Multi-model Data Warehouse for Multidimensional Analysis of Flavescence Dorée: Application to the Bordeaux Region

Sandro Bimonte[1(✉)], Fagniné Alassane Coulibaly[1], Frederic Fabre[2], Sylvie Malembic-Maher[3], and Stefano Rizzi[4]

[1] TSCF, INRAE, University Clermont Auvergne, Clermont-Ferrand, France
{sandro.bimonte,fagnine.coulibaly}@inrae.fr
[2] UMR SAVE, INRAE, Bordeaux Sciences Agro, Villenave d'Ornon, France
frederic.fabre@inrae.fr
[3] UMR BFP, INRAE, Université Bordeaux, Villenave d'Ornon, France
sylvie.malembic-maher@inrae.fr
[4] DISI, University of Bologna, Bologna, Italy
stefano.rizzi@unibo.it

Abstract. Flavescence Dorée is a highly contagious grapevine disease that poses a significant economic threat to European viticulture. Its current management relies on symptom monitoring, uprooting infected plants, insecticide treatments, and reservoir management. However, effective disease control is hindered by the disease's asymptomatic period and symptom similarities with Bois Noir, necessitating complex data-driven monitoring. Currently, the reliance on traditional spreadsheet software limits the scalability and spatio-temporal analysis of the massive datasets collected. To address these issues, this paper proposes a Business Intelligence architecture based on Spatial OLAP (OnLine Analytical Processing). By integrating Data Warehousing with Geographic Information Systems, the proposed system enables intuitive, multi-dimensional analysis of diverse data types——including graphs for dispersal modeling and JSON for scientific flexibility. Relying on a Multi-Model Data Warehouse with schema-on-read extensibility, this architecture provides an agile, "drag-and-drop" platform for extension services, farmers and researchers in epidemiology. This approach could facilitate rapid detection, predictive modeling, and optimized resource allocation, ultimately shifting pest control towards a more ecological and economically sustainable model.

Keywords: Flavescence dorée · Multi-model data · Spatial OLAP

1 Introduction

Flavescence Dorée (FD) is a grapevine disease caused by a wall-less bacterium (phytoplasma) that circulates in the plant's phloem. It is transmitted by the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 509–525, 2027.
https://doi.org/10.1007/978-3-032-30527-5_33

leafhopper *Scaphoideus titanus* or through grafting [13]. First identified in Armagnac in the 1950s, FD has since spread across France and much of Southern Europe. FD phytoplasma is classified as a quarantine organism in European Union. Symptoms include leaf rolling and discoloration (yellow or red), drying of clusters, and a failure of new shoots to lignify. However, visual diagnosis is difficult because the non-epidemic phytoplasma associated to Bois Noir (BN) disease induces identical symptoms; only molecular testing can distinguish them. Furthermore, symptoms in plants only appear at least one year after infection. This means plants infected in year n act as asymptomatic sources of spread throughout the spring of year $n+1$ before they are even detected. The strategy to control FD relies on four pillars: (i) monitoring symptomatic plants and insect vector populations, (ii) uprooting infected plants to eliminate the inoculum source; (iii) using insecticide treatments to reduce vector populations; (iv) managing uncultivated grapevines to remove potential disease reservoirs. The *Groupement de Défense contre les Organismes Nuisibles* (GDON) of Bordeaux monitors 74,000 hectares in the French department Gironde to control FD. They conduct leafhopper counts and symptom inspections from May to October of each year. If symptoms are detected in new areas, laboratory samples confirm if the cause is FD or BN. The monitoring protocol of GDON of Bordeaux is described in [4]. This targeted monitoring allows for a more ecological and economic approach by adjusting treatment levels based on real-time insect vector and disease pressure. Although GDON collect huge volume of data, data management and analysis is conducted by means of classical spreadsheet software (like Excel), which present several limitations in terms of scalability, management of complex data, and visual spatio-temporal analysis capabilities.

Data Warehouses (DWs) and OLAP systems are first-class citizens of Business Intelligence systems and have been used for several years in the agricultural context [6]. A DW acts as a centralized repository that consolidates large volumes of diverse, historical data from multiple sources. OLAP is the visual metaphor used to interact with this data, allowing non-technical users to perform multidimensional analysis through intuitive drag-and-drop operations. Together, they enable the creation of interactive dashboards by structuring data into cubes that can be easily aggregated, sliced, and diced. This combination transforms raw data into actionable business intelligence, facilitating rapid decision-making without requiring complex programming or data science expertise. The integration of OLAP and Geographic Information Systems (GIS) leads to the concept of Spatial OLAP (SOLAP), which enables spatial and OLAP queries as well as the visualization of query results by means of interactive maps [6].

Data variety characterizes agricultural pest monitoring applications since they need to analyze diverse formats of data like spatio-temporal files, satellite rasters, and connectivity graphs [18]. Data is obtained from field surveys, crowdsourcing, IoT, and remote sensing to capture complex ecological dynamics. To maintain data expressivity and avoid complex Extraction-Transformation-Loading processes, it is crucial to store information in its native model rather than forcing a relational transformation. In particular, graphs have been proved

to be effective for modeling epidemiological dispersal and social knowledge-sharing networks [14], while JSON offers the schemaless flexibility needed for hierarchical and evolving scientific data. Ultimately, centralizing these multi-model data in their original formats within a DW ensures better integrity and higher-quality analysis for plant disease surveillance. Moreover, DW extensibility is mandatory for agricultural epidemiology, meaning that new schema elements can be incrementally added without redesigning the entire system [5]. While traditional schema-on-write approaches are rigid and time-consuming, a *Multi-Model Data Warehouse* (MMDW) that relies on a schema-on-read approach allows for agile, near-instant OLAP queries on new data [11,12]. This agility is vital for epidemiological surveillance under the *One Health* framework, where scientific knowledge and pathogen life cycles evolve rapidly [18,27]. For example, recent findings on FD regarding landscape geometry, heavy-tailed dispersal kernels, and phytoplasma genotypes necessitate flexible systems to refine control strategies [3,16].

Finally, analytical tools must be inherently extensible to integrate these diverse, emerging data types and multidisciplinary insights without technical bottlenecks. To support diverse actors like epidemiologists, policy makers and farmers who may lack data science expertise, OLAP tools provide intuitive drag-and-drop interfaces for multidimensional analysis. However, standard OLAP lacks the cartographic visualization and spatial analysis methods essential for tracking disease outbreaks [6]. SOLAP bridges this gap by integrating GIS capabilities, allowing for sophisticated spatio-temporal queries and interactive mapping of massive datasets [18]. This integration is crucial for identifying infection hotspots, understanding dispersal dynamics, and optimizing resource allocation for pest control. Ultimately, combining OLAP and GIS creates a powerful analytical platform for epidemic detection, predictive modeling, and the overall protection of crops.

Motivated by the lack of proposals of DW and OLAP systems for agricultural epidemiology, we propose in this paper a Business Intelligence architecture that supports the SOLAP analysis of FD observations according to a multi-model and extensible approach. The paper is organized in the following way: Sect. 2 presents the related work in OLAP and plant health management, Sect. 3 outlines our solution dedicated to the analysis of FD observations, and Sect. 4 details the implementation of each layer of our architecture to the FD monitoring data of GDON of Bordeaux between 2017 and 2020. Some examples of SOLAP analyses offered to the decision-makers are shown in Sect. 5.

2 Related Work

The role of OLAP tools for managing plant pests and diseases is emphasized in [10,17,22,25]. Although these works are conceptual, they highlight urgent needs: for example, [22] proposed a web solution for Florida citrus diseases using algorithms to merge agricultural data for real-time mapping and OLAP reporting. Similarly, [10] identified the multidimensional and spatio-temporal nature

of pest data, leading them to evolve their database schema towards an OLAP approach to analyze pest pressure by farm, crop, and region. [25] also advocates for OLAP to enrich the analysis of environmental variables and crop growth. At the implementation level, [26] developed an OLAP system for olive crop Integrated Pest Management, integrating the Growing Season Index [15] to farming data in order to compare farm performance under similar climatic conditions; this system uses a relational storage model with a Mondrian server but lacks GIS integration.

In the context of cotton management, [2] established the *Pest-Pesticide-Metereology Data Warehouse* (PPM-DWH) in Pakistan. Using a constellation schema, it correlates major facts (production and pest quantity) with dimensions like weather and pesticide use. This work was expanded by [1] into the ADSS-OLAP system, a MOLAP (Multidimensional OLAP) implementation using Microsoft Analysis Services. Similarly, [19] utilized Microsoft Business Intelligence to provide multidimensional exploration of rice pest data in Vietnam for farmers and researchers.

Addressing Big Data challenges, [20] introduced an architecture to manage the "four Vs" (volume, variety, velocity, veracity). Critiquing the rigidity of star schemata [21], they adopted a flexible constellation schema featuring 3 fact tables and 19 dimensions. They argue that traditional relational approaches [23,24] and RDF triplestores suffer from performance issues during complex analytical queries. The polyglot architecture they propose utilizes Hive, MongoDB, and Cassandra for better scalability, though it relies on predefined HQL/SQL queries rather than on a dedicated OLAP server. While handling diverse data like JSON, this approach faces typical polyglot limitations regarding extensibility.

Overall, these works highlight the potential of OLAP systems in the agricultural field in general, and specifically for plant disease surveillance. The integration of these technologies in agriculture marks a significant advance in how agricultural data is analyzed and used for decision-making. OLAP systems allow for the exploration and analysis of large volumes of agricultural data across multiple dimensions, thereby facilitating decision-making. The studies described in this section demonstrate a diversity of applications in agriculture, ranging from the monitoring of plant diseases and pests to water resource optimization and crop management. These systems utilize data from various sources, notably those provided by weather stations and sensors, as well as data related to different crops or soil characteristics.

Table 1 lists some case studies regarding the use of OLAP tools for plant disease surveillance. To evaluate the case studies listed in Table 1, we have identified the following criteria: the type of pest/bio-aggressor, the type of data stored in the DW, the OLAP server used (either MOLAP or ROLAP, i.e., Relational OLAP), and the integration of a GIS.

The nature of the pest and the associated crop are important for assessing the specificity of each study as well as its socio-economic impact. In particular, [1,2, 19,22,26] respectively study pests affecting olive trees, rice, cotton, and citrus. These studies cover a fairly wide range of pests, from insects (mealybugs, flies,

Table 1. Existing works on OLAP and pest analysis

Paper	Pest	Data type	OLAP Server	GIS
[1]	Mealybug (cotton)	Relational	MOLAP	No
[2]	Cotton pests	Relational	MOLAP	No
[10]	Not specified	Relational	Pre-built queries	No
[17]	Not specified	Relational	Pre-built queries	No
[19]	Rice pests	Relational	MOLAP	No
[20]	Not specified	Rel./Document	Pre-built queries	No
[22]	Citrus insects and pests	Relational	Pre-built queries	No
[25]	Not specified	Relational	Pre-built queries	No
[26]	Olive fruit fly and moth	Relational	ROLAP	No

and moths) to pathogens (microorganisms associated with cankers and bacteria responsible for citrus "greening"). These works present many similarities with our case study, which focuses on the design of a SOLAP tool for the surveillance of FD, a phytoplasma disease of the grapevine spread by a leafhopper.

The second criterion chosen for evaluating these works is the type of data stored within the DW. This criterion is essential for assessing the proposed solution's ability to adapt to the evolution toward big data, thus, toward data that is voluminous and varied. The majority of the mentioned studies are limited to the storage of relational data, except for [20], which stores and analyzes JSON documents in addition to relational data. Thus, in most cases, the issue of agricultural data variety is addressed during the transformation phase of the Extraction-Transform-Load (ETL) process. [20] provides an innovative response to this problem by proposing a polyglot system; however, it is not accompanied by an OLAP server implementation capable of performing multidimensional analysis operations.

The type of OLAP server constitutes the third evaluation criterion, as it represents the central core of an OLAP system. MOLAP servers, such as those used in [1,2,19], generally offer better performance than ROLAP servers for multidimensional analysis operations. However, it is important to emphasize that ROLAP servers are better suited for integration with GIS [8], as well as for managing data variety [9] and extensibility [7,12]. Nevertheless, the consideration of variety and extensibility in these works remains limited, as they use either MOLAP servers or ROLAP servers with exclusively relational data storage.

Although not integrated into any of the works mentioned above, the inclusion of a GIS could significantly enrich OLAP analyses in the field of epidemiological surveillance, because it would allow for the integration of a spatial dimension into the analyses. This is crucial for understanding the dispersal of pests and the impact of numerous spatially structured factors on their development. This spatial dimension is particularly relevant for the implementation of targeted surveillance or control strategies and for the optimization of field interventions.

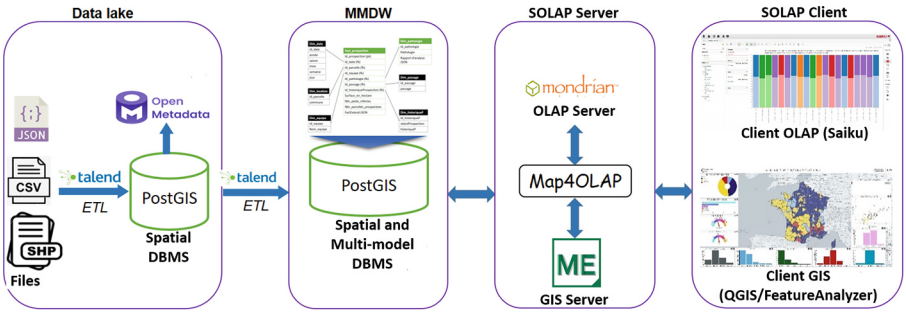

Fig. 1. Overall architecture.

In conclusion, it should be noted that despite the advantages offered by SOLAP tools, their adoption in agriculture faces several challenges. The complexity of agricultural data, the need to integrate GIS for spatial analysis, and the limitations of existing architectures are significant obstacles [6]. Furthermore, the design and development of these systems require close collaboration between agronomists, computer scientists, and decision-makers to ensure that the solutions developed effectively meet their needs.

3 Architecture

This section presents the architecture of our SOLAP solution dedicated to epidemiological surveillance. Inspired by previous works [7,8,12], this architecture is centered on an MMDW.

3.1 Overview

The architecture is an extension of the architecture proposed in [7,12]. It integrates the Map4OLAP tool [8], aimed at taking into account the geographic component of the spatial dimension, which is particularly important in the context of epidemiological surveillance. As illustrated in Fig. 1, it relies on four main layers: (i) Data Lake, (ii) MMDW, (iii) SOLAP server, and (iv) SOLAP client. It is a multi-model spatial data processing and analysis architecture, combining spatial data management technologies (PostGIS), ETL tools (Talend), as well as an OLAP front-end (Saiku) and GIS (QGIS or Feature Analyser) for decision-making analysis.

3.2 Data Lake Layer

In the Data Lake layer, raw data from different sources are stored in various formats, such as SHP files (Shapefile) for geospatial data, JSON files for semi-structured data, and CSV for structured data. This data is then integrated into

PostGIS, a spatial and multi-model database management system, via a simple ETL process performed with Talend Open Studio.

An important element of this layer is the management of metadata via the OpenMetadata tool. OpenMetadata is an open-source metadata management and cataloguing solution that centralizes descriptive information of the data in the data lake. It also stores the schemata of this data and a sample of the data itself. It thus allows for the cataloguing of data and searching via tags to facilitate identification and reuse. Data cataloguing facilitates the discovery of additional data that could be integrated into the MMDW layer to implement extensibility in case of evolving analytical needs for end-users.

Certain data present in the Data Lake layer are integrated into the next layer, MMDW, through a more complex ETL process that ensures not only data transfer but also its transformation and organization according to a multi-model star schema. This ETL process also ensures the high quality of data imported into the MMDW layer to guarantee the reliability of OLAP analyses; among other things, it includes managing missing and erroneous values, harmonizing municipality names, and replacing special characters.

3.3 MMDW Layer

The primary goal of the MMDW layer is to prepare data for decision-making exploitation through the SOLAP client layer. Here, data is stored in PostGIS and organized according to a multi-model star schema designed to meet specific analytical needs. This star schema will be described in more detail in Subsect. 4.1; it combines relational data and JSON documents, making it more flexible and enabling extensibility [9].

The MMDW layer communicates directly with the SOLAP server layer using *XML for Analysis* (XMLA), a computing standard designed to allow access to data in analytical systems, particularly in OLAP systems. This standard relies on other pre-existing standardized web technologies such as XML, SOAP, and HTTP.

The MMDW layer executes queries coming from the OLAP Server. These multi-model OLAP queries are implemented through a mix of languages like SQL for relational data, Cypher for graphs, and the appropriate language for querying JSON documents (see Sect. 4.3). After the execution of these queries by PostGIS, the MMDW layer returns the corresponding results to the OLAP server, which writes them into a well-structured file in the form of dynamic cross-tabulations.

3.4 SOLAP Server Layer

The SOLAP Server layer consists of an OLAP server (Mondrian), the Map4OLAP tool, and a GIS server (M.App Enterprise). These components ensure the exploitation of data from the MMDW layer for spatial and multi-dimensional analyses.

- Mondrian allows the exploration and analysis of multidimensional data via OLAP cubes; it is responsible for translating MDX (Multidimensional Expressions) queries from the OLAP client into queries adapted to the MMDW layer where the data is stored. Mondrian also reads the XMLA file resulting from the OLAP queries and sends it to the OLAP client.
- Map4OLAP is a Java program that links the OLAP server (Mondrian) and the GIS server (M.App Enterprise). It retrieves the OLAP server results written in XMLA and transforms them into GeoJSON files compatible with the GIS server. This process also includes decision rules on how the GIS client should visualize interactive maps. For example, it allows defining specific graphical representations (such as pie charts or histograms) for polygons (such as municipalities), while specifying how to handle missing or null values.
- M.App Enterprise is used to manage cartographic aspects. It is responsible for making the cartographic data (created by *Map4OLAP*) available to the GIS client. Furthermore, it implements features associated with interactive maps.

3.5 SOLAP Client Layer

The SOLAP Client layer consists of two main elements:

- The OLAP Client we adopt is Saiku. It works in a traditional manner, allowing end-users to interact with OLAP cubes and perform multidimensional analyses via an intuitive user interface. It translates dynamic cross-tabulations into MDX queries sent to the OLAP server.
- The GIS Client is used to visualize the maps created by Map4OLAP. We currently rely on QGIS, but we plan to replace it with Feature Analyser to implement all the functionalities of Map4OLAP.

The integration of OLAP and GIS clients in this layer ensures that the analysis of multidimensional data and its cartographic visualization are carried out in an efficient and fluid manner.

4 Implementation

This section describes in more detail the implementation of the MMDW layer (Sect. 4.1) and of the OLAP server component of the SOLAP layer (Sect. 4.2). The extension of the DW with additional complex data is described in Sect. 4.3. Data used are collected by GDON from 2017 and 2020.

4.1 MMDW Layer

Usually, a DW schema is composed by the fact table, which contains measures, and a set of dimension tables that represent the dimensions levels. The logical schema of our MMDW is shown in Fig. 2.

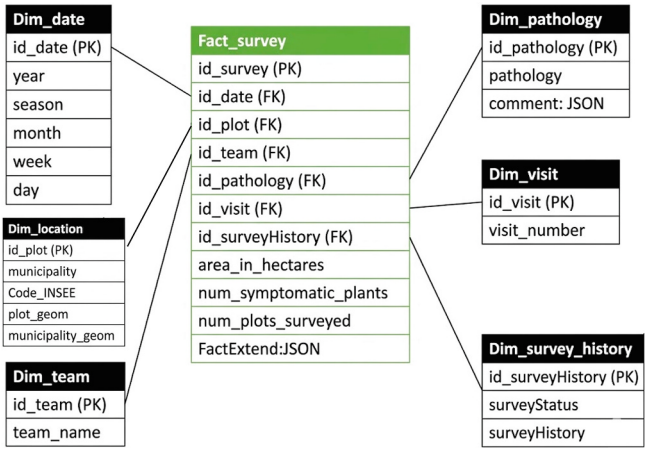

Fig. 2. Logical schema of the MMDW for FD Monitoring.

It hinges on a fact table, Fact_survey, defined at the finest possible granularity, i.e., at the plot and day level (differently from the dashboards defined by the GDON, which show data by year and across the intervention zone of the GDON Bordeaux in Gironde). Our approach allows for analyzing data with full detail for each survey and explains aggregated data. This fact table includes a primary key, the set of foreign keys from the dimensions, the measures, and a JSON-type column to ensure extensibility. Its columns are:

- id_survey is the primary key.
- id_date is a foreign key referencing the Dim_date dimension table.
- id_plot is the identifier of the plot where the survey took place. It is assigned by the GDON to each plot and is a foreign key referencing the Dim_location dimension table.
- id_team identifies the team that performed the survey (foreign key referencing Dim_team).
- id_pathology identifies the disease detected during the survey and subsequent analyses (foreign key referencing Dim_pathology).
- id_visit identifies the visit/pass (foreign key referencing Dim_visit).
- id_survey_history, area_in_hectares, num_symptomatic_plants, and num_plots_surveyed are measures; their explanation will be given in Sect. 4.2.
- FactExtend is a JSON field that facilitates the integration of additional information specific to each survey.

As to the dimension tables:

- The time dimension table, Dim_date, provides different aggregations of the date when the survey was conducted.

- The spatial dimension table, Dim_location, provides geographical information about the plot. Besides the INSEE code and name of the municipality, it includes a spatial representation of the geometry of the plot and of the municipality.
- The Dim_team dimension table provides details about the GDON team performing the survey.
- The Dim_pathology dimension table represents information related to the detected pathologies.
- The Dim_visit dimension table provides information regarding the visits (passes) conducted during the survey (visit_number is a Boolean value that represents if it is the first or the second visit).
- The Dim_surveyHistory dimension table represents the survey history, specifying whether a plot has been surveyed or not. If the plot has already been surveyed, it specifies whether it is a first survey or not; it also indicates if the plot has been surveyed in both the current and the previous year.

4.2 OLAP Server

The OLAP server uses an XML file to generate queries to be sent to the MMDW layer. This file defines multidimensional elements such as measures, dimensions, and hierarchical levels; it also specifies the mappings between these multidimensional elements and the data stored in the MMDW layer. For example, it indicates to the OLAP server that Fact_survey is the fact table and that Dim_location contains data related to the location dimension. The OLAP server also specifies the measures and their aggregation functions. In our implementation we have defined the following measures:

- Number of symptomatic plants (corresponding to num_symptomatic_plants in the fact table) is an additive measure (i.e., it can be aggregated using the sum on all the dimensions) referring to the number of symptomatic plants observed by agents during survey visits.
- Number of surveyed plots (num_plots_surveyed in the fact table) represents the number of plots surveyed. It is calculated using the `distinct count` operator on the result of concatenating the plot id and the survey year. This ensures that, when a plot is surveyed in two different years, it is counted twice, whereas if it is surveyed twice in the same year (in the case of a second pass), it is counted only once.
- Area in hectares (area_in_hectares in the fact table) is a semi-additive measure (i.e., it cannot be aggregated using the sum along all the dimensions) because, although the area surveyed at a municipality level is obtained by adding the areas of the plots in that municipality, and the area surveyed during a year corresponds to the sum of the areas of plots surveyed that same year, this does not apply to second passes. Indeed, if a plot is surveyed twice in the same year, its area should not be counted twice. Thus, this measure is additive across spatio-temporal dimensions but not across the visit dimension.

- Number of symptomatic plants per hectare is derived by dividing the number of symptomatic plants by the surveyed area in hectares to measure the density of symptomatic plants.
- Multiplication rate of symptomatic plants measures the year-over-year change in the number of symptomatic plants.

All these measures are implemented using the `Measure` and `Calculated Member` statements of Mondrian.

4.3 Extensibility

To give our MMDW the extensibility required by the application, we adopted the *xCube* approach, proposed in [7] to provide on-demand, supply-driven extensibility of multidimensional cubes. *xCube* allows users to select a specific element within the schema of the DW and expand it using supplementary data, often sourced from a data lake. The system then extends the multidimensional schema by incorporating the functional dependencies found within this new data, providing users with an updated schema ready for immediate OLAP analysis. *xCube* follows these steps:

1. Selection of the target element: the user identifies the specific component within the schema of the DW to be expanded——either the fact table (to add new measures or dimensions) or a specific level (to introduce new hierarchies). The columns used for extension must be defined as a JSON type.
2. Data sourcing: the user selects supplementary data from the data lake, which may include relational tables, document collections, or graphs. These datasets are then automatically imported into the MMDW layer.
3. Schema extension: the system extends the schema of the DW by identifying functional dependencies between the selected schema element and the newly imported data.

The additional data we used for the extension of our MMDW are taken from the Casier Viticole Informatisé (CVI), a digital tool established in 1998 by the viticulture services of the French directorate general of customs. It lists all vineyard cadastral plots in France along with their characteristics, such as owner, area, grape variety, planting year, planting density, and spacing, among others. We used an anonymized version of the 2016 CVI for the Bordeaux territory. Figure 3 shows an excerpt of the JSON CVI data, where the French terms "refCadastrale", "cepage", "typeCepage" and "densite" represent the plot identifier, the grape variety, the type of variety, and the number of plants, respectively. The protocol for attributing CVI variables to fields sampled by GDON is described in [4]. This data is stored in the Data Lake layer, and it was not integrated at the time the MMDW of Fig. 2 was created.

Following the *xCube* approach, it was decided to extend the MMDW with a new measure quantifying the density of plants, which corresponds to the densite attributes in the CVI JSON data. Therefore, the MMDW was extended by designers with a new measure number_of_plants, fed via an SQL query in the

```
{
    "refCadastrale": "refCadastrale1",
    "cepage": "semillon B",
    "typeCepage": "cepage blanc",
    "densite": 5000
},
{
    "refCadastrale": "refCadastrale2",
    "cepage": "CABERNET FRANC N",
    "typeCepage": "cepage noir",
    "densite": 4000
},
```

Fig. 3. Additional CVI data.

XML Mondrian file as shown in Fig. 4. It is important to note that, since this measure is computed from the JSON document stored in the FactExtend column of the fact table, it is defined as a query over a JSON document. This is technically possible since the PostGIS extension of the PostgreSQL DBMS used for the implementation of the MMDW layer supports multi-model data storage and querying.

```
<Measure name="Number_of_plants" column="nb_plants" aggregator="avg">
  <SQL>
    SELECT
      COALESCE((Fact_survey.FactExtend->>'densite')::INTEGER, 4000) AS nb_plants
    FROM Fact_survey
  </SQL>
</Measure>
```

Fig. 4. Fact extended with a new measure.

5 SOLAP Analysis

In this section, we show the results of some analyses that could be conducted by decision-makers for the monitoring of FD. Due to confidentiality requirements, the data presented here has been modified and does not represent real values.

Figure 5 shows the number of symptomatic plants identified per plot and per day. Note that plot "0002" was surveyed on 7th September 2018, 3rd September 2019, and 7th August 2020, with 14, 8, and 70 symptomatic plants identified, respectively. Thus, Saiku allows for a precise analysis of a plot's surveying history. Figure 6 shows the distribution of the number of symptomatic plants per plot for three years (only five plots are shown for better readability), plot "0005" recording the highest number of symptomatic plants. Thus, these pie charts help identify the most affected plots.

Figure 7 shows a choropleth map of the vineyard plots in a specific municipality. In this map, the color of the plots varies according to the total number of symptomatic plants observed between 2017 and 2020. The data presented in this

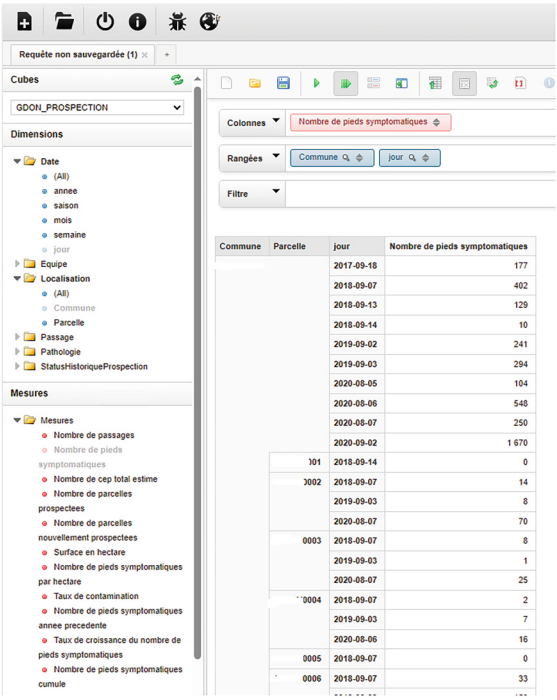

Fig. 5. Number of symptomatic plants per plot and per day.

map were obtained by aggregating the data illustrated in Fig. 5 across all years. This corresponds to a *Roll-Up* operation applied to the Date dimension. The value categories are defined using the quantiles method, so as to distribute the number of plots equally across each of the 10 categories. This approach prevents the majority of plots from falling into the first class, given that they generally have a very low number of symptomatic plants.

Thanks to this representation, it is easy to identify the plots requiring special attention, particularly those belonging to the bottom classes (in red), characterized by more than 75 symptomatic plants. Furthermore, unlike visualization techniques such as pie charts or evolution curves, which lose readability as the number of plots increases, the choropleth map remains legible. Finally, this map allows for the identification of potential contamination hotspots (clusters of geographically close plots with many symptomatic plants). For example, it appears that several plots located in the northeast of the municipality (top and right part of Fig. 7) stand out due to a high presence of symptomatic plants.

Finally, Fig. 8 shows a map of the vineyard plots in the same municipality, where each plot is associated with a bar chart consisting of four bars, each representing the number of symptomatic plants observed between 2017 and 2020. The data presented in this map were obtained by aggregating the data in Fig. 5 by year rather than by day. This corresponds to a *Roll-Up* operation applied

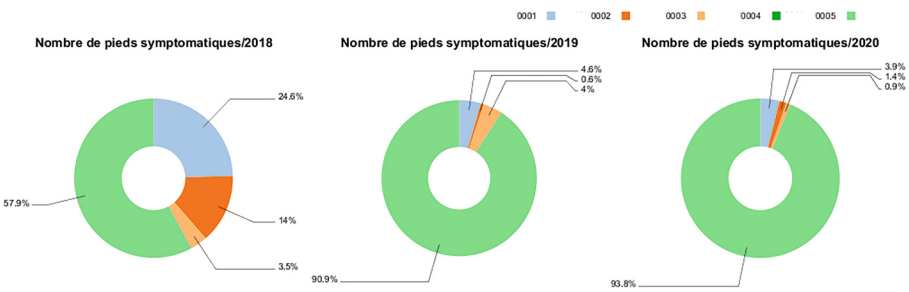

Fig. 6. Evolution of the distribution of symptomatic plants per plot.

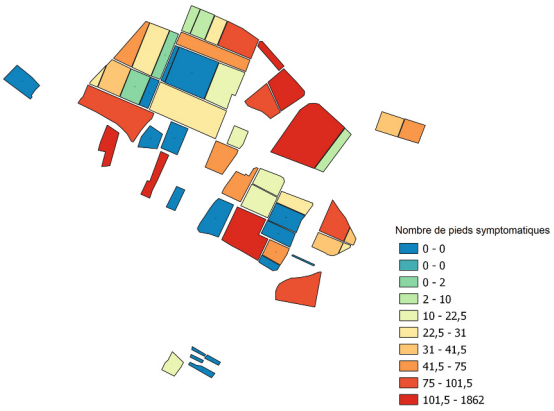

Fig. 7. Choropleth map of the total number of symptomatic plants (observed between 2017 and 2020) per plot in a specific municipality. (Color figure online)

to the Date dimension. This figure complements Fig. 7 by showing, for example, that the plots in the northeast of the municipality identified previously record a relatively low number of symptomatic plants compared to plot "0005" to the west, where more than 1,700 symptomatic plants were identified in 2020. Indeed, the northeast plots have symptomatic vine counts closer to 100 than to 1,800.

6 Conclusion

In this work, we described the design and implementation of a MMDW dedicated to the monitoring of FD in the Bordeaux wine region. The proposed SOLAP architecture integrates the key components necessary for the analysis of spatial and multidimensional data. In particular, the use of the Map4OLAP tool enriched the DW functionalities by providing cartographic visualization capabilities and interaction with spatial data. The results obtained show that the designed MMDW could meet the GDON's needs for managing and analyzing survey data. It facilitates the production of detailed reports, the spatio-temporal

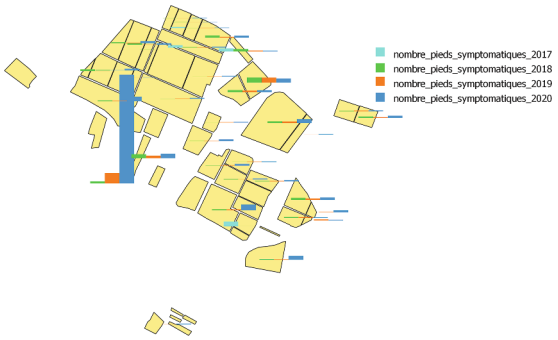

Fig. 8. Evolution map of the number of symptomatic plants (observed between 2017 and 2020) per plot in a specific municipality.

tracking of FD, and can adapt to the evolving analytical needs of users. Furthermore, the established architecture provides a foundation for integrating new data sources and extending the tool's analytical capabilities, thus reinforcing its relevance in the context of epidemiological surveillance.

In conclusion, beyond offering a dashboard of FD distribution, the MMDW has the potential to drive the development of predictive risk maps. By cross-referencing domain expertise with model outputs, operational stakeholders could use these maps to significantly improve FD monitoring strategies. For instance, risk mapping could leverage graph-based methods to explicitly model infectious "contacts" between plots, offering a robust framework for simulating epidemic spread over real-world vineyard landscapes. The probability of FD propagation would be dynamically updated year-over-year at the local level through a model incorporating: (i) the structural graph of the vineyards (derived from CVI data); (ii) empirical knowledge of FD dispersal (e.g., [3]; (iii) historical data from prior field surveys; and (iv) key risk factors (e.g., cultivar type, landscape composition and configuration, proximity to abandoned vines; [4]).

Acknowledgements. This work is partially funded by the BEYOND ANR-20-PCPA-0002 project and the 101226371-HORIZON MSCA DN-JD grant "IoRT Data Management and Analysis for Sustainable Agriculture (GreenFieldData)". We sincerely thank Sophie Bentejac and Charlotte Labit from the GDON of Bordeaux who sampled and provided data and for their help for formatting data.

References

1. Abdullah, A.: Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool. Comput. Electron. Agric. **69**(1), 59–72 (2009)
2. Abdullah, A., Brobst, S., Umer, M., Khan, M.F.: The case for an agri data warehouse: enabling analytical exploration of integrated agricultural data. In: Proceedings of Databases and Applications, pp. 139–144 (2004)

3. Adrakey, H.K., Gibson, G.J., Eveillard, S., Malembic-Maher, S., Fabre, F.: Bayesian inference for spatio-temporal stochastic transmission of plant disease in the presence of roguing: a case study to characterise the dispersal of flavescence dorée. PLoS Comput. Biol. **19**(9), e1011399 (2023)
4. Adrakey, H.K., et al.: Field and landscape risk factors impacting flavescence dorée infection: insights from spatial Bayesian modeling in the Bordeaux vineyards. Phytopathology® **112**(8), 1686–1697 (2022)
5. Bazza, H., Bimonte, S., Rizzi, S., Laneurit, J., Badir, H.: A UML profile for variety and variability awareness in multidimensional design: an application to agricultural robots. In: Proceedings of DOLAP, pp. 1–10 (2022)
6. Bimonte, S.: Current approaches, challenges, and perspectives on spatial OLAP for agri-environmental analysis. Int. J. Agric. Environ. Inf. Syst. **7**(4), 32–49 (2016)
7. Bimonte, S., Coulibaly, F.A., Rizzi, S.: An approach to on-demand extension of multidimensional cubes in multi-model settings: application to IoT-based agro-ecology. Data Knowl. Eng. **150**, 102267 (2024)
8. Bimonte, S., Edoh-Alove, E., Coulibaly, F.A.: Map4OLAP: a web-based tool for interactive map visualization of OLAP queries. In: Proceedings of Big Data, pp. 3747–3750 (2021)
9. Bimonte, S., Gallinucci, E., Marcel, P., Rizzi, S.: Data variety, come as you are in multi-model data warehouses. Inf. Syst. **104**, 101734 (2022)
10. Boll, R., Bout, A., Doise, S., Bernardet, X., Poncet, C.: "Data_market" as a realistic tool for decision-making in the integrated pest management challenge. In: Proceedings of IHC2010, pp. 163–170 (2010)
11. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: Mad skills: new analysis practices for big data. Proc. VLDB Endow. **2**(2), 1481–1492 (2009)
12. Coulibaly, F.A., Sandro, B., Rizzi, S.: Towards a multi-model approach to support user-driven extensibility in data warehouses: agro-ecology case study. In: Proceedings of DataPlat, vol. 3379, pp. 1–5 (2023)
13. (EFSA), E.F.S.A., Tramontini, S., Delbianco, A., Vos, S.: Pest survey card on flavescence dorée phytoplasma and its vector scaphoideus titanus. EFSA Support. Public. **17**(8), 1909E (2020)
14. Gent, D.H., Mahaffee, W.F., McRoberts, N., Pfender, W.F.: The use and role of predictive systems in disease management. Annu. Rev. Phytopathol. **51**, 267–289 (2013)
15. Jolly, W.M., Nemani, R., Running, S.W.: A generalized, bioclimatic index to predict foliar phenology in response to climate. Glob. Change Biol. **11**(4), 619–632 (2005)
16. Malembic-Maher, S., et al.: When a palearctic bacterium meets a nearctic insect vector: genetic and ecological insights into the emergence of the grapevine flavescence dorée epidemics in Europe. PLoS Pathog. **16**(3), e1007967 (2020)
17. Mir, S., et al.: Decision support systems in a global agricultural perspective - a comprehensive review. Int. J. Agric. Sci. **7**(1), 403–415 (2015)
18. Morris, C.E., Géniaux, G., Nédellec, C., Sauvion, N., Soubeyrand, S.: One health concepts and challenges for surveillance, forecasting, and mitigation of plant disease beyond the traditional scope of crop production. Plant. Pathol. **71**(1), 86–97 (2022)
19. Nghia, P.Q.: Building OLAP application to exploit database of rice pests. Int. J. Emerg. Technol. Adv. Eng. **3**(11), 130–137 (2013)
20. Ngo, V.M., Le-Khac, N.A., Kechadi, M.T.: Designing and implementing data warehouse for agricultural big data. In: Proceedings of Big Data, pp. 1–17 (2019)
21. Nilakanta, S., Scheibe, K., Rai, A.: Dimensional issues in agricultural data warehouse designs. Comput. Electron. Agric. **60**(2), 263–278 (2008)

22. Rogers, S.: Real-time internet-PDA-based survey and modeling systems for agricultural insects and diseases: integrated fruit fly, citrus canker and greening surveys. In: Proceedings of Florida State Horticultural Society, vol. 119, pp. 107–119 (2006)
23. Schuetz, C.G., Schausberger, S., Schrefl, M.: Building an active semantic data warehouse for precision dairy farming. J. Organ. Comput. Electron. Commer. **28**(2), 122–141 (2018)
24. Schulze, C., Spilke, J., Lehner, W.: Data modeling for precision dairy farming within the competitive field of operational and analytical tasks. Comput. Electron. Agric. **59**(1–2), 39–55 (2007)
25. Wang, C., Wang, L., Dong, P., Qiao, X.: Research of decision support system (DSS) for greenhouse based on data mining. In: Proceedings of CCTA, pp. 1071–1076 (2008)
26. Zaza, C., Bimonte, S., Faccilongo, N., La Sala, P., Contò, F., Gallo, C.: A new decision-support system for the historical analysis of integrated pest management activities on olive crops based on climatic data. Comput. Electron. Agric. **148**, 237–249 (2018)
27. Zinsstag, J., Schelling, E., Crump, L., Whittaker, M., Tanner, M., Stephen, C.: One Health: The Theory and Practice of Integrated Health Approaches. CABI (2020)