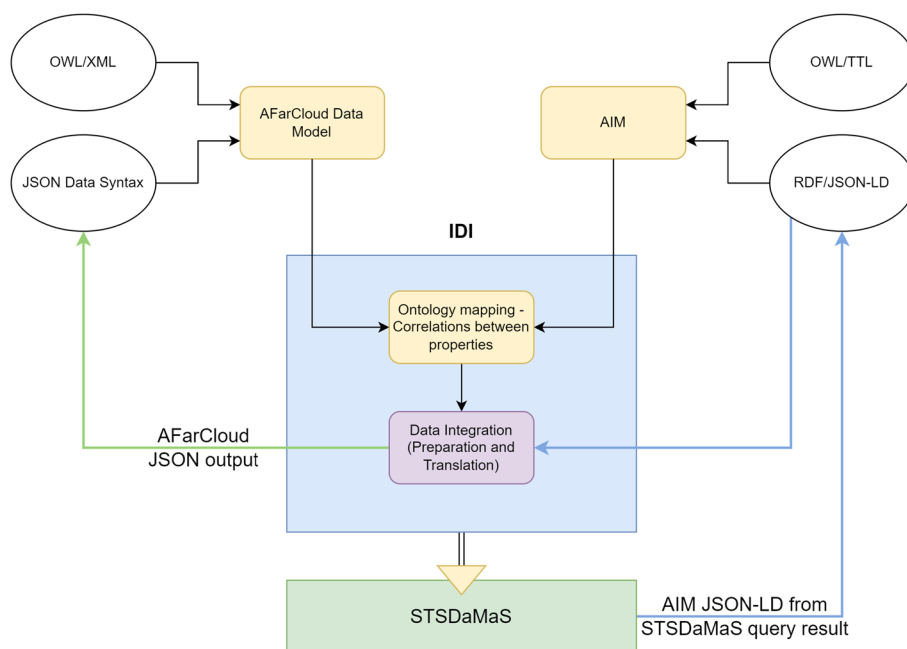


Fig. 5 Interoperability through Data Integration (IDI). AFarCloud Use Case

Table 4 RDF mapping graph

SELECT * WHERE {?subject ?predicate ?object}		
PREFIX sosa:	https://www.w3.org/ns/sosa/	
PREFIX qudt:	https://qudt.org/schema/qudt/	
PREFIX owl:	https://www.w3.org/2002/07/owl#	
PREFIX afc:	https://w3id.org/afarcloud/	
PREFIX afc:	https://www.w3.org/2006/time#	
Subject	Predicate	Object
sosa:hasFeatureOfInterest	owl:hasKey	afc:latitude afc:longitude
sosa:madeBySensor	owl:equivalentClass	afc:AfarcloudSensors
sosa:observedProperty	owl:equivalentClass	sosa:Observation
sosa:resultTime	owl:equivalentClass	time:TemporalEntity
qudt:numericValue	owl:equivalentClass	afc:sensorValue
qudt:unit	owl:equivalentClass	afc:unit

The IDI component operates as a wrapper for the STSDaMaS framework, allowing data input and output to be used for the understanding and coupling of other services.

STSDaMaS Interoperability validation

To validate the interoperability of the STSDaMaS service through the IDI component, an asset tracking and monitoring application has been developed. This application uses the queries offered by the STSDaMaS interface to represent the temporal evolution of observations associated with geolocated devices. However, it cannot understand and

```

{
  "@graph": [
    {
      "@id": "http://www.w3id.org/demeter/pCoord?lat=64.05029&long=24.72468",
      "type": "Point",
      "asWKT": "POINT(24.72468 64.05029)"
    },
    {
      "@id": "http://www.w3id.org/demeter/poi?lat=64.05029&long=24.72468",
      "type": "Feature",
      "hasGeometry": "http://www.w3id.org/demeter/pCoord?lat=64.05029&long=24.72468"
    },
    {
      "@id": "urn:afc:fct:environmentalObservations:UPM:soil:KotipeltoFarmWeatherStation",
      "type": ["SoilSensor"]
    },
    {
      "@id": "urn:afc:fct:environmentalObservations:UPM:soil:KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653",
      "type": "Observation",
      "hasFeatureOfInterest": "http://www.w3id.org/demeter/poi?lat=64.05029&long=24.72468",
      "hasResult": "urn:afc:fct:environmentalObservations:UPM:soil:KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653/q1",
      "madeBySensor": "urn:afc:fct:environmentalObservations:UPM:soil:KotipeltoFarmWeatherStation",
      "observedProperty": "http://www.w3id.org/demeter/obs_air_humidity",
      "resultTime": "2018-01-01T12:36:12Z"
    },
    {
      "@id": "urn:afc:fct:environmentalObservations:UPM:soil:KotipeltoFarmWeatherStation:obs_air_humidity-1579311537653/q1",
      "type": "QuantityValue",
      "identifier": "KotipeltoFarmWeatherStation/obs_air_humidity-1579311537653/q1",
      "numericValue": "3.5",
      "unit": "http://qudt.org/vocab/unit/DEG_C"
    }
  ],
  "@context": ["https://w3id.org/demeter/agri-context.jsonld"]
}

```

Listing 3 Interoperability through Data Integration (IDI). AFarCloud Use Case

process the data modeled using AIM. The application will represent the data only when the query response message is valid against the JSON schemas defined in the AFarCloud model.

To take advantage of the power of spatio-temporal semantic data management offered by the proposed system, the repositories should be nourished with data from observations, sensors, and devices, providing relevant information for each of the defined dimensions. For this experimental case, queries will be performed whose predicate describes the observations associated with environmental and soil measurements, such as temperature and humidity. For this purpose, it will be filtered by the type of service offered. Additionally, the search is limited to sensors installed on mobile devices, including information on batteries and other properties. Correct representation and monitoring of the

```

{
  "sensorData": {
    "resourceId": "KotipeltoFarmWeatherStation",
    "resourceType": "soil",
    "resourceUrn": "urn:afc:fct:
      environmentalObservations:UPM:soil:
      KotipeltoFarmWeatherStation",
    "latitude": 24.72468,
    "longitude": 64.05029,
    "altitude": 0.0
    "resultTime": 1514806572,
    "observations": [{
      "observedProperty": "http://www.w3id.org/
        demeter/obs_air_humidity",
      "uom": "http://qudt.org/vocab/unit/DEG_C"
    }
  ]
},
"multiVariableObservationData": {
  "resourceId": "urn:afc:fct:
    environmentalObservations:UPM:soil:
    KotipeltoFarmWeatherStation",
  "sequenceNumber": 304,
  "observations": [{
    "observedProperty": "http://www.w3id.org/
      demeter/obs_air_humidity",
    "resultTime": 1514806572,
    "result": {
      "value": 3.5
    }
  }
]
}
}

```

Listing 4 Interoperability through Data Integration (IDI). AFarCloud Use Case

temporal and geospatial evolution of devices and associated measurements are offered through the exploitation of the time-series oriented repository.

To build the scenario, real data were extracted from the AFarCloud project repositories and submitted to STSDaMaS via IDI for “translation” and subsequent injection into the repositories. To provide visualization of the evolution of the data in its temporal dimension, injection is performed continuously throughout the experimentation. The monitoring application is connected to the query interface offered by the STSDaMaS framework through the IDI component, which offers a response to queries according to

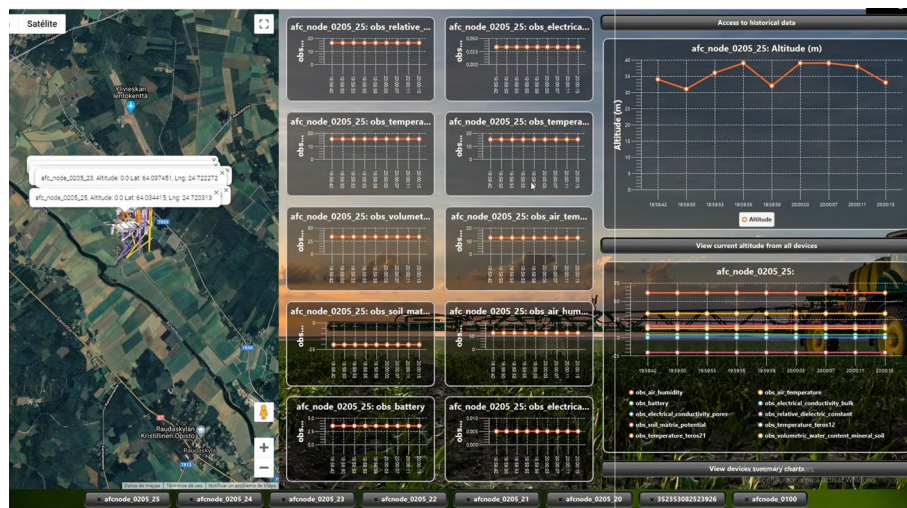


Fig. 6 Monitoring application that displays updated data through STS-DaMaS



(a) Cow tracking

(b) UAVs tracking

Fig. 7 Representation of assets and cow tracking

the AFarCloud data model. The information provided by the monitoring application is represented in Fig. 6.

Thanks to the query interface offered by STSDaMaS, the developed graphical application can represent historical and updated data in a near real-time environment. The application is especially useful for the representation of routes described by assets (see Fig. 7a) or even livestock (see Fig. 7a), through a purely user-oriented interface.

These asset monitoring applications provide added value to users, allowing them to check the current status or historical values captured by their devices in a simple and effective way. Even livestock behavior can be studied to improve animal welfare and optimize production. However, the potential of STSDaMaS lies in the possibility of performing more complex data management, providing answers to user-oriented queries, facilitating decision-making, or nurturing learning models. “[System validation: applying neural network for fodder legume harvesting and bailing prediction](#)” section presents the validation of the system against the need for more elaborate queries for a specific use case: increase in yield in forage crops.

System validation: applying neural network for fodder legume harvesting and baling prediction

The main agricultural sectors include livestock farming and agriculture. However, in many scenarios, the dividing line between the two activities is practically non-existent. An example is the cultivation of crops for livestock feed. High crop quality and optimal harvesting and drying directly impact the quality of livestock, whether for meat or milk.

There is an extensive list of supervised and unsupervised machine learning algorithms [44–47]. Due to the temporal nature offered by the agricultural data and the data management offered by the proposed framework, the use of neural networks has been chosen for the evaluation and validation of the proposal as a “feeder” for training and prediction.

In this chapter, the purpose of this article is to apply the data management framework proposed in the article to train and feed an MLP (Multi-Layer Perceptron) feedforward neural network to predict the optimal moment for harvesting, ensuring optimal drying of the forage legume crop for baling. An MLP neural network has been chosen based on its performance for prediction or forecasting solutions due to its multi-layer architecture. By ensuring optimal conditions for the harvest and drying of the forage, the quality of the product increases greatly, eliminating toxins and the loss of nutrients.

The application of the system for feeding neural networks is proposed to predict the optimal mowing time for the fodder legume crop alfalfa. Alfalfa is a protein source and is characterized as one of the best forage crops for livestock feed. However, the optimal harvest time and the drying time to reduce moisture before packing are key factors in the quality of the harvested fodder and, therefore, in the development of livestock for meat or dairy.

Thanks to the system proposed in this article, the extraction of historical information for neural network training is made possible, as well as a query system for the prediction of observations such as temperature and humidity, which are key for the determination of optimal times for mowing and drying of the fodder crop.

To evaluate the STSDaMaS framework, the experiment focuses on forecasting the ideal harvest time of the fodder legume. After mowing, the desiccation stage begins. Natural desiccation of green forages occurs effectively at temperatures above 15 °C, with relative humidity in the atmosphere below 70%. The most favorable conditions are offered with temperatures above 25 °C and relative humidity below 60%. The objective of fodder drying is to reach a maximum moisture between 18 and 20%, as once packed it is difficult to dry it later. A higher moisture level in the forage, between 20 and 35%, favors the microbial development of fungi, so that, among others, the loss of digestible nutrients, energy, and the possible production of harmful toxins to the animal will critically deteriorate the quality of the product.

The drying process is a critical step to ensure maximum forage quality. However, the most common technique to determine the drying time by the farmer is to wait during a 24–48 h time window until harvest. This technique is not very precise. Other solutions include the use of sensors to constantly monitor product moisture. However, the moisture margin for optimal forage quality is very narrow, requiring constant monitoring costs and high availability for the farmer to carry out the baling at the right time.

Mowing forage without adequate weather conditions will make the drying process impossible. Accurate forecasting of temperature and moisture conditions can ensure an effective forage drying process and provide the farmer with a valuable decision-support solution.

To solve the forage mowing and drying process and to validate the STSDaMaS framework, a neural network is developed to predict the exact moment when the optimal conditions for mowing will occur. The neural network will be fed through the STSDaMaS query interface for training, validation, and production status. For the development of this experiment, the following tasks are performed:

1. Nourish STSDaMaS repositories with real data captured by moisture and temperature sensors in a forage crop.
2. Design and development of an MLP (Multi-Layer Perceptron) feedforward neural network.
3. Neural network training through the STSDaMaS query interface.
4. Evaluation of the prediction provided by the neural network.

The model presented for the validation of STSDaMaS offers the service of temperature and humidity prediction based on historical data. However, the development of an accurate weather prediction model is a complex task due to the variability and susceptibility to sudden changes in the weather. Therefore, as a test evaluation of the system proposed in the article, the model focuses on prediction based on historical data to provide decision support to the farmer or user.

STSDaMaS repository nourishment

Due to the large amount of data collected by hundreds of devices and sensors deployed in more than 10 scenarios across Europe, the AFarCloud repositories are used as a data source.

As discussed in [“Achieving interoperability, use case with the AFarCloud Data Model”](#) section, the data from the AFarCloud repositories are structured and formatted according to their model, so their injection into the STSDaMaS repositories requires the IDI component, which is responsible for providing the interoperability of the service. Once the data are prepared and integrated with the AIM model used by STSDaMaS, they are written to the repositories through the injection interface.

To ensure that the assessment scenario is as close as possible to a real environment, all data contained in one of the AFarCloud scenarios, located in the surroundings of Gamleby, Sweden, are uploaded. Comprising real observations captured by a multitude of sensors between 2020 and 2021. In turn, several active sensors provide service to STSDaMaS to keep humidity and temperature data in the crop up-to-date. In this way, the observations collected between 2020 and 2021 will be used to train the neural network and verify the forecast offered. The updated data will serve as input for the temperature and humidity forecast in the coming days.

Neural network design and development

STSDaMaS will provide the data set to train the neural network developed for the experiment through query interfaces. The first step in developing a suitable neural network for the prediction of humidity and temperature is the correct characterization of the data set.

The prediction is oriented to time-series forecasting, so it must be taken into account that the data set is time-dependent (this breaks the requirement of linear regression that its observations be independent) and has some kind of stationarity (or tendency, for instance, according to the time of day, month, or time of year).

For this environment, a simple FeedForward² neural network architecture is used, with 7 inputs, a hidden layer of 7 neurons, and an output of a single neuron. Based on the available data set, the frequency of measurement of agricultural devices and sensors, and the evolution trend throughout the day, the 7/7/1 architecture has been chosen due to the number of days of the week. It is intended to forecast humidity and temperature on the next consecutive day, displacing the input for the prediction of the following days. The hyperbolic tangent activation function is chosen since the values will be normalized between -1 and 1 . The Adam optimizer and the loss metric Mean Absolute Error are configured. To calculate the accuracy, the Mean Squared Error is used, since the prediction will be a continuous value and not a discrete one.

Neural network training and validation

A total of 8,430,118 observations captured by static and mobile devices are available in the repository, together with the semantic description of the devices and the farm or crop with which they are associated. This data set comprises observations from sensors, observations from sensors embedded in collars fitted to the neck of livestock, and state vectors from vehicles or UAVs. The values associated with temperature and atmospheric humidity for the forage crop comprise a total of 511,738 time points. For neural network training, consecutive queries are made to the STSDaMaS Data Query interface to extract the timestamps and values of humidity and temperature captured for the forage crop.

For neural network training, it has been decided to collect the humidity observations in sets of 7 consecutive days and forecast the humidity values on the eighth day. Therefore, requests to the STSDaMaS query interface will extract the samples for 8 consecutive days, starting January 1, 2020, until September 1, 2022 (the rest of the data, until January 1, 2022, will be used for validation). To train the network with backpropagation [48], the time series provided by the output of the STSDaMaS query interface must be converted into a “supervised type problem”.

In Fig. 8 the values obtained after 40 EPOCHS in model training, directly fed by the STSDaMaS query interface, are shown.

Once the model has been trained, it is validated with the samples that have not been used in the training set. For the validation of the humidity prediction, the samples captured between 1 September 2021 and 1 January 2022 are available, obtaining the following graph as a result of the validation (see Fig. 9). The measurements shown in green

² Also denominated MLP for Multi-Layered Perceptron.

```

Epoch 38/40
28608/28608 [=====] - 33s 1ms/step - loss: 0.0138 - mse: 0.0019 - val_loss: 0.0131 - val_mse: 4.9646e-04
Epoch 39/40
28608/28608 [=====] - 35s 1ms/step - loss: 0.0137 - mse: 0.0019 - val_loss: 0.0132 - val_mse: 4.7382e-04
Epoch 40/40
28608/28608 [=====] - 34s 1ms/step - loss: 0.0136 - mse: 0.0019 - val_loss: 0.0126 - val_mse: 4.6247e-04
    
```

Fig. 8 Training parameters

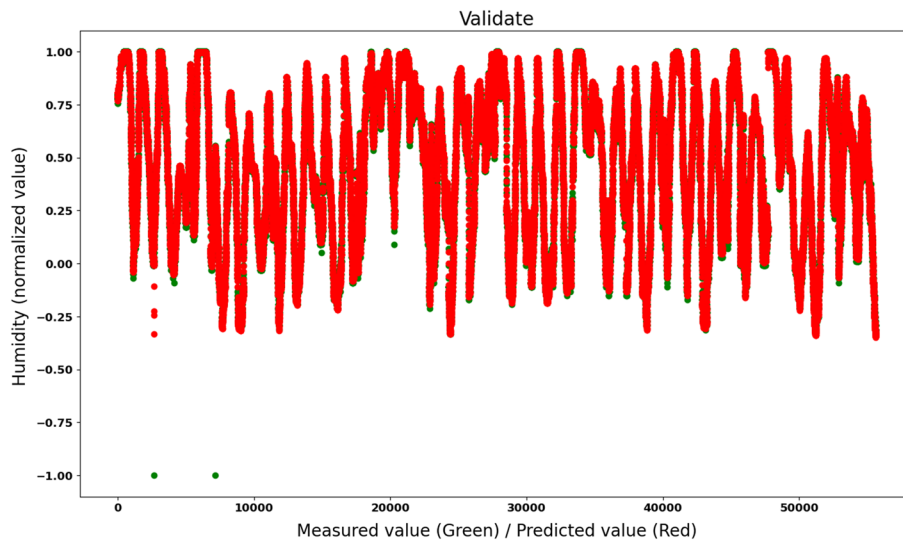


Fig. 9 Moisture prediction by the neural network fed from STSDaMaS

correspond to the actual value obtained from the sensors, and the measurements shown in red correspond to the values provided by the model.

The process is repeated for the temperature data, performing sequential queries through the STSDaMaS query interface, and training the model. Finally, the remaining set of temperature samples from the training set is used for model validation. Figure 10 shows the result of the validation of the temperature forecast between 1 September 2021 and 1 January 2022. The values represented in green correspond to the actual measurement, and the values represented in red correspond to the prediction provided by the model.

Humidity and temperature forecasting for predicting forage desiccation

Once the neural network has been trained and validated, it is time to evaluate its response in the production environment. Given the following case: A certain crop has developed correctly and the user or farmer wants to know the optimal time for harvesting, ensuring that the atmospheric humidity and temperature conditions are correct for the drying of the fodder and its subsequent packaging. For this purpose, the neural network is fed through the STSDaMaS query interface to forecast the evolution of humidity and temperature in the following days.

To provide a useful forecast for the farmer, historical data are queried for the previous 30 days and the evolution of humidity and temperature for the next 7 days is predicted.

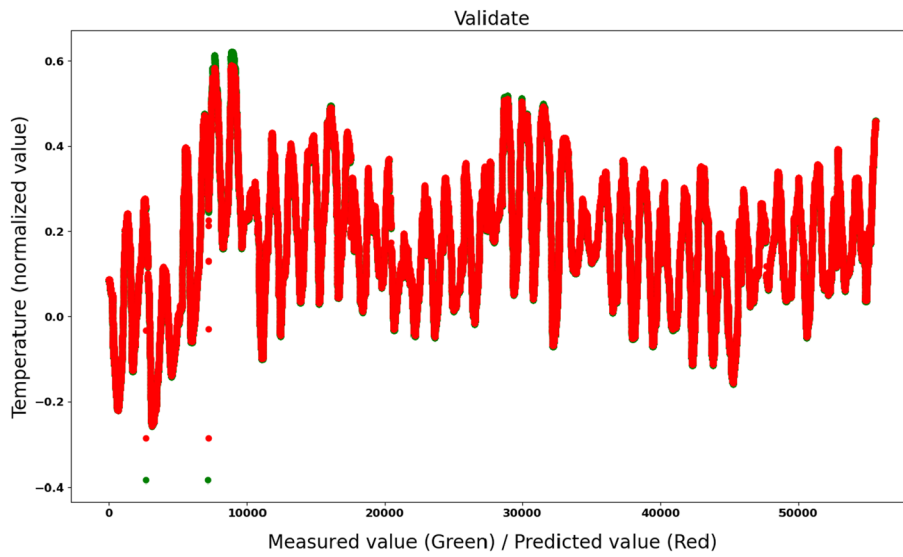
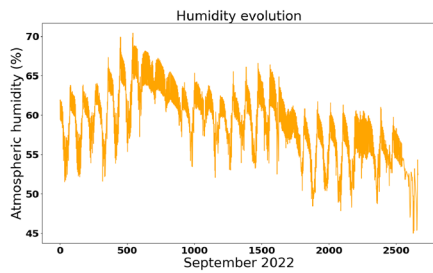
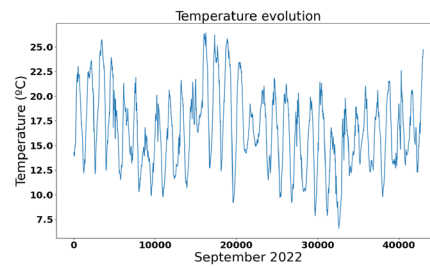


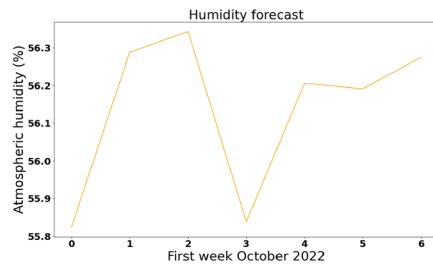
Fig. 10 Temperature prediction using the neural network fed by STSDaMaS



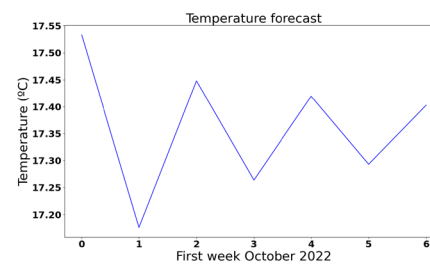
(a) Humidity (%) evolution in September 2022.



(b) Temperature (°C) evolution in September 2022.



(c) Humidity forecast in the first week of October



(d) Temperature forecast in the first week of October

Fig. 11 Evolution of temperature and humidity provided by STSDaMaS and forecast provided by NN

If the average humidity and temperature values are in the right range for effective forage drying, harvesting will be recommended, and the ideal day and time to start will be indicated. Figure 11 shows the evolution of humidity Fig. (11a) and temperature Fig. (11b) for September. For the forecast offered by the neural network, the average humidity Fig. (11c) and temperature Fig. (11d) for the first week of October are shown.

The forecast provided by the neural network determines that the first week of October presents optimal conditions for the forage drying stage. It is recommended that the mowing be performed from 11:00 A.M. on the first day of October to take advantage of the temporary window of optimal weather conditions. In turn, through the STSDaMaS query system, the current observations captured by the field devices will be represented in a near real-time environment, providing up-to-date information.

Through the set of forage moisture observations, the neural network can be trained and validated to predict the exact moment for the baling of the forage, ensuring optimal forage quality conditions. However, the objective of this experiment is to validate the STSDaMaS framework, the data management offered, and the interoperability offered by its service; therefore, the implementation and improvement of decision support systems or monitoring or automation applications, among others, will be the target of future developments and research.

The work carried out for the validation of the model aims to show the effectiveness and applicability of the proposal to feed prediction models that support decision-making. However, the accuracy obtained in the development of the neural network presented in this article could be optimized through some changes in its development: (i) Development of a new model taking the time stamp as a new input variable. (ii) Using the time stamp as an additional variable but using it with *Embeddings*.

Temperature and humidity vary according to the year, month, and time of day at which the measurement is taken. Therefore, adding the time stamp as an input variable to the model can substantially improve the accuracy of the forecast, reducing the number of data needed for training. Through the STSDaMaS query interface, the evolution of atmospheric humidity presented for the first week of August 2020 and 2021 Fig. (12a) and the temperatures presented for the last week of July 2020 and 2021 Fig. (12b) are shown, demonstrating a clear similarity in their evolution.

Thanks to the neural network developed, the STSDaMaS framework has been validated as a spatio-temporal semantic data management system. The framework can provide service for machine learning models, decision-making algorithms, and monitoring agents and crop conditions. In turn, the development of the IDI component provides the service interoperability necessary for its inclusion in other precision agriculture solutions.

The STSDaMaS framework could be of high utility in other application environments such as Industry 4.0, but for this, both its model, query functions, and repository structure must be adapted.

Conclusion

The authors of this article propose an innovative framework that, after describing the spatio-temporal semantic character of the data collected in the precision agriculture domain, offers the service of advanced data management and querying. The core of the STSDaMaS framework focuses on repository distribution in a knowledge base and time-series oriented database. Due to the repository distribution, STSDaMaS offers high performance in query resolution and complete data description. The distributed architecture of the repository offers: (A) Resolution of temporal and/or spatial queries through access to a historical database. (B) Resolution of semantic queries, through

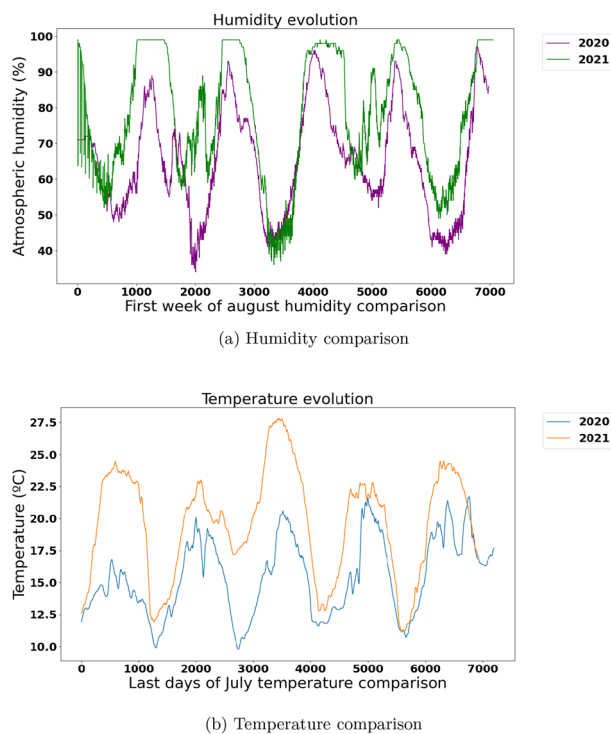


Fig. 12 Comparisons of humidity and temperature between 2020 and 2021 (output provided by the STSDaMaS query interface)

SPARQL on a triple-store, allowing the extraction of information and semantic description of data, as well as its relations with other individuals. (C) Resolution of complex queries through the relationships between the knowledge database and the historical database employing devices and their captured observation UIDs.

STSDaMaS proposal constitutes a spatio-temporal semantic data management system that offers service interoperability through semantic interoperability. This innovative approach is particularly important in view of the European data strategy [49], which aims to make Europe a leader in the data-driven society.

To guarantee the interoperability of the service offered by the proposed framework, the *Interoperability through Data Integration* component is described. The proposal has been validated by demonstrating the interoperability between the STSDaMaS framework, which works with DEMETER’s AIM data model, and the AFarCloud platform and its data model.

In this way, projects such as DEMETER, which does not have a management system or native repositories, can include data ingestion and query services. In turn, it guarantees interoperability to increase the number of solutions available within the project framework, as shown by guaranteeing interoperability with the AFarCloud platform. The data preparation and transformation between the AIM and AFarCloud data model allow solutions such as decision-making or machine learning models fed by data in the AFarCloud framework, to be integrated with DEMETER through the STSDaMaS, increasing the number of available solutions.

The article presents a specific use case for the validation of the proposal as input for machine learning models. It describes the complexity and importance of the task of mowing and desiccating legume fodder crops and the possibility of making a forecast for its correct development. A neural network is developed and fed through STSDaMaS for training and validation. An accurate weather forecast is provided for the recommendation of forage mowing, ensuring effective drying. The management of these crops is essential to ensure the quality of livestock feed, which includes a direct impact on animal welfare and the quality of beef or dairy products.

The proposal presented in this article has been developed in the framework of the European DEMETER project, in collaboration with and expansion of the research and development carried out in the European AFarCloud project. This scenario allows for experimentation and validation of the proposal against a real scenario and data. The novel data management framework proposed in this article gathers the needs exposed in the precision agriculture framework, the literature, and the needs of specialists, engineers, farmers, or agricultural administrators themselves. The proposal has been applied as an spatio-temporal semantic data manager with service interoperability, enabling the annexation of AFarCloud project solutions in the DEMETER architecture.

Future work

The design and implementation of the STSDaMaS system focus on the application domain of precision agriculture. In this state, its adaptability to other application domains such as Smart Industry, Smart Cities, or Smart healthcare would imply modifications or extensions to the native data model of the system or the adaptability of some of its query or data injection functions. However, precision agriculture presents a scenario with common characteristics to those of other domains versed in sensor networks. Our current line of research focuses on extending the proposal to ensure its adaptability to other application domains of the Internet of Things framework.

The temperature and humidity prediction model developed for the validation of the proposed system offers a valuable solution to the farmer or user. Once the proposal has been validated, the resilience of the model to abrupt changes in weather conditions could be improved. To increase prediction accuracy against sudden changes in climatic conditions, the inclusion of other data sources outside of the specific assessment scenario should be considered. Data provided by weather stations or data captured by devices deployed in neighboring scenarios offer the ability to anticipate abrupt changes in current conditions. The evaluation of observations such as precipitation amounts or wind speed and direction is currently under study.

Acknowledgements

This publication is part of the Horizon 2020 DEMETER project (857202) supported by the European Union.

Author contributions

Conceptualization, MSEP and NLM; Data curation, MSEP and JFMO; Formal analysis, MSEP and VHD; Funding acquisition, JFMO; Investigation, MSEP, NLM and VHD; Methodology, MSEP, NLM and JFMO; Project administration, MSEP and JFMO; Software, MSEP, NLM and JFMO; Supervision, MSEP and JFMO; Validation, MSEP, NLM, VHD and JFMO; Visualization, VHD; Writing—original draft, MSEP; Writing—review and editing, MSEP, NLM, VHD and JFMO. All authors read and approved the final manuscript.

Funding

This work was supported by the DEMETER project, which has received funding from INDUSTRIAL LEADERSHIP—Leadership in enabling and industrial technologies—Information and Communication Technologies (ICT) under grant

agreement No. 857202. ICT receives support from the European Union's Horizon 2020 research and innovation program, and Italy, Luxembourg, Spain, Germany, Greece, United Kingdom, Norway, Czechia, Montenegro, Belgium, Romania, Ireland, Finland, Serbia, Portugal, Poland, Georgia, and Slovenia.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 10 January 2023 Accepted: 11 April 2023

Published online: 28 April 2023

References

1. dpicampaigns: Take Action for the Sustainable Development Goals. <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Accessed 11 Oct 2022.
2. Animal welfare. https://food.ec.europa.eu/animals/animal-welfare_en. Accessed 11 Oct 2022.
3. Home | Food and Agriculture Organization of the United Nations. <https://www.fao.org/home/en>. Accessed 11 Oct 2022.
4. International Fund for Agricultural Development. <https://www.ifad.org/en/>. Accessed 11 Oct 2022.
5. Agricultural research for development. <https://www.ifad.org/en/agricultural-research-for-development>. Accessed 11 Oct 2022.
6. The State of Food Security and Nutrition in the World 2020 | FAO | Food and Agriculture Organization of the United Nations. <https://doi.org/10.4060/CA9692EN>. <https://www.fao.org/publications/sofi/2020/en/>. Accessed 11 Oct 2022.
7. Martin: Goal 2: Zero Hunger. <https://www.un.org/sustainabledevelopment/hunger/>. Accessed 11 Oct 2022.
8. Kim W-S, Lee W-S, Kim Y-J. A Review of the Applications of the Internet of Things (IoT) for Agricultural Automation. *J Biosyst Eng*. 2020;45(4):385–400. <https://doi.org/10.1007/s42853-020-00078-3>.
9. Tiddi, I., Lécué, F., Hitzler, P. (eds.): Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges. Studies on the semantic web, vol. volume 047. IOS Press, Amsterdam (2020)
10. Zinke, C., Ngomo, A.-C.N.: Discovering and Linking Spatio-Temporal Big Linked Data. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 411–414 (2018). <https://doi.org/10.1109/IGARSS.2018.8519025> ISSN: 2153-7003
11. N. Zhang, R.K.T.: APPLICATIONS OF A FIELD-LEVEL GEOGRAPHIC INFORMATION SYSTEM (FIS) IN PRECISION AGRICULTURE. <https://doi.org/10.13031/2013.6829> Accessed 2023-01-04
12. Burlacu G, Costa R, Sarraipa J, Jardim-Golcalves R, Popescu D. A Conceptual Model of Farm Management Information System for Decision Support. In: Camarinha-Matos LM, Barrento NS, Mendonça R, editors. Technological Innovation for Collective Awareness Systems. IFIP Advances in Information and Communication Technology. Berlin: Springer; 2014. p. 47–54. https://doi.org/10.1007/978-3-642-54734-8_6.
13. Leroux C, Jones H, Pichon L, Guillaume S, Lamour J, Taylor J, Naud O, Crestey T, Lablee J-L, Tisseyre B. GeoFIS: An Open Source, Decision-Support Tool for Precision Agriculture Data. *Agriculture*. 2018;8(6):73. <https://doi.org/10.3390/agriculture8060073>.
14. Li R, He H, Wang R, Huang Y, Liu J, Ruan S, He T, Bao J, Zheng Y. JUST: JD Urban Spatio-Temporal Data Engine. In: 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 1558–1569 (2020). <https://doi.org/10.1109/ICDE48307.2020.00138>. ISSN: 2375-026X.
15. Ren Y, Huang D, Wang W, Yu X. BSMD: A blockchain-based secure storage mechanism for big spatio-temporal data. *Fut Gen Comput Syst*. 2023;138:328–38. <https://doi.org/10.1016/j.future.2022.09.008>.
16. Isomura A, Shigematsu N, Ueno I, Oki N, Arakawa Y. Real-time Spatiotemporal Data-management Technology (Axispot™). *NTT Techn Rev* 20(7), 54–60 (2022). <https://doi.org/10.53829/ntr202207fa8>.
17. Deeken H, Wiemann T, Lingemann K, Hertzberg J. SEMAP - a semantic environment mapping framework. In: 2015 European Conference on Mobile Robots (ECMR), pp. 1–6. <https://doi.org/10.1109/ECMR.2015.7324176>. 2015.
18. Deeken H, Wiemann T, Hertzberg J. A Spatio-Semantic Model for Agricultural Environments and Machines. In: Mouhoub M, Sadaoui S, Ait Mohamed O, Ali M, editors. Recent Trends and Future Technology in Applied Intelligence. Lecture Notes in Computer Science. Cham: Springer; 2018. p. 589–600. https://doi.org/10.1007/978-3-319-92058-0_57.
19. Deeken H, Wiemann T, Hertzberg J. A spatio-semantic approach to reasoning about agricultural processes. *Appl Intell*. 2019;49(11):3821–33. <https://doi.org/10.1007/s10489-019-01451-2>.
20. Wisnubhadra, I., Baharin, S., Herman, N., Open Spatiotemporal Data Warehouse for Agriculture Production Analytics. *Int J Intell Eng Syst* 13(6), 419–431 (2020). <https://doi.org/10.22266/ijies2020.1231.37>
21. Murlidharan S, Shukla VK, Chaubey A. Application of Machine Learning in Precision Agriculture using IoT. In: 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), pp. 34–39. 2021. <https://doi.org/10.1109/ICIEM51511.2021.9445312>.

22. Sharma A, Jain A, Gupta P, Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*. 2021;9:4843–73. <https://doi.org/10.1109/ACCESS.2020.3048415>.
23. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine Learning in Agriculture: A Review. *Sensors*. 2018;18(8):2674. <https://doi.org/10.3390/s18082674>.
24. Meshram V, Patil K, Meshram V, Hanchate D, Ramkteke SD. Machine learning in agriculture domain: A state-of-art survey. *Artif Intell Life Sci*. 2021;1: 100010. <https://doi.org/10.1016/j.aillsci.2021.100010>.
25. Mee J, Boyle L. Assessing whether dairy cow welfare is “better” in pasture-based than in confinement-based management systems. *New Zealand Veter J*. 2020;68(3):168–77. <https://doi.org/10.1080/00480169.2020.1721034>.
26. Mwangi FW, Charmley E, Gardiner CP, Malau-Aduli BS, Kinobe RT, Malau-Aduli AEO. Diet and genetics influence beef cattle performance and meat quality characteristics. *Foods*. 2019;8(12):648. <https://doi.org/10.3390/foods8120648>.
27. Moore KJ, Lenssen AW, Fales SL. Factors Affecting Forage Quality. In: *Forages*, pp. 701–717. Wiley, New York. 2020. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119436669.ch39> Accessed 5 May 2022.
28. Johann AL, de Araújo AG, Delalibera HC, Hirakawa AR. Soil moisture modeling based on stochastic behavior of forces on a no-till chisel opener. *Computers Electr Agric*. 2016;121:420–8. <https://doi.org/10.1016/j.compag.2015.12.020>.
29. Nahvi B, Habibi J, Mohammadi K, Shamshirband S, Al Razgan OS. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. *Computers Electr Agric*. 2016;124:150–60. <https://doi.org/10.1016/j.compag.2016.03.025>.
30. Vamseekrishna A, Nishitha R, Kumar TA, Hanuman K, Supriya CG. Prediction of Temperature and Humidity Using IoT and Machine Learning Algorithm. In: Bhattacharyya S, Nayak J, Prakash KB, Naik B, Abraham A, editors. *International Conference on Intelligent and Smart Computing in Data Analytics Advances in Intelligent Systems and Computing*. Singapore: Springer; 2021. p. 271–9. https://doi.org/10.1007/978-981-33-6176-8_30.
31. Kumar YJN, Spandana V, Vaishnavi VS, Neha K, Devi VGRR. Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. In: *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 736–741. 2020. <https://doi.org/10.1109/ICCES48766.2020.9137868>.
32. Jena A. Apache Jena - Reasoners and rule engines: Jena inference support. <https://jena.apache.org/documentation/inference/#rules>. Accessed 29 Mar 2023.
33. W3C: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. <https://www.w3.org/Submission/SWRL/> Accessed 2023-03-29
34. W3C: RIF Overview (Second Edition). <https://www.w3.org/TR/rif-overview/>. Accessed 29 Mar 2023.
35. W3C: Notation 3 Logic. <https://www.w3.org/DesignIssues/Notation3.html>. Accessed 29 Mar 2023.
36. W3C: RuleML - W3C RIF-WG Wiki. <https://www.w3.org/2005/rules/wg/wiki/RuleML>. Accessed 29 Mar 2023.
37. Demeter - EMPOWERING FARMERS. 2019. <https://h2020-demeter.eu/>. Accessed 10 Nov 2023.
38. Palma R, Roussaki I, Döhmen T, Atkinson R, Brahma S, Lange C, Routis G, Plociennik M, Mueller S. Agricultural Information Model. In: Bochtis DD, Sørensen CG, Fountas S, Moysiadis V, Pardalos PM, editors. *Information and Communication Technologies for Agriculture-Theme III: Decision Springer Optimization and Its Applications*. Cham: Springer; 2022. p. 3–36. https://doi.org/10.1007/978-3-030-84152-2_1.
39. 14:00-17:00: ISO/IEC 21823-1:2019. <https://www.iso.org/standard/71885.html>. Accessed 11 Mar 2022.
40. Khatoun, P.S., Ahmed, M.: Semantic Interoperability for IoT Agriculture Framework with Heterogeneous Devices. In: Gunjan, V.K., Zurada, J.M. (eds.) *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*. *Advances in Intelligent Systems and Computing*, pp. 385–395. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-7234-0_34
41. Aydin S, Aydin MN. Semantic and Syntactic Interoperability for Agricultural Open-Data Platforms in the Context of IoT Using Crop-Specific Trait Ontologies. *Appl Sci*. 2020;10(13):4460. <https://doi.org/10.3390/app10134460>.
42. Castillejo P, Johansen G, Cürüklü B, Bilbao-Arechabala S, Fresco R, Martínez-Rodríguez B, Pomante L, Rusu C, Martínez-Ortega J-F, Centofanti C, Hakojarvi M, Santic M, Häggman J. Aggregate Farming in the Cloud: The AFarCloud ECSEL project. *Microprocess Microsyst*. 2020;78:10321103218. <https://doi.org/10.1016/j.micpro.2020.103218>.
43. Parte MSE, Serrano SL, Díaz, VH, Martínez-Ortega J-F. grys-upm/Spatio-Temporal-Semantic Data Model for Precision Agriculture. Zenodo (2022). <https://doi.org/10.5281/zenodo.7263254>. <https://zenodo.org/record/7263254> Accessed 29 Oct 2022.
44. Ray S. A Quick Review of Machine Learning Algorithms. In: *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 35–39. 2019. <https://doi.org/10.1109/COMITCon.2019.8862451>
45. Mahesh B. Machine Learning Algorithms –A Review. *International Journal of Science and Research (IJSR)*. 2019. <https://doi.org/10.21275/ART20203995>.
46. Bonaccorso G. *Machine Learning Algorithms*. New York: Packt Publishing Ltd; 2017.
47. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315. 2016.
48. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533–6. <https://doi.org/10.1038/323533a0>.
49. Communication from the commission to the European Parliament, the council, the European Economic and Social Committee and the committee of the regions. *A European strategy for data*. 2020). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX>. Accessed 17 Oct 2022.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.