Mo.Re.Farming: A hybrid architecture for tactical and strategic precision agriculture

Mo.Re.Farming: A Hybrid Architecture for Tactical and Strategic Precision Agriculture[☆]

Enrico Gallinucci, Matteo Golfarelli, Stefano Rizzi

DISI - University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

Abstract

In this paper we propose an innovative architecture, called *Mo.Re.Farming*, for handling agricultural data in an integrated fashion and supporting decision making in the precision agriculture domain. This architecture is oriented to data analysis and is inspired by Business Intelligence 2.0 approaches. It is hybrid in that it couples traditional and big data technologies to integrate heterogeneous data, at different levels of detail, from several owned and open data sources; its goal is to demonstrate that such integration is feasible and beneficial in supporting situ-specific and large-scale analyses. The proposed architecture has been developed in the context of the Mo.Re.Farming project, aimed at providing a Decision Support System for agricultural technicians in the Emilia-Romagna region and to enable analyses related to the use of water and chemical resources. The architecture is fully deployed and serves as a hub for agricultural data in Emilia-Romagna; the integrated data are made available in open access mode and can be accessed through web interfaces and through a set of web services. The paper describes the architecture from the technological and functional points of view and discusses the Mo.Re.Farming project outcomes and lessons learnt.

Keywords: BI 2.0, precision agriculture, data integration

1. Introduction

With the rise of precision agriculture, the world of agriculture has become a major producer and consumer of data. Indeed, recent technologies allow satellite images and sensor data to be generated with higher detail and frequency [1]; the set of available open data regularly increases in both quantity and quality, and new approaches to data collection, such as *crowd sensing* [2], are applied to the agriculture area as well. Properly handling such a mass of data requires

[☆]Partially supported by the Mo.Re.Farming Project (`www.morefarming.it/`) funded by the POR FESR Program 2014-2020.

Email addresses: `enrico.gallinucci2@unibo.it` (Enrico Gallinucci), `matteo.golfarelli@unibo.it` (Matteo Golfarelli), `stefano.rizzi@unibo.it` (Stefano Rizzi)

emerging digital technologies, such as big data and IoT, to be adopted. The interest in adopting big data approaches for precision agriculture and precision farming is confirmed by increasing research activities in this area [3]. However, a careful analysis of the 34 research projects surveyed in [3] shows that most efforts are focused on applying machine learning techniques to ad hoc agricultural datasets, whereas data collection and integration systems have attracted less interest —which may give rise to the problem of *information silos* (i.e., information that can be hardly shared and reused), a phenomenon that has already been observed in other contexts [4].

Although the definition of a comprehensive and integrated architecture for precision agriculture has been poorly addressed in the scientific literature, both free and commercial data services are already available to obtain high-value data such as recommendations for irrigation and vegetation indices. While most of these solutions are based on web applications to deliver data services, they strongly differ in the way data are stored, processed, and made available, as well as in the type of data provided and in the professional figures and services they are oriented to.

In this paper we propose an innovative architecture, called *Mo.Re.Farming* (*MOnitoring and REmote system for a more sustainable FARMING*), for handling agricultural data in an integrated fashion. This architecture is oriented to data analysis and is inspired by Business Intelligence 2.0 (BI) approaches [5]; it is hybrid in that it couples traditional and big data technologies to integrate heterogeneous data, at different levels of detail, from several owned and open data sources. Using the BI terminology [6], we distinguish between *tactical* and *strategic* services provided by Mo.Re.Farming.

- **Tactical services** typically exploit data from a limited area and within a restricted time-span to provide detailed information to the users. An example is the current vegetation index of a specific field, which can be used to modulate the quantity of fertilizer to be spread on its surface.

- **Strategic services** aggregate and analyze data from broader areas, spanning on longer time intervals. An example is the time series of the average vegetation index for all the corn fields in the different provinces of a given region during the whole corn farming season.

Clearly, the production of these two kind of information involves a different quantity of raw data and a different level of detail. As agreed in the BI literature, these differences call for separated repositories and schemata to properly store information for the tactical and strategic levels.

A further feature of precision agriculture systems is the inherent presence of spatial information such as georeferenced satellite images, maps of fields, positions of sensors on the ground, and so on. To handle this feature, the adoption of a spatially enabled technology, typically a Geographic Information System (GIS), is required. In Mo.Re.Farming we exploit georeferencing as the basis to carry out an integration of the different data sources.

The proposed architecture, developed in the context of the Mo.Re.Farming project, aims at providing a Decision Support System (DSS) for agricultural technicians in the Emilia-Romagna region (Italy) and to enable analyses related to the use of water and chemical resources. The Mo.Re.Farming project provided not only the requirements for the architecture but also a case study to test it; thus, in this paper we describe the architecture from the technological and functional points of view with specific reference to its deployment within the project. In particular we focus on its most innovative elements: its hybrid features, the data representation, the exploitation of open data information, and the overall integration process. The deployed architecture serves now as a hub for agricultural data in the Emilia-Romagna region; the integrated data are made available in open access mode and can be accessed through web interfaces and through a set of web services.

In a previous version of the paper [7] we gave a preliminary description of the architecture. Here we further elaborate on it mainly by adding:

1. a more detailed and comprehensive description of the architecture and of the technical solutions adopted;
2. a comparison of the possible alternative technological solutions;
3. an evaluation of the performance of the processes used to feed the different levels of the architecture.

The paper outline is as follows. In Section 2 we discuss the project goals and main features, while in Section 3 we present the underlying architecture. Section 4 describes in detail the data models and the main processes involved. Section 5 describes the user interface for the system, and Section 6 discusses the performance of the ETL processes. Section 7 summarizes the related literature. Finally, in Section 8 we discuss the main lessons learnt from the Mo.Re.Farming project.

2. The Mo.Re.Farming Project

The main goal of the Mo.Re.Farming project is to verify the feasibility of a data integration approach to deliver precision agriculture services to a plethora of different stakeholders. The idea from which the project comes is that the higher the number of effectively integrated data sources, the higher the number of services to be delivered. In other words, *putting together information useful for single services enables and empowers further services.* The data sources that are currently ingested in Mo.Re.Farming are listed below:

- **Satellite images**: images are taken from Sentinel-2 satellites, designed to deliver land remote sensing data that are central to the European Commission Copernicus program [8]. The Sentinel-2 mission consists of two satellites developed to support vegetation, land cover, and environmental monitoring. The Sentinel-2 MultiSpectral Instrument (MSI) acquires 13

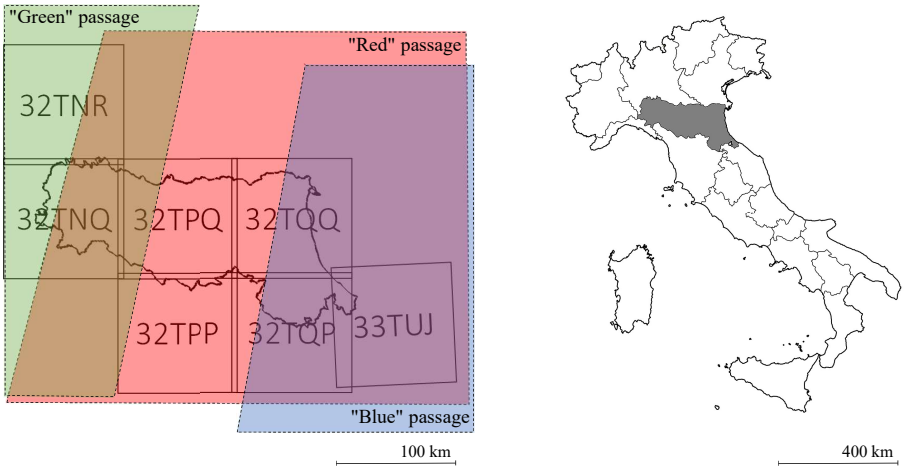


Figure 1: On the left, the seven 100 km^2 granules covering the Emilia-Romagna region; the three layers (green, red, and blue) represent the portions covered by the satellites with each passage. On the right, the position of Emilia-Romagna in Italy

spectral bands ranging from Visible and Near-Infrared (VNIR) to Shortwave Infrared (SWIR) wavelengths along a 290 km orbital swath. Satellite images are provided in the L-1C format, meaning that they are affected by the reflection of solar light against the atmosphere; they provide 12 bands and a maximum spatial resolution of 10 meters. The spherical surface of the Earth is subdivided into partially overlapped squares of 100 km^2 (i.e., *granules*) according to the USA National Grid (`www.fgdc.gov/usng`), and each image corresponds to a granule. The Emilia-Romagna region is covered by a total of 7 granules: 32TNR, 32TNQ, 32TPQ, 32TPP, 32TQQ, 32TQP, 33TUJ. As shown in Figure 1, the satellites make three kinds of passages during which they never cover the whole area, but only a specific portion of it. The current frequency is 2-3 passages per week, where each passage (i.e., the green, red, and blue layers in the figure) is made every 10 days. Images occupy up to 1GB and they also contain quality indicators, auxiliary data, and metadata to enable cloud screening, georeferencing, and atmospheric corrections.

- **Field Sensors**: part of the Mo.Re.Farming project is aimed at developing field sensors to complement satellite data with on-the-ground data. In particular, two sensors have been developed and installed on a set of sample fields. Figure 2 shows the smart pheromone trap and the waveguide-based spectrometry at the core of the humidity sensor. Unlike traditional sensors based on impedance measures, where probes should be inserted into the terrain, the Mo.Re.Farming humidity sensor simply uses a waveguide faced to the soil surface [9]. One humidity value per hour is collected. In


Figure 2: An in-field prototype of the smart pheromone trap (left) and the waveguide-based spectrometry at the core of the humidity sensor (right)

the smart pheromone trap, insects are captured through an adhesive strip with pheromones. A smart camera inside the trap captures one image per day and analyses it to classify and count culture-specific insects (e.g., Grapholita molesta). All field sensors are connected through a GPRS network to the Mo.Re.Farming server and data are downloaded only daily for energy saving purposes.

- **Crop Register**: yearly filled by farmers, it includes —for each rural land— a 49-valued classification of crops (e.g., tomato and forage) and a binary information about irrigation of the fields. The crop register is not available as open data.

- **Administrative boundaries**: a vector layer including municipal, provincial, and regional boundaries. This layer is freely downloadable from the website of the Italian institute for statistics (ISTAT, `www.istat.it/it/archivio/222527`).

- **Rural Land Register**: a vector layer including municipal, field, and farm boundaries tagged with additional information such as field surface. This layer is made available by the Regional agency for agricultural funding (AGREA, `agrea.regione.emilia-romagna.it/`) and CER regional associations (`www.consorziocer.it`).

- **Weather data**: a vector layer with daily data about minimum, maximum, and average temperatures and rainfall on a regional grid of 858

Table 1: A list of the services possibly enabled by the Mo.Re.Farming architecture

Service	*Description*	*Stakeholder*	*Service level*
Infestation alerting	Early detection of harmful insects	Agr. technician, Farmer	Tactic
Monitoring of the vegetation level	Define and apply automatic and differentiate treatments (e.g., irrigation, fertilization) on specific fields based on the diagnosis of colture growth	Agr. technician, Farmer	Tactic
Irrigation forecast	Identify the need for irrigation in the succeeding days for a specific field, based on vegetation indices, field humidity and weather forecast	Agr. technician, Farmer	Tactic
Irrigation monitoring	Check the proper use of irrigation resources with respect to the consumption level declared by farmers	Water procurement and management agencies	Tactic
Drought monitoring	Evaluation of the historical and current state of drought during drought crisis	Councilor for agriculture	Strategic
Optimization of water resources	Planning of irrigation resources according to the crops and to the farmer irrigation declarations, possibly compared with decisions made in previous years	Water procurement and management agencies	Strategic
Analysis of the vegetation level	Culture or wide-area long term analysis of the vegetation level	Councilor for agriculture	Strategic

sensors. This layer is made available by the ARPAE regional association (`www.arpae.it`).

All the data listed above are inherently spatial, thus a natural way to analyze and visualize them is through a GIS system.

Table 1 shows a non-exhaustive list of the services possibly enabled by the data described above, including the name and description of the services as well as the stakeholder to whom it is directed and the level of the services.

As to quality we remark that, based on our experience, the selected sources have all proven to be highly reliable. Specifically, the ESA provides a detailed report about the quality standards met by satellite images (`https://sentinel.esa.int/documents/247904/685211/Sentinel-2_L1C_Data_Quality_Report`). The data collected from ISTAT, AGREA, and CER (i.e., rural land registers and administrative boundaries) are the result of yearly census activities; the static nature of these data allowed us to manually validate them on random samples. Conversely, the quality of the data coming from other sources is not certified: the ARPAE regional association warns about the possible inaccuracy of weather data (`https://dati.arpae.it/`), and field sensors do not guarantee the correctness of their data. A manual validation of these data is clearly unfeasible; however, the risk of relying on inaccurate data could be mitigated by coupling the owned field sensors with open data and mutually verifying the consistency of each measurement. Although this strategy has not been adopted in the Mo.Re.Farming project (where we do not have multiple sources for the same kind of data), the underlying architecture would in principle be able to

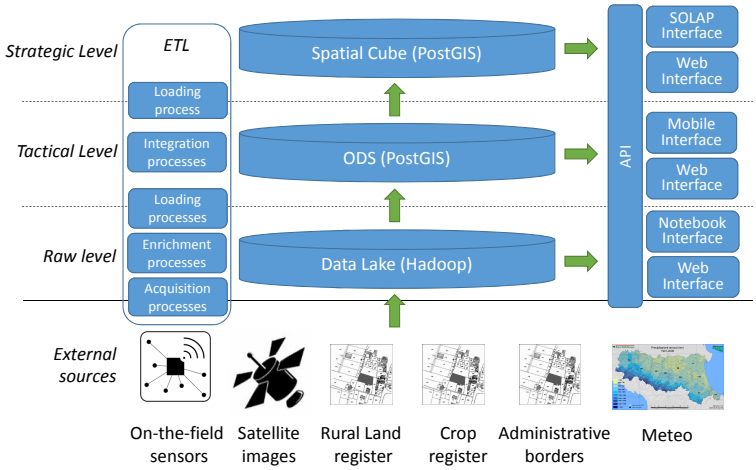


Figure 3: The Mo.Re.Farming architecture

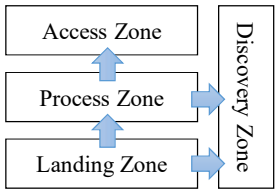


Figure 4: The multi-zone architecture of the Mo.Re.Farming data lake

support this validation approach.

3. The Mo.Re.Farming Architecture

To meet both the tactical and strategic goals, in the Mo.Re.Farming project we adopted a three-tier architecture (often used for BI applications), where each tier relies on its own storage. The architecture is sketched in Figure 3 and is composed as follows.

- The lowest tier hosts a *data lake* [10], i.e., a storage repository that holds a vast amount of raw data in its native format, including structured, semi-structured, and unstructured data. The data lake is used to store all the data coming from the external sources in their raw format and to host the enrichment activities required before their integration (as described in Section 4.1). We rely on a multi-zone architecture [11] for the data lake, so as to logically separate the subsequent processing and enrichment activities. The architecture is shown in Figure 4. It mainly consists of a

Landing Zone (to store data in its raw format), a *Process Zone* (to store intermediate data generated by the processing activities), and an *Access Zone* (to store the data that is ready to be consumed); also, a *Discovery Zone* is allocated to provide a safe environment for data scientists' ad-hoc analyses.

- The middle tier, called *Operational Data Store* (ODS), stores structured data at the finest level of detail for in-depth analysis and monitoring. Whereas in the data lake no fixed schema is defined a priori, the ODS schema is defined at the design time. Indeed, data integration takes place at this level and is mainly based on the spatial features of data; as described in more detail in Section 4, the relationships between heterogeneous data are found thanks to geopositioning even in absence of direct references.

- Finally, the top tier of the architecture consists of a *spatial cube* to enable SOLAP (Spatial OnLine Analytical Processing). The term SOLAP refers to geo-business intelligence technologies allowing online analysis of a massive volume of multidimensional data. Multidimensional data are organized in cubes describing domain events called *facts*, which are characterized by numerical indicators called *measures* (e.g., the NDVI of a field) and *dimensions* to be used for analyses (e.g., space, time, crop). Each dimension is described by a *hierarchy* of concepts that describes the dimension of analysis at different granularity levels (e.g., a field belongs to a municipality, which in turn belongs to a province). Data can be analyzed using SOLAP operators such as *spatial slice* and *spatial drill*, which allow for aggregating measure values along hierarchies with SQL operators.

Table 2 shows the information stored for each type of ingested data and for each architectural level. Spatial data are first-class citizens as they are the basis of data integration (which takes place at the tactical level). However, the storage and manipulation of spatial data requires the adoption of software tools specifically designed to handle these kind of data. The computational needs and the quantity of data to be stored pushes towards the adoption of big data solutions, but the maturity level of the latter is still limited. Whereas vector data (e.g., points and polygons) are relatively easy to manage, the same cannot be said for raster images. In particular, current big data solutions are mainly limited in properly handling continuous field geographic data, that is, spatial phenomena that are perceived as having a value at each point in space and/or time [12, 13, 14]. It is the case of satellite images, which need to be handled at the pixel level. For instance, vegetation indices are computed through *map algebra* manipulations [15], i.e., algebraic operations applied on raster layers (ranging from arithmetical to statistical and trigonometric operations). Whereas map algebra manipulations are a mature feature of *array DBMS*s [16] such as Rasdaman and PostGIS, few big data solutions support this feature and with several limitations.

In light of the above, two opposite technological solutions emerge:

Table 2: Data distribution across architecture levels

	Raw level	*Tactical level*	*Strategic level*
Satellite	ESA raw images (L-1C), Corrected ESA images (L-2A), GeoTIFFs	GeoTIFF pyramids, pixel-level indices (e.g. NDVI)	Crop-level indices
Sensor	Raw images	Enriched images and image-related indices	Crop-level indices
Meteo	–	Vector layer	Crop-level indices
Rural land & Crop Registers, Municipalities	–	Vector layers	Rural land dimension

- *Traditional solution.* Array DBMSs such as Rasdaman and PostGIS are mature open-source technologies that fully support the storage and manipulation of spatial data in both vector and raster forms. The adoption of an array DBMS would be enough to implement all the three levels of the architecture; however, the triviality of this solution is hindered by its lack of support to a big data environment.

- *Big data solution.* The big data landscape is currently characterized by a mosaic of tools that support only some of the features required to handle spatial data. A big data solution is technically feasible, but it requires significant effort to properly manage and integrate the different tools across the different levels of the architecture, depending on the features that they provide. Table 3 shows a summary of the most common technologies that enable the storage and manipulation of spatial data. The tools for data storage (i.e., HDFS, Hive, HBase, Accumulo, Elasticsearch) do not support the storage and manipulation of rasters, which must be stored as simple files. Some software libraries have been released for the manipulation of spatial data, but none of these libraries fully covers the range of spatial data manipulations: GeoSpark and GeoWave mainly focus on vector data; GeoTrellis and GeoMesa are able to process both vector and raster data, but with limited support (GeoTrellis mainly focuses on raster data, GeoMesa mainly focuses on vector data); Rasterframes is still in its initial development. To the best of our knowledge, the only big data solution (comparable to the traditional Rasdaman and PostGIS) is ArcGIS, which is not an open-source software.

At the time of project development, the big data solution was not even an option, as the mentioned tools for raster data manipulations where either not available or too immature to be adopted (even at the time of writing, an open-source big data architecture cannot be implemented without significant effort). For this reason we relied on a *hybrid architecture* that combines the capability to scale to large volumes of data and to enable all the required spatial operations without the hassle of configuring and integrating several not-yet-mature tools. In particular, the data lake is Hadoop-based, while the upper levels rely

Table 3: Summary of the most common softwares for spatial data storage and manipulation

Software	*Storage*	*Manipulation*	*Open-source*	*Big data*
PostGIS	✓	✓	✓	—
Rasdaman	✓	✓	✓	—
GeoSpark	—	vectors	✓	✓
GeoWave	—	✓	✓	✓
GeoTrellis	—	✓	✓	✓
GeoMesa	—	✓	✓	✓
Rasterframes	—	✓	✓	✓
HDFS + GeoJinni	✓	vectors	✓	✓
Hive + GIS tools	vectors	—	✓	✓
HBase	vectors	—	✓	✓
Accumulo	vectors	—	✓	✓
Elasticsearch	vectors	vectors	✓	✓
ArcGIS	✓	✓	—	✓

Table 4: Hardware and software features of the Mo.Re.Farming architecture

	Raw level	*Tactical & Strategic levels*
Hardware	11-node cluster	1 server
	6TB disk (per node)	2TB disk
	32GB RAM (per node)	64GB RAM
	4-core CPU @3.4GHz (per node)	12-core CPU @2.6GHz
Software	Centos 6.9	Windows Server 2012
	Cloudera 5.10.0	PostgreSQL 9.5
	Apache Hadoop 2.6.0	PostGIS 2.3.2
	Sen2cor 2.3.1 [17]	
	GDAL 2.2.0 [18]	

on two PostGIS DBMSs running on a centralized server. This solution enables scalability by relying on a distributed storage and by parallelizing the acquisition and enrichment of raw data, while the PostGIS DBMSs provide a mature environment to implement the tactical and strategic levels. The summary of the hardware and software features of the Mo.Re.Farming architecture is reported in Table 4.

4. Data Model and ETL Processes

In this section we provide an in-depth discussion about the data models adopted for data representation and storage in each architectural level, as well as the ETL (Extract, Transform, and Load) processes that drive the flow of data between such levels. First, we focus on the acquisition of data from the external sources and their enrichment processes (i.e., *External Sources* → *Raw Level*); then we explain how their integration is carried out and modeled at the tactical level (i.e., *Raw Level* → *Tactical Level*); as for the strategic level, we discuss the conceptual model of the spatial cube. Table 2 shows the information stored for each type of ingested data and for each architectural level. We close this section with considerations on the performance of the different ETL processes.

4.1. Acquisition and Enrichment

At the data lake level, data are stored in files on the Apache Hadoop distributed file system (HDFS), which ensures system robustness and enables parallel processing. The processes that run in parallel are those concerning the acquisition and enrichment of satellite images, which are also the most computationally demanding ones. Parallelization is achieved on the 11-node cluster in compliance with the bag-of-task paradigm, where 10 slave workers are coordinated by 1 master process and share a set of independent tasks, each producing an independent output. Specifically, the tasks (implemented in Python) are the following:

- *Satellite images download*: the system periodically verifies through the ESA web services if new satellite images are available for the considered region. If this is the case, the involved granules are downloaded in parallel and stored on HDFS.

- *Atmospheric correction*: since satellites orbit above the Earth atmosphere, the captured images are affected by the reflection of solar light against the atmosphere itself; these image are said to contain *top-of-atmosphere* reflectances and are published by ESA as Level-1C (L-1C) products. Then, atmospheric correction is the task required to cleans images from such reflectances. This is done by relying on the Sen2cor software, made available by ESA (`step.esa.int/main/third-party-plugins-2/sen2cor`), which delivers Level-2A (L-2A) products for L-1C ones. The corrected images (which are said to contain *bottom-of-atmosphere* reflectances) are stored back on HDFS. This particular process is necessary only for the images published before April 2017; since then, the ESA web services also expose the pre-computed L-2A images.

- *GeoTIFF creation*: once the images of every granule for a given date have been downloaded and corrected, they are merged into a single file image and translated from the JPEG 2000 format to the GeoTIFF format, which embeds georeferenced data. This process relies on GDAL (Geospatial Data Abstraction Library, `www.gdal.org`), a translating library for raster and vector geospatial data formats, and is carried out by first extracting the layers of every band in every granule image, then merging the layers from the different granule images for every band. The resulting GeoTIFF is stored on HDFS.

- *Raster pyramid creation*: to enable its visualization through a web interface, the GeoTIFF image is transformed into a pyramid, i.e., a multi-resolution hierarchy of tiled levels. This means that the original image is first split into several tiles, i.e., squares of fixed dimension (256 px). Then, lower resolution levels are progressively built by creating tiles (of the same pixel width) that merge four tiles of the lower level; in other words, each pixel at a higher level consists of the average of four pixels at the lower

level. This process relies on the GDAL library and builds a pyramid consisting of 8 levels overall; the lower level is made of a maximum of 24 400 tiles at maximum resolution (10 m), while the higher level is made of 1 or 2 tiles at the lowest resolution (1.28 km). The result is a set of 8 GeoTIFF tiled images, all stored on HDFS.

The acquisition and enrichment of data from the remaining external data sources is less complex: the volume of data is significantly lower, while the required enrichment (if any) is more trivial.

- On-field sensors are connected to each other through a Robust Wireless Sensor Network [19] and transmit their data via a GPRS connection. Daily acquired trap images and humidity streams are stored on HDFS. Enrichment of images (i.e., the recognition of the captured bugs and their counting) is done by running an ad-hoc image recognition software written in Java, which proceeds in four steps: i) comparison of the new image with the last one to identify the areas that have changed; ii) segmentation of candidate bugs; iii) feature extraction for each bug; iv) classification of the bugs. The classifier is a deep neural network trained on hundreds of bug images, and it provides results with an 82% precision and a 92% recall. The results are stored on the ODS.

- Weather data are made available as open data on a daily basis. A simple ETL process downloads the data and stores them on the ODS, as no enrichment is necessary.

- The rural land and crop registers and the administrative boundaries are all published on a yearly basis. These data are manually downloaded and stored on the ODS; no enrichment is necessary here either.

The above-mentioned sources are downloaded through calls to web services depending on their update frequencies.

4.2. Integration

The ODS was created following a bottom-up approach, i.e., by integrating the different data sources available at the time of the project. Though the users were not directly involved in the design of the ODS, the selection of the data sources was preliminarily done by strictly following the recommendations given by users during a macro-analysis phase, which ensures a good coverage of their requirements. The actual completeness of the ODS with reference to the users specific requirements was verified a-posteriori, when they validated the multidimensional schema of the spatial cube obtained from the ODS.

Figure 5 shows the relational schema of the ODS, where integrated data are stored. Relations are grouped according to the corresponding data sources (shown as grey bounding boxes); primary and foreign keys are denoted with PK and FK, respectively. Solid links represent many-to-one relationships modeled by foreign keys, while dashed links represent many-to-many relationships that

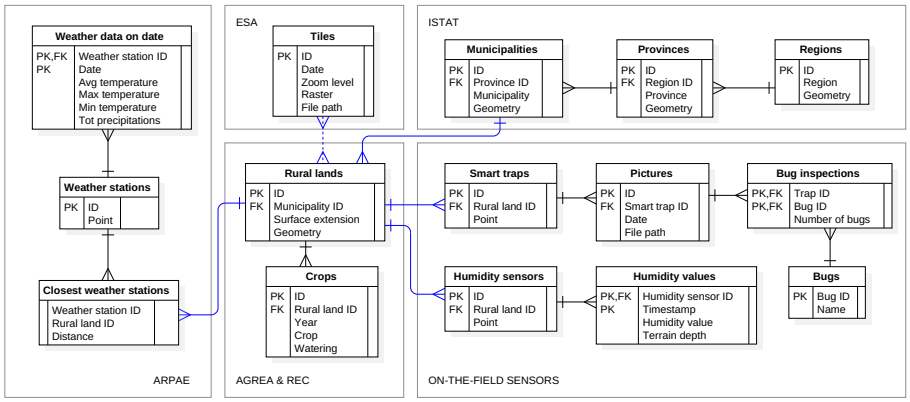


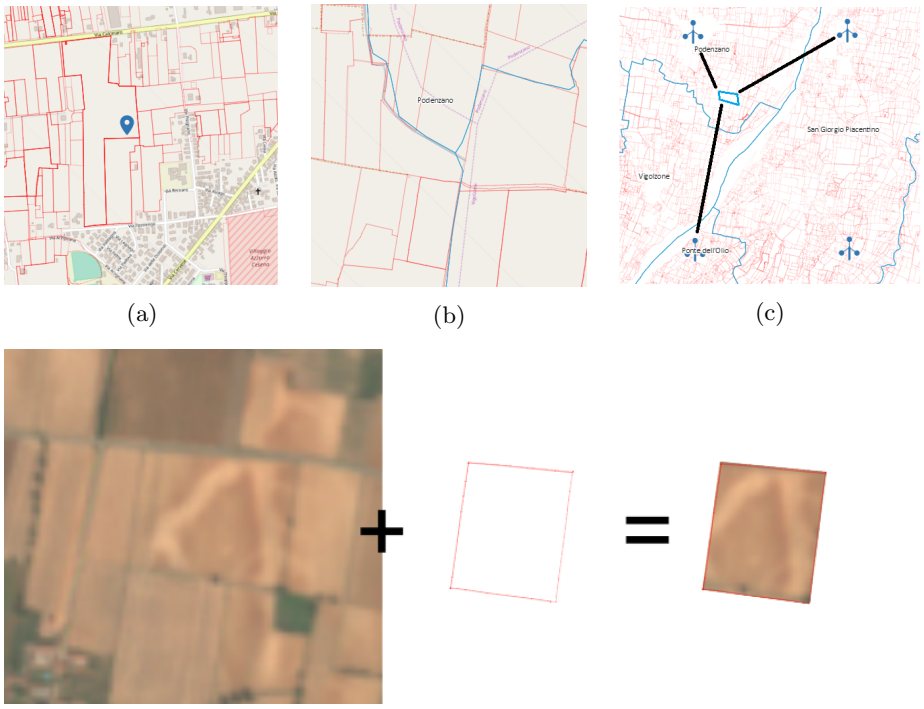
Figure 5: Data schema of the ODS

are not materialized, i.e., they are computed on-the-fly by queries; links are shown in blue if they are determined by spatial join operations. The central role is clearly played by the Rural lands relation, which enables the integration of the different data sources. Remarkably, integration is carried out by exploiting the spatial features of the GIS DBMS. In particular:

- Rural lands & Smart traps/Humidity sensors. Both traps and sensors (identified by a spatial point in the map) are associated to the rural land (represented by a multipolygon) they are contained in, determined using the `contains` spatial operator. Figure 6.a shows with a blue marker the smart trap located in Martorano (Forlì-Cesena) and the rural land it is located in.

- Rural lands & Municipalities. Each rural land is associated to the municipality (represented by a multipolygon) it is contained in. Determining this association is less simple, as the precision of municipal boundaries is lower than the one of rural lands; this results in rural lands being often intersected by different municipalities. For instance, Figure 6.b shows how the boundaries of the municipality of Podenzano (Piacenza) (in blue) do not match exactly the boundaries of rural lands (in red); noticeably, the administrative boundaries of OpenStreetMap (dotted purple lines) are even more distant. To integrate these data, i.e., to determine the municipality of each rural land, the multipolygon of a rural land is spatially intersected with the multipolygons of municipalities using the `intersect` spatial operator; then, the rural land is assigned to the municipality whose area intersects the highest percentage of the rural land area.

- Rural lands & Weather stations. Weather stations are uniformly distributed across the national territory, but their density is (obviously) lower than the one of rural lands; thus, assessing the weather conditions for a given



(a) (b) (c)

(d)

Figure 6: Spatial integration between rural lands and (a) smart traps, (b) weather stations, (c) administrative borders, and (d) tiles; in the latter, the tile at the highest resolution is overlapped by the boundaries of a rural land and then clipped

rural land on a given date requires to locate the closest weather stations by means of the `distance` spatial operator; then weather conditions at rural land granularity can be computed as a weighted triangulation from the closest stations. To efficiently compute it, as shown in Figure 6.c, relation Closest weather stations stores, for each rural land, the three weather stations with the lowest spatial distance —unless a rural land contains a weather station, in which case only that station is stored.

- Rural lands & Tiles. Unlike the previous cases, where only vector data are involved, tiles are represented by raster images. Thus, integrating them with rural lands requires to trim (using the `clip` spatial operator) the portion of the rasters based on the shape of the multipolygons; an example of this operation is shown in Figure 6.d. This overlap is not materialized into a relation for space reasons, but it is computed on-the-fly when answering queries at the tactic level (e.g., to compute vegetation indices on a specific date and rural land) and when loading data to the spatial cube.

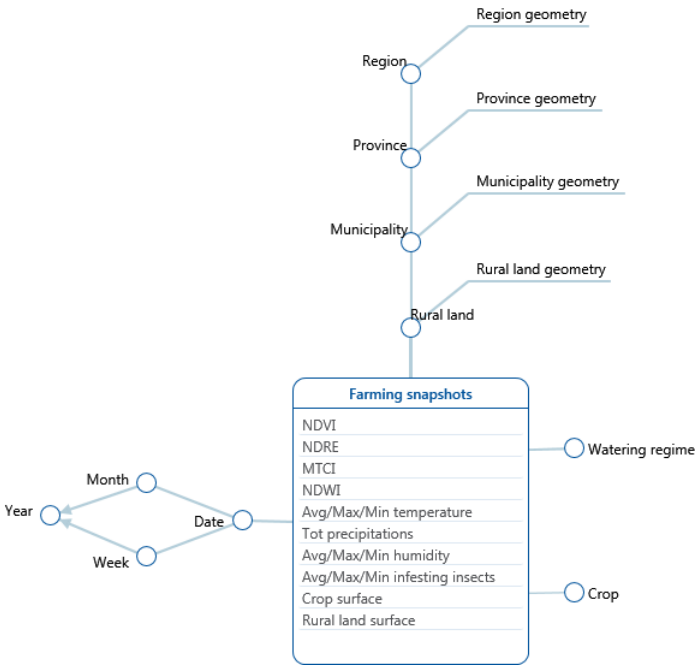


Figure 7: Multidimensional schema of the spatial cube

The ODS is updated daily for data from on-field sensors, satellite data, and weather data, while it is updated yearly for registers.

Finally, it is worth mentioning that different data sources adopt different coordinate reference systems (EPSG) to encode spatial information. Since the amount of data obtained from ESA is the most significant, we chose to keep the ESA reference EPSG (i.e., 32632) and to convert spatial data from the other sources to the same EPSG (ISTAT and AGREA adopt EPSG 23032, while ARPAE adopts EPSG 4326).

4.3. Loading

The functions offered at the strategic level are centered on a spatial cube, whose multidimensional schema is shown in Figure 7 using the DFM notation [20]. The schema was obtained with a data-driven approach, i.e., by choosing a table in the ODS (namely, Crops) as the fact of interest, then following the functional dependencies coded within the ODS schema, and finally validating the result during a meeting with users with expertise in the field of agriculture. Data-driven approaches to multidimensional modeling are widely recognized to achieve an excellent coverage to the users' requirements while keeping the design effort and the probability of errors/misunderstandings to a minimum [20]. Noticeably, the validation of the cube schema by the users also allows to indirectly certify the completeness of the ODS. Although two other tables

in the ODS could have been chosen as facts (Bug inspections and Humidity values), the users declared to be not interested in a fine-grained analysis of bugs and humidity measurements at this stage; conversely, they suggested to compute some statistical metrics, such as the average number of infesting insects, which were added to the multidimensional schema as measures (see below for a description of these metrics).

The Farming snapshot cube features four dimensions: Rural land, Date, Crop, and Watering regime. The Date dimension develops into a temporal hierarchy and contains the date of every satellite image downloaded. The Rural land dimension develops into a spatial hierarchy and provides the geometries of rural lands and administrative divisions. The Crop and Watering regime dimensions contain the list of crops and irrigation regimes (either watered or non-watered), respectively. The events represented in this cube consist of snapshots, one for each satellite image downloaded, providing statistics for the crop of a given rural land in a given date. The available statistics, represented as measures in the cube, are:

- NDVI [21], NDRE [22], MTCI [23], NDWI [24]: a set of vegetation indices, computed by analyzing —for each given date— the portion of satellite image that falls within the spatial boundaries of the rural land. Vegetation indices are used to measure the status of vegetation on the ground; the rationale is to compare the reflectance values on different spectral bands, as the chlorophyll in plants' leaves reflects light in different ways under different conditions. As mentioned in Section 2, satellite images contain 13 spectral bands that range from the visible range (i.e., red, green, and blue bands) to the shortwave infrared (SWIR). Each index is calculated by applying a map algebra operation on selected bands provided in the image. In particular:
  - NDVI (Normalized Difference Vegetation Index) ranges from -1 to 1 and is calculated as $\frac{NIR-Red}{NIR+Red}$, where $NIR \in [0,1]$ is the ratio of the reflected over the incoming radiation in near-infrared light, and $Red \in [0,1]$ is the ratio of the reflected over the incoming radiation in the visual red light. It is one of the most common vegetation indices and its goal is to verify whether the observed region contains living green vegetation.
  - NDRE (Normalized Difference Red Edge index) ranges from -1 to 1 and is calculated as $\frac{NIR-RE}{NIR+RE}$, where RE is a frequency band that sits on the transition region between Red and NIR. It is a slight variation of NDVI and gives better insight on some kind of crops (e.g., cereals) and growth stages (i.e., later stages of growth), where plants are more rich of leaves and it is more difficult to observe the state of their lowest levels.
  - MTCI (Meris Terrestrial Chlorophyll Index) ranges from 0 to $+\infty$ and is calculated as $\frac{NIR-RE}{RE-RED}$. It is used to observe the chlorophyll content in vegetation.

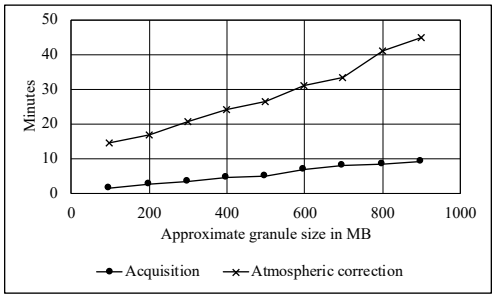


Figure 8: Average execution times of ETL processes involving single granules by granules' approximate size

  - NDWI (Normalized Difference Water Index) ranges from -1 to 1 and is calculated as $\frac{NIR-SWIR}{NIR+SWIR}$, where $SWIR \in [0,1]$ is the shortwave infrared band. It is used to measure the hydration of plants, as $SWIR$ reflects changes in both the vegetation water content and the mesophyll structure in vegetation canopies.

- Avg/Max/Min temperature, Tot precipitations: weather information computed by weighing the temperatures measured at the three stations closest to the rural land. Since satellite images are not available for every date, the snapshots for some dates are missing from the cube; thus, measure values are actually representative of the whole time range between the previously available snapshot and the current one. For instance, if no snapshots were downloaded from January 5 to 9, 2016, measure Max temperature for January 10, 2016 is computed as the maximum of the maximum temperatures registered since January 5, 2016.

- Avg/Max/Min humidity, Avg/Max/Min infesting insects: depending on the density of the sensor network, these measures could be given a value either only for the specific field the sensor is installed in (low density), or for all the fields following a weighing approach similar to the one adopted for weather information. Due to the low number of sensors currently available, in Mo.Re.Farming we adopt the first solution.

- Crop surface, Rural land surface: the extent of the land for the given crop in the given rural land, and the overall extent of the rural land. Since these measures are time-independent, they cannot be aggregated along the time dimension.

5. User Interface for Data Access

In Mo.Re.Farming we devised two approaches for users to access the data collected and produced. The first one is a web interface that guides the user

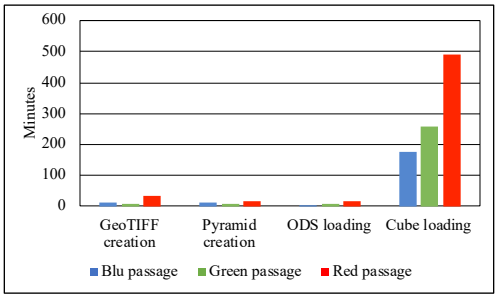


Figure 9: Average execution times of ETL processes involving whole images (colors refer to Figure 1)

experience through a set of dashboards, depending on the chosen architectural level. The second one is a set of public web services that expose the available data by means of standard RESTful APIs. Such connectivity is particularly relevant to make the Mo.Re.Farming repository a node of a wider network of open data and open services for precision agriculture.

5.1. Web Interface

The web interface provides different experiences to users depending on the architectural level.

The data lake is mainly used for back-end computations, nonetheless it can be queried through notebook technology, which is a perfect tool for data scientists since it allows to couple advanced visualizations, data retrieval, programming, and documenting of research procedures. In Mo.Re.Farming we adopted Apache Zeppelin (`zeppelin.apache.org/`).

The main access to the ODS level is through an ad-hoc dashboard, implemented in PHP and Javascript; specifically, a map component is delivered by relying on the open-source OpenLayers3 Javascript library. The interface (available at `semantic.csr.unibo.it/morefarming` and shown in Figure 10) is intended for both agricultural technicians and farmers, and allows several different information to be visualized. The option panel on the left allows to select: (i) the kind of image to be displayed (either the raw satellite image or its translation to one of the vegetation indices), possibly coupled with OpenStreetMap in the background; (ii) the reference date for the image; and (iii) one or more vector layers, including the rural land borders, administrative borders, and markers that highlight the location of smart traps, humidity sensors and weather stations. The user can obtain statistics about any point or field by simply clicking on the map: the right panel will show the ID of the selected rural land (if any), geographical information (exact coordinates, province, municipality), the values for the vegetation indices on the selected date (for both the single point and the whole rural land), and the overall trend of vegetation

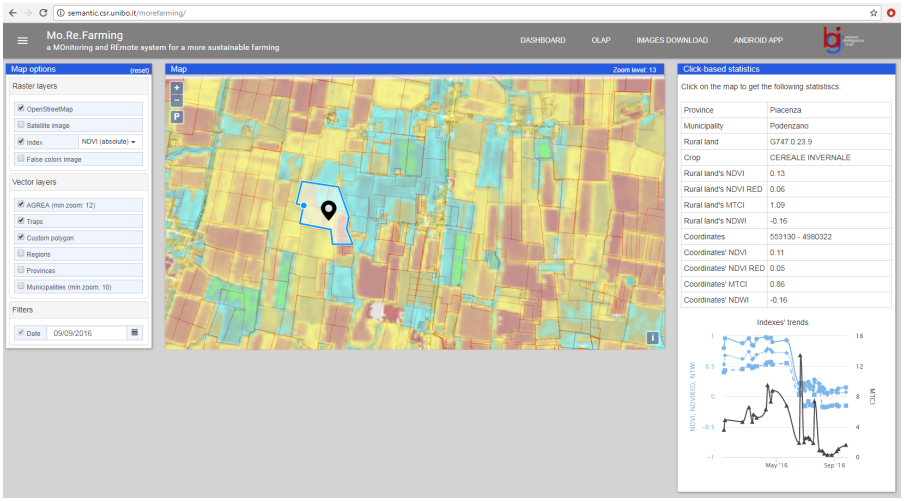


Figure 10: Tactical-level web application showing statistics on a field in the municipality of Podenzano on September 9, 2016; the colors of the map are representative of the NDVI

indices for a single point[1]. The "P" icon on the map allows to draw a custom polygon on the map and to obtain statistics on this polygon instead of a single point or rural land. Finally, the blue markers that locate smart traps can be clicked to visualize the photos in inverse chronological order, indicating the number and kinds of bugs recognized. All the information available on the web interface are also available on a mobile app for Android, whose goal is to exploit the smartphones geolocalization capabilities to allow in-field visualization of the relevant information.

Finally, a SOLAP solution is implemented on the top layer through the Saiku software (`meteorite.bi/products/saiku`), which enables the execution of SOLAP queries on the spatial cube. The spatial features can be exploited by first drawing a polygon on the map through the tactical-level interface; then, as shown in Figure 11, the polygon can be used as a spatial filter in the SOLAP query to select only the rural lands that intersect that polygon.

5.2. Web Services

Each level of the Mo.Re.Farming architecture is exposed by a set of public web services, i.e., APIs that are based on the well-known Representational State Transfer (REST) software architecture.

The first layer of APIs exposes most of the contents of the data lake: satellite images in every available format (i.e., the raw L-1C image with top-of-atmosphere reflectances, the corrected L-2A image with bottom-of-atmosphere

[1] This information is actually more strategic than tactical; providing this view at the tactical level is possible only because the scope is limited to a tiny portion of the map.

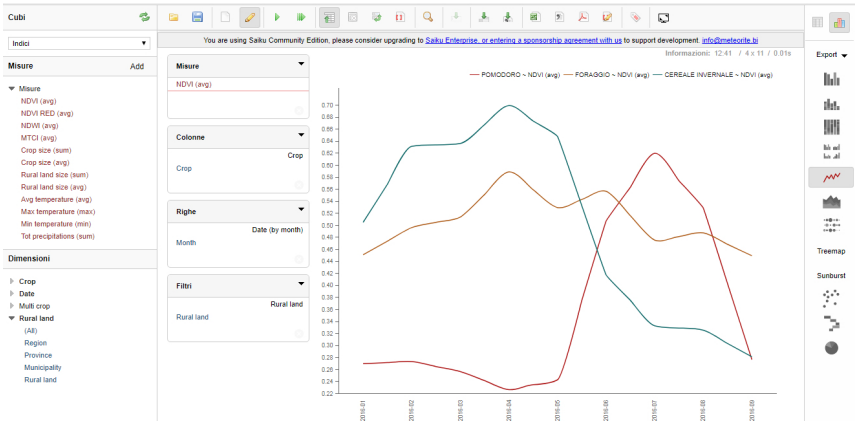


Figure 11: Strategic-level web application, showing average monthly trends for NDVI on the crops available in a set of fields in the municipality of Podenzano. The fields have been selected by spatially intersecting the polygon manually drawn in the tactical-level interface with the rural land geometries. The red line represents tomato crops, the orange one forage, the green one winter cereal.

reflectances, and the final GeoTIFF image that covers the whole region after being tiled and turned into a pyramid of multi-resolution levels), raw data and results of elaborations from on-the-field sensors, the adopted georeferenced administrative borders obtained from ISTAT, and the meteo data obtained from ARPAE. Unfortunately, due to license regulation, rural land and crop data are not available for open access at the time of writing. The entry point for these APIs is `semantic.csr.unibo.it/morefarming/api/data-lake`; the web service is built in a browsable way, i.e., data can be downloaded by simply following the links from the entry point.

The second layer of APIs exposes the contents of the ODS. In particular, it allows to obtain punctual information about vegetation indices for a given point on the map on a given date (i.e., the same information shown in the main dashboard by clicking on the map) in a RESTful way. Users can access these data by connecting to `semantic.csr.unibo.it/morefarming/api/ods`.

The third layer of APIs exposes the contents of the spatial cube. Since the Saiku software (which has been deployed in the web interface) already comes with its own web services, we reuse them in order to enable the execution of OLAP (On Line Analytical Processing) queries on the cube. In particular, these APIs allow users to understand the multidimensional structure of the cube and to issue custom queries formulated in the MDX[2] language.

Documentation and examples on how to use the APIs are provided in the web interface at `semantic.csr.unibo.it/morefarming/opendata.php`.

[2]Multidimensional expressions; `mondrian.pentaho.com/documentation/mdx.php`

6. Performance

As discussed in Section 4, a series of ETL processes controls the flow of data, from their acquisition into the data lake to their loading into the Farming snapshot cube. The average execution times of these processes are shown in Figures 8 and 9; the first one is focused on the ETL processes involving single granules, the other one those involving the whole images. The size of images varies depending on the kind of satellite passage (take Figure 1 as a reference): the red passage is the one covering the greatest area (including all 7 granules, although some only partially), while the blue and green passages cover similar areas. We can comment the figures as follows:

- The performances of the ETL processes scale linearly with the size of the data. This is observable in both Figures 8 and 9. In the former, the size of granules depends on the portion covered by the satellite and on the atmospheric conditions (i.e., cloudy or clear); in the latter, the size of the whole images mostly depends on the area covered by the satellite passage.
- Overall, the most expensive process is the loading of the spatial cube, even though it is incremental (i.e., the cube is updated with the changes in the ODS since latest iteration). This is due to the spatial integration of rural land vectors and tile rasters required to compute the vegetation indices. In particular, the single operation of clipping the raster image on the rural land boundaries takes about 80% of the execution time of this process; another 10% is due to the calculation of weather data for every rural land in a specific time range (usually 2-3 days); the remaining time is taken by other join and insertion operations.
- The amount of time spent doing atmospheric correction on a whole image is also significant. However, since this task can be carried out independently on the single granules (which can be up to 7 per image), its overall impact can be considerably reduced by parallelizing its execution.
- Execution times for the acquisition of granules are subject to variation depending on the quality of the internet connection and the speed of ESA servers.

The hardware features listed in Table 4 determine adequate performance for the area of interest of the project. Obviously, scaling to a nation-wide (or greater) level makes it necessary to either scale-up or scale-out. Conversely, moving to a coarser data resolution would make some of the services proposed in Table 1 unfeasible (e.g., the monitoring of the vegetation level).

7. Related Literature

The literature on big data approaches to precision farming and precision agriculture is growing quickly. Most efforts are focused on applying machine

learning techniques to ad hoc agricultural datasets [3]. Another relevant part of the literature is devoted to sensors [1] and remote sensing [25]. All these works only marginally focus on the data management architecture and on the features of the platform in charge of providing an integrated and univocal view of farming data. The only work going in this direction is [26], which proposes an open and integrated *cyber-physical infrastructure*, i.e., a coordinated environment that includes several hardware components, software, and interactions. This infrastructure builds on open standards to define an integrated middleware between heterogeneous monitoring sensors and different precision agriculture applications. The focus of this work is mainly on the data collection problem, which is addressed through a service-oriented architecture. However, the applications proposed are exclusively at the tactical level; validation on a real case study involving a large quantity of data is limited, and no OLAP-like functionalities are discussed or implemented. In [27] the authors stress the advantages of making available to farmers heterogeneous but integrated data in near-real time. To this end they propose the PRIDE (Progressive Rural Integrated Digital Enterprise) business model, but no technical details on the architecture and data model are provided.

Although the scientific community has paid little attention to the issues related to architectures, integration, and data handling, some free and commercial data services are already available. While most solutions are based on web applications to deliver data services, they strongly differ in the way data are stored, processed, and made available, as well as in the type of data provided and in the professional figures and services they are oriented to. Such services are briefly surveyed below to emphasize differences and similarities with reference to Mo.Re.Farming (see Table 5). We did not restrict ourselves to services strictly related to precision agriculture, but we also considered some general purpose services for distributing remotely-sensed satellite data.

- *Global Land Cover* [28] relies on an architecture similar to Mo.Re.Farming, though it is implemented on a relational engine coupled with an image-publishing DBMS. The system integrates LandSat satellite images with a large set of ancillary data (e.g., ecological zones, digital elevation models) and in-situ data (e.g., local photos). The goal is to deliver world-wide information of land-use data production. Besides traditional visualization and navigation of the information, the system includes a *collaborative management* module that supports project participants, possibly around the world, to collaboratively work on the data. For example, it can design the workflow of data verification, assign them different roles and tasks, and further maintain the execution of the workflow. It does not implement any OLAP-like functionality.

- *CropScape* [29] is a DSS for agriculture in the USA. Similarly to Mo.Re.Farming it provides information about historical data and crops, even though the granularity is coarser (30-56 m). It does not implement any OLAP-like functionality and relies on a traditional three-tier architecture (Application/Service/Data) where the data layer is directly implemented on a file

system.

- *Mundialis* [30, 31] is a commercial system providing a large set of land-related services based on satellite and radar images. Services span from precision agriculture (e.g., plant health and growth, spatial pattern of productivity, fields history using archive satellite data) to insurance (e.g., monitoring crop health and productivity for index insurance, post-event damage assessment) and developmental work (e.g., project region targeting, assessment of land suitability). Mundialis leverages on open source software coupled with specialized user interfaces. The main open source packages involved in the back-end are GDAL, PROJ4, GRASS GIS, and GeoServer. The front-end is based on the JavaScript libraries GeoExt and OpenLayers. To ensure an adequate computation power, the system runs on an HPC (High Performance Computing) infrastructure. No OLAP-like functionality is implemented.

- *Moses* [32] is mainly devoted to developing machine learning and statistical techniques for the optimization and monitoring of the irrigation resources. It integrates satellite images with macro-crop types and weather data. In particular, the main functionalities in the scope of the projects are: early crop mapping from space, seasonal prediction of irrigation water demand, in season evapotranspiration and crop water status monitoring from space, and mid-term irrigation numerical forecasting. The Moses system is the outcome of an ongoing EU-H2020 project and is mainly devoted to developing machine learning techniques. Although the project goals are very ambitious in terms of functionalities to provide, the set of available features is still limited and the current implementation is at a beta-testing stage. No OLAP-like functionality is implemented.

- *Earth Observation Data Services* [33] is based on the Rasdaman data manager [34] and provides worldwide coverage for different families of satellites. It enables data navigation and downloading and it pre-calculates some features derived from the the images, such as the NDVI and a cloud mask.

Mo.Re.Farming presents several distinguishing features with reference to the other projects in Table 5. In particular it integrates a high number of different layers while delivering detailed information in terms of both spatial and temporal dimensions (i.e., satellite images are captured daily at 10 m resolution). Furthermore, it provides support for both tactical and strategic precision agriculture by including a SOLAP interface. The comprehensive survey proposed in [36] describes the relevant projects in SOLAP for Agri-Environmental Analysis and classifies them according to different coordinates such as design methodology, spatio-temporal model, analysis type, and architecture implementation. The Mo.Re.Farming SOLAP module would be classified as *simple* on all the coordinates except for architecture, since it allows *drill-through* analyses that cross the border between the tactical and strategic levels to get an insight about a specific component of an OLAP report.

Table 5: Main features of agriculture and terrain-health related projects

	Mo.Re. Farming	*Global Land Cover* [28]	*CropScape* [29]	*Mundialis* [30]	*Moses* [35]	*Earth Obs. Data Service* [33]
Technologies	hybrid: relational and big data	relational	file system & web-services	file system[3] & web-services	ESRI GIS & Postgres	array DBMS
Main layers beyond sat. images	crop boundaries, crops, weather, vegetation indices	ecological zones, digital elevation, in-situ data	crops	vegetation indices	crop boundaries, crops, weather	vegetation indices
Number of crops	50	-	> 100	-	16	-
Satellite	Sentinel-2	LandSat	Landsat5/7/8,Sentinel-2		Landsat8, Sentinel-2	Sentine2, Landsat8 & Modis
Sat. images definition	10 m	30 m	30/56 m	10 m	10 m	10 m
Historical depth	2 years	2 years (2000, 2010)	20 years	30 years	1 year	17 years
Sat. revisit time	2-3 per week	once per year	once per year	once per day	once per day	
Goal	tactical and strategic precision agriculture	general purpose	tactical precision agriculture	general purpose	water resource optimization	general purpose
OLAP	yes	no	no	no	no	no

8. Conclusion and Discussion

In this paper we described the architecture supporting the Mo.Re.Farming project. In particular, we showed how the integration of data related to weather, crops, and fields, enabled by the adoption of a hybrid architecture, boosts precision agriculture both at the strategic and tactical levels.

The Mo.Re.Farming experience elicited several interesting aspects related to the development of precision agriculture systems.

- Although the computational needs and the quantity of data to be stored pushes towards the adoption of big data solutions (e.g., Apache Hadoop), the level of development of spatio-temporal features on those platforms is still limited. In particular, no support to continuous field data is given in general-purpose big data solutions. At the time of writing, some open-source libraries for spatial big data operations (i.e., GeoSpark and GeoTrellis) have matured, and their combined functionalities cover the set of operations required for the management of satellite images. This makes it possible to migrate the project to a full big data solution, although some effort would be necessary to properly integrate the available libraries.

- The precision agriculture landscape is currently characterized by several solutions, models, and services, each working on a subset of the available data and providing complementary but non-integrated information and services. Creating integrated hubs of information is mandatory to overcome these limitations and to deliver more effective information and services [37].

- As emphasized in Section 4, integration tasks are computationally demanding and can hardly be carried out on-the-fly. Specifically, while on-the-fly integration could still be feasible at the tactical level thanks to the low quantity of information involved, it can hardly be done at the strategic level where a huge amount of data and processing is needed.

- The Mo.Re.Farming architecture represents an example of a data hub, conceived as a starting point for delivering integrated information and services. The advantages of carrying out integration at the physical level, i.e., by materializing the integrated data in the ODS, are (i) the possibility of late data reworking, (ii) a 360-degrees exploitation of data with no limitations due to different data owners, and (iii) the possibility of efficiently running complex analytics on heterogeneous data [37].

- Physical integration and materialization also comes with a few drawbacks, namely higher costs for storing and handling data and the risks related to data replication.

- Precision agriculture architectures should be as open as possible to enable complementary data exchange and fruition. From this point of view we perceived the need for a standard and machine-readable terminology. The issue of standard terminology has been addressed by some research papers [38, 39, 40], but no complete and accepted solution is available yet. As to machine-readability, [41] proposes a framework for SOLAP analyses on the semantic web and presents a case study in the environmental and agriculture domain.

- The number of potential data sources and data collection approaches will further increase in the coming future. Interesting insights come from applying the crowdsourcing approach to agriculture (*farmsourcing* [2]). This model opens to the possibility of collecting new types of data that complements big data. For example, satellite remote sensing data can be better exploited in areas with large, homogeneous, and flat agricultural parcels, and may work not properly in presence of small-scale parcels and mixed crop cover. A step forward is the adoption of a technological farmsourcing approach, that is, an IoT of farm machinery and sensors that send data to a hub. We are also working to build an intelligent and flexible data lake system, where new data in different formats can be ingested and processed based on the automatic recognition of the content.

- The farmers involved in the project have reported a decreasing interest towards historical data due to the ongoing climate change, which decreases the accuracy of forecasting based on such data. This is further amplified by the absence of detailed data (due to the high cost of sensor technologies, that hinders their dissemination in a poor and often underdeveloped sector). Conversely, farmers would be more interested into a *telemetry-like* system, that continuously monitors large areas and promptly identifies patterns of unexpected situations, so that a fast and timely action can be adopted.

- Stakeholders in the agriculture field are characterized by completely different skills, culture, and mindset as to digitization and data analysis. While technicians and managers (e.g., councilors for agriculture and agriculture technicians) are already willing to move towards a data-driven agriculture, most farmers are not; a way to overcome such reluctance is to deliver easy-to-use analytics and identify a small set of "killer applications", i.e., applications that deliver an immediate and relevant advantage to the user. In this direction, the Mo.Re.Farming mobile interface has been strongly appreciated by in-field users since it delivers geolocalized information. As emphasized in [2], the main recipients are new farmers or early-adopters of cutting-edge technologies who are more sensible to the added value that these solutions can offer and can appreciate a win-win exchange of data, information, and services.

The Mo.Re.Farming project is still ongoing. Our future work goes in two directions: on the one hand, we plan to integrate further data sources (e.g., the soil map); on the other, we plan to develop analytics that can exploit the information power of integrated data. In particular, our users are interested in water resource optimization, which requires both historical and current data.

References

[1] T. Ojha, S. Misra, N. S. Raghuwanshi, Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges, Computers and Electronics in Agriculture 118 (2015) 66–84.

[2] J. Minet, Y. Curnel, A. Gobin, J.-P. Goffart, F. Mélard, B. Tychon, J. Wellens, P. Defourny, Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach, Computers and Electronics in Agriculture 142 (2017) 126–138.

[3] A. Kamilaris, A. Kartakoullis, F. X. Prenafeta-Boldú, A review on the practice of big data analysis in agriculture, Computers and Electronics in Agriculture 143 (2017) 23–37.

[4] A. Jhingran, N. Mattos, H. Pirahesh, Information integration: A research agenda, IBM Systems Journal 41 (4) (2002) 555–562.

[5] J. Trujillo, A. Maté, Business intelligence 2.0: A general overview, in: Business Intelligence, Springer, 2012, pp. 98–116.

[6] M. Golfarelli, S. Rizzi, I. Cella, Beyond data warehousing: what's next in business intelligence?, in: Proc. DOLAP, Washington DC, USA, 2004, pp. 1–6.

[7] E. Gallinucci, M. Golfarelli, S. Rizzi, A hybrid architecture for tactical and strategic precision agriculture, in: Proc. DaWaK, Linz, Austria, 2019, pp. 13–23.

[8] Copernicus project (October 2017).
URL `https://scihub.copernicus.eu/`

[9] G. Luciani, A. Berardinelli, M. Crescentini, A. Romani, M. Tartagni, L. Ragni, Non-invasive soil moisture sensing based on open-ended waveguide and multivariate analysis, Sensors and Actuators A: Physical 265 (2017) 236–245.

[10] B. Stein, A. Morrison, The enterprise data lake: Better integration and deeper analytics, PwC Technology Forecast: Rethinking integration 1 (2014) 1–9.

[11] F. Ravat, Y. Zhao, Data lakes: Trends and perspectives, in: Proc. DEXA, Linz, Austria, 2019, pp. 304–313.

[12] A. A. Vaisman, E. Zimányi, A multidimensional model representing continuous fields in spatial data warehouses, in: Proc. ACM-GIS, Seattle, USA, 2009, pp. 168–177.

[13] L. I. Gómez, A. A. Vaisman, E. Zimányi, Physical design and implementation of spatial data warehouses supporting continuous fields, in: Proc. DaWak, 2010, pp. 25–39.

[14] R. Laurini, L. Paolino, M. Sebillo, G. Tortora, G. Vitiello, A spatial SQL extension for continuous field querying, in: Proc. COMPSAC, Vol. 2, 2004, pp. 78–81.

[15] C. Tomlin, J. Berry, Mathematical structure for cartographic modeling in environmental analysis, in: Proc. American Congress on Surveying and Mapping Annual Meeting, 1979, pp. 269–283.

[16] P. Baumann, A database array algebra for spatio-temporal data and beyond, in: Proc. NGITS, 1999, pp. 76–93.

[17] U. Muller-Wilm, J. Louis, R. Richter, F. Gascon, M. Niezette, Sentinel-2 level 2A prototype processor: Architecture, algorithms and first results, in: Proc. ESA Living Planet Symposium, Edinburgh, UK, 2013, pp. 9–13.

[18] F. Warmerdam, The geospatial data abstraction library, Open source approaches in spatial data handling (2008) 87–104.

[19] A. Giorgetti, M. Lucchi, E. Tavelli, M. Barla, G. Gigli, N. Casagli, M. Chiani, D. Dardari, A robust wireless sensor network for landslide risk analysis: system design, deployment, and field testing, IEEE Sensors Journal 16 (16) (2016) 6374–6386.

[20] M. Golfarelli, S. Rizzi, Data warehouse design: Modern principles and methodologies, McGraw-Hill, Inc., 2009.

[21] D. W. Deering, Rangeland reflectance characteristics measured by aircraft and spacecraft sensors, Ph.D. thesis, Texas A&M Univ., College Station (1978).

[22] A. Gitelson, M. N. Merzlyak, Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. spectral features and relation to chlorophyll estimation, Journal of Plant Physiology 143 (3) (1994) 286–292.

[23] J. Dash, P. J. Curran, The MERIS terrestrial chlorophyll index, IJRS 25 (23) (2004) 5403–5413.

[24] B.-C. Gao, NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space, Remote sensing of environment 58 (3) (1996) 257–266.

[25] A. Vibhute, S. Bodhe, Applications of image processing in agriculture: a survey, International Journal of Computer Applications 52 (2).

[26] N. Chen, X. Zhang, C. Wang, Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring, Computers and Electronics in Agriculture 111 (2015) 78–91.

[27] M. Sawant, R. Urkude, S. Jawale, Organized data and information for efficacious agriculture using PRIDE model, International Food and Agribusiness Management Review 19 (A).

[28] G. Han, J. Chen, C. He, S. Li, H. Wu, A. Liao, S. Peng, A web-based system for supporting global land cover data production, ISPRS 103 (2015) 66–80.

[29] W. Han, Z. Yang, L. Di, R. Mueller, CropScape: A web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support, Computers and Electronics in Agriculture 84 (2012) 111–123.

[30] Mundialis (October 2017).
URL `https://www.mundialis.de/`

[31] M. Neteler, T. Adams, H. Paulsen, Combining GIS and remote sensing data using open source software to assess natural factors that influence poverty, in: Proc. World Bank Conference on Land and Poverty, Washington DC, USA, 2017, pp. 1–7.

[32] A. Di Felice, M. Mazzolena, V. Marletto, A. Spisni, F. Tomei, S. Alfieri, M. Menenti, G. Villani, G. Scarpino, A. Battilani, Climate services for irrigated agriculture: Structure and results from the MOSES DSS, in: Proc. Symposium Società Italiana per le Scienze del Clima, Bologna, Italy, 2017, pp. 1–4.

[33] Earth observation data services (October 2017).
URL `https://eodataservice.org/`

[34] P. Baumann, A. Dehmel, P. Furtado, R. Ritsch, N. Widmann, The multidimensional database system RasDaMan, in: Proc. SIGMOD, 1998, pp. 575–577.

[35] MOSES project (October 2017).
URL `http://moses-project.eu/moses_website/`

[36] S. Bimonte, Current approaches, challenges, and perspectives on spatial OLAP for agri-environmental analysis, IJAEIS 7 (4) (2016) 32–49.

[37] S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, Big data in smart farming– a review, Agricultural Systems 153 (2017) 69–80.

[38] G. Song, M. Wang, X. Ying, R. Yang, B. Zhang, Study on precision agriculture knowledge presentation with ontology, AASRI Procedia 3 (2012) 732–738.

[39] S. D. K. Tomic, W. Wöber, S. Hörmann, W. Auer, D. Drenjanac, Enabling semantic web for precision agriculture: a showcase of agriOpenLink project, in: Proc. SEMANTiCS, Vienna, Austria, 2015, pp. 26–29.

[40] K. Charvat, M. A. Esbri, W. Mayer, A. Campos, R. Palma, Z. Krivanek, Foodie – open data for agriculture, in: Proc. IST-Africa Conference, 2014, pp. 1–9.

[41] N. Gür, K. Hose, T. B. Pedersen, E. Zimányi, Enabling spatial OLAP over environmental and farming data with QB4SOLAP, in: Proc. JIST, Singapore, Singapore, 2016, pp. 287–304.