Open Spatiotemporal Data Warehouse for Agriculture Production Analytics

Irya Wisnubhadra[1, 2]*   **Safiza Suhana Kamal Baharin[2]**   **Nanna Suryana Herman[2]**

[1]*Informatics Engineering Department, Universitas Atma Jaya Yogyakarta, Indonesia*
[2]*Centre for Advanced Computing Technology, Faculty of Information and Communication Technology,*
Universiti Teknikal Malaysia Melaka, Malaysia
* Corresponding author's Email: irya.wisnubhadra@uajy.ac.id

Abstract: Business Intelligence (BI) technology with Extract, Transform, and Loading process, Data Warehouse, and OLAP have demonstrated the ability of information and knowledge generation for supporting decision making. In the last decade, the advancement of the Web 2.0 technology is improving the accessibility of web of data across the cloud. Linked Open Data, Linked Open Statistical Data, and Open Government Data is increasing massively, creating a more significant computer-recognizable data available for sharing. In agricultural production analytics, data resources with high availability and accessibility is a primary requirement. However, today's data accessibility for production analytics is limited in the 2 or 3-stars open data format and rarely has attributes for spatiotemporal analytics. The new data warehouse concept has a new approach to combine the openness of data resources with mobility or spatiotemporal data in nature. This new approach could help the decision-makers to use external data to make a crucial decision more intuitive and flexible. This paper proposed the development of a spatiotemporal data warehouse with an integration process using service-oriented architecture and open data sources. The data sources are originating from the Village and Rural Area Information System (SIDeKa) that capture the agricultural production transaction in a daily manner. This paper also describes the way to spatiotemporal analytics for agricultural production using a new spatiotemporal data warehouse approach. The experiment results, by executing six relevant spatiotemporal query samples on DW with fact table contains 324096 tuples with temporal integer/float for each tuple, 4495 tuples of field dimension with geographic data as polygons, 80 tuples of village dimension, dozens of tuples of the district, subdistrict, province dimensions. The DW time dimension contains 3653 tuples representing a date for ten years, proved that this new approach has a convenient, simple model, and expressive performance for supporting executive to make decisions on agriculture production analytics based on spatiotemporal data. This research also underlines the prospects for scaling and nurturing the spatiotemporal data warehouse initiative.

Keywords: Spatiotemporal, Data warehouse, Open data, Agriculture production analytics.

1. Introduction

The Indonesia Government policy in the last decade still concerns about agricultural sector development with information technology support. Agricultural growth in Indonesia is associated with village growth because the location of the traditional agrarian land is mostly in the village. At this moment, the primary producer of farming products is coming from the village. Farming products make a significant contribution to the Gross Domestic Product (GDP). Between 2013 and 2018, the GDP of the agricultural sector consistently showed a positive trend. In 2013

the Agriculture Sector's GDP was 847.8 trillion Indonesian Rupiah and increased to 1000.54 trillion in 2018 and contributed 13.53% of the Indonesia total GDP.

Indonesia, as an agrarian country, is still facing difficulties in achieving sustainable food safety. These problems relate to numerous aspects of social, political, economic, cultural, and technology. Food safety is referred to as food resources and food needs. Food resources data and information currently are officially centralized under the Indonesian Statistics Central Bureau and the Agriculture Ministry generation and supervision. But these institutions still

do not have an efficient approach to capture the agricultural data and predict the food from farming production for decision making related to the fulfillment of domestic food needs.

In the context of sustainable food safety, many queries have to be answered as a foundation for decision making support related to spatial and temporal dimensions, such as (a) How many tons of rice farming production in District X in the first quarter this year? (b) How much fertilizer needs in the sub-district X for the next two months? (c) How many areas of soybeans farming failed because of the pests last year, in a predetermined location, and aggregated on a sub-district or district basis?

The queries could be answered using the data that captured using the conventional way: (a) collecting data at the rice production each rice fields/parcels in an old-fashioned way. This technique gives excellent accuracy, but this approach needs many resources and costly because it must be implemented massively and regularly. (b) using satellite imagery to assess land crop productivity [1], but this is the expensive approach because the remote sensing process has to be carried out periodically and using costly satellite technology. This paper will present the new way of collecting the spatiotemporal data, processing it to become information, and visualizing the information with spatial and temporal dimensions with open data concept.

Integration agricultural data in Indonesia has been pioneered with the One Data movement (https://data.go.id), but this data still have general attributive data, did not have a detailed location and temporal aspect. This data is rarely having the complete progress of agriculture production step.

The farming cycle, from the farmer's point of view, has three cycles, e.g.: (1) Pre-planting: comprises credit/loan finding, planting time provision, land selection, seed assortment process. (2) Planting and harvesting: provides soil preparation and planting, planting management, management of water, fertilization, pesticide sowing process, and harvesting (3) Post-harvest: includes marketing, packaging, transporting, and product post-processing [2]. In each step of the agricultural cycle needs IT intervention to make the process more efficient and productive. Figure. 1 shows the cycle of agriculture [2].

The idea of open data drew the attention of the Government and researchers since the 5-star open data concept emerge. Many governments participate in sharing and publishing their data in an open concept, such as the US, UK, France, Australia, Indonesia, etc. Open Data Government gives the

Figure. 1 Farming cycle

beneficial value in transparency, citizen service, effectiveness and efficiency, innovation, and boost economic growth. Indonesia has a 25% score of Global Open Data Index rating and is positioned at 61 of 94 countries.

Open Data has a five-category index, namely: (a) 1-Star: Data is stored in various format, accessible on the internet, and pursues the open license rules, (b) 2-Stars: Data is accessible as computer-readable structured data (e.g., spreadsheet instead of picture scan), (c) 3-Stars: Data accessible as 2-Star and have non-proprietary formats (e.g. CSV, instead of Excel), (d) 4-Stars: Data accessible as 3-Stars and have W3C open standard usage to identify things so that individuals can point at that things, and (e) 5-Stars: Data available as 4-Stars and also can link to people's data to provide context (Linked Open Data).

This research explains the development of the Open Data Warehouse system for agriculture production spatiotemporal analytics that incorporates multiple data sources with 4 and 5-stars open data format. The source data has distributed in nature and was provided by the Village and Rural Area Information System (SIDeKa) agricultural module that has been studied earlier.

Daily agricultural data production steps from SIDeKa are publicly shared using application programming services and as Linked Open Data (LOD) that could be integrated and aggregated into the Data Warehouse (DW) for analytical purposes. The visualization of the fact of agricultural production could be used for descriptive and diagnostic analytics for supporting decision making in the context of food security.

For supporting spatiotemporal analysis, some data warehousing techniques can be used, including spatiotemporal DW or Trajectory DW. Trajectory

DW usually has a fact table with the segments of spatiotemporal or trajectories as a measure associated with spatial and non-spatial dimensions. This paper goes beyond this concept by including, as measures, the moving states of objects at any point in time. In this way, online analytical processing (OLAP) queries, typically including aggregation, can be combined with spatiotemporal queries, to express queries like "Find out how much average crop potential agricultural loss (currency) suffered by the Boyolali district because of the eruption of Mount Merapi on January 21, 2020. It is assumed that the affected area is at a radius of 5 km." in a concise and elegant way. Existing proposals for trajectory data warehouses do not support queries like this, since they are based on either the segmentation of the trajectories or a pre-aggregation of measures.

This research has organization as, in Section 2, studies the related works and state of the art of open spatiotemporal DW. Section 3 describes the research methodology of this research. Section 4 describes the discussion and results of this research, and Section 5 concludes the paper and describes the future works.

2. Related works

2.1 Data warehouse for government services

In the last two decades, the Business Intelligence (BI) implementation has increased significantly with the growing number of BI tools available on the market. The core concept of BI is the Data Warehouse (DW) as an enterprise data repository of the organizations. DW is defined as a technology for assisting decision-makers in making a tactical or strategic decision that has characteristics as subject-oriented, non-volatile, time-varying, and integrated [3].

DW has proven to provide technology for supporting decision making, including in the governmental institution. Liu et al. built e-GovMon Data Warehouse for e-government data repository service. The E-GovMon developed using open source technologies that deploy PostgreSQL DBMS, operational service data as a data source, and a right-time ETL tool to get, transform, and populate the data. This DW has distributed architecture, but this project does not consider the shareability of the data source as open data[4]. Furthermore, Maliappis et al. (2015) proposed the implementation of OLAP DW for agricultural decision making in Greek. This research tries to integrate numerous agricultural data sources that have different natures, multiple dimensions, granularity, and user requirements. In this research, the authors develop a suitable approach for producing information efficiently using the data warehouse.

The study utilizes Pentaho Data Integration Tools for ETL processes and Open Source RDBMS as datastore [5]. Arora and Gupta (2017) describe the implementation of DW in India. The DW could support the Government in increasing efficiencies, decreasing cost across some domains, including the agriculture sector, transportation, education, healthcare, etc. [6]. Wisnubhadra et al. (2017) develop BI for sustainable development planning as a part of e-government with distributed architecture in Indonesia. However, this paper is still using isolated data and not using the open data concept for shareability[7]. The implementation of the DW in [4-7] has an impressive result for giving government support for decision making, but all studies still have disadvantages for supporting the shareability of data sources. Even these researches using open source technology, but the open concept of the data still in limited exposure.

2.2 Open data warehouse

The significant growth of internet connection is changing the approach of the design of BI and DW. In the era of Web 2.0, new ideas such as Linked Open Data and Open Government Data bring many organizations, including the government body, willing to share the data and information they have.

Salas et al. proposed a way to extract, publish, and manage statistical data on the web. This research develops two approaches for extracting and publishing statistical data using OLAP2DataCube and CSV2DataCube. The two methods use the Resource Description Framework (RDF) Data Cube vocabulary and are linked with the GeoNames for geographical dimensions and DBpedia. The statistical data also have been linked and visualized for describing ten years of life in Brazil [8].

Research by Asano et al. (2016) build an approach for publishing and sharing open data for the Japan Ministry of Economy, Trade, and Industry. This research proposed two site designs for share and published data, including Linked Open Data (LOD) searching and data directory site. The first feature provides users with searching capability for LOD, and the second feature offers a data catalog display that could be customized by the needs of the institution [9]. This research is extended by Matsuda et al. (2018), who proposed the Japanese Statistic Centre that published approximately 1.3 billion triples. This publication of the statistical data warehouse introduces the design policy of LOD and methods for representing a geographical area. The result of the data could be linked with data from domestic or global data[10].

Furthermore, the development of tools for Linked Open Statistical Data (LOSD) dataset management was conducted by Stasiewicz et al. (2018). These tools are taking advantage of marine open statistical datasets that coming from maritime rescue digital services. This research emphasizes the growing interest to policymakers, academia, and industry for producing and publishing digital public services using multidimensional data cubes [11]. Moreover, Digital Earth Australia (DEA) project, a project funded by the Australian Government to improve, shield, and evaluate the environment annually, is using big data technology and published the data as an Open Cube. This project has various plans, including Great Barrier Reef protection by land management, farming efficiency evaluation, and critical wetland protection in Murray -Darling Basin[12].

Moreover, Klimek et al. (2018) proposed the Czech Social Security Administration publication of LOD of official pension statistical data. The statistical data is represented by the RDF Data Cube Vocabulary (DCV) and by the Simple Knowledge Organization System (SKOS) vocabulary model. The data was persisted as RDF and CSV files for expert and novice users, respectively. This study also demonstrates that this data accessible and reusable for application development. This data also could be linked with other LOD to create statistical measures [13].

The LOD and the Semantic Web (SW) has drawn much attention from data researchers since the finding of new vocabulary for multidimensional data cubes. This vocabulary is called QB, became the W3C standard for distributing statistical data on the web. QB initially follows models for analyzing statistical information[14]. Alternatively, Etcheverry et al. created Open Cube (OC) vocabulary on RDF that closely follows the conventional MD models for OLAP. However, these two approaches have limitations. The QB vocabulary has problems for representing hierarchical data, but OC did not support data cubes data representation over QB[15]. The new vocabulary has been proposed, namely QB4OLAP. QB4OLAP has useful features for accomodating OLAP technology, like slice, dice, roll-up, and drill down that has limited functionality in QB. Linked Data represented in QB4OLAP could easily access using SPARQL queries[16]. Bouza et al. proposed the implementation of QB4OLAP to transforms data from relational DW into RDF DW. This implementation applies realistic case of health services, citizen wage, and housing survey carried by the Government of Uruguay[17]. However, writing efficient analytical queries over LOD or SW data cube needs a deep understanding of SPARQL query language. This knowledge is difficult for the end-user. That is why Etcheverry et al. proposed a Cube Query

Language (CQL), a query language with a high-level abstraction to help end-user interacts over data cubes. The CQL is using QB4OLAP vocabulary representation and make improvement on efficient SPARQL query[18]. The project proposed in [8]–[18] have important support for sharing, publishing statistical data or data cubes using LOD or open data concept. However, this implementation did not explore the possibilities to include spatiotemporal measure or spatial dimension in their DW. Exploring the spatiotemporal DW makes some new opportunities to execute complex queries with impressive and intuitive results with geographical visualization and its hierarchies.

2.3 Open spatiotemporal data warehouse

In the last decade, research on spatiotemporal or Mobility DW increases with the advancement of GPS, sensors, IoT, and mobile technology. In the descriptive, diagnostic, predictive, and prescriptive analysis, there is a need for impressive visualization with geographical and spatiotemporal dimensions. Many researchers proposed their concepts on an open spatiotemporal data warehouse.

Mijovic et al. proposed Exploratory Spatio-Temporal Analysis – Linked Data (ESTA-LD) tool to publish, distribute, and share spatiotemporal Open Government Data for location and spatial analytics and could help decision-maker for planning, regulating, and make a prediction. This tool has inspiring reporting functions, such as visualizing data cubes with spatial dimensions, reporting a number of measures, generate statistical maps, several measures comparison by time and regions. Multiple data sources aggregates and integrates using some information integration mashups. The study also discusses the representation model for spatiotemporal data cubes in the LOD format. This research gives the contribution of open-source tools for linked open data cubes analytics with a SPARQL endpoint easy access[19].

On the other hand, the majority of DW with the spatiotemporal feature is referred to as trajectory DW. Trajectory DW manages and evaluates trajectory data collected from moving objects (MOs) and can be exploited using the OLAP and data mining techniques[20]. The proposal of Wagner et al. (2018) offered Mob-Warehouse as a Trajectory DW with semantic enrichment. This study proposed the movement unit described as a spatiotemporal point with semantics, including patterns, activity segments, hierarchies, categories, and transportation means[21]. Vaisman et al. (2019) built Mobility DW for mobility analytics. The DW store the movement of objects as a prolonged series of spatiotemporal coordinates and

the positions of moving objects at any point in time in (x, y, t) format. The mobility DW stores MOs data that may have mobility measures in the fact table and mobility dimensions. The instances of the mobility measures are trajectory and route. The examples mobility dimension are road network, a region of districts, a region of villages, and etc [22]. Zhao et al. (2019) proposed a data cube model and integrated data model and organization methods for remote sensing data and location-based social sensing data based on spatiotemporal data cube. This research implements the spatiotemporal data cube model on an array DBMS, and show a disaster analysis scenario for integrated analysis of multi-source spatiotemporal big data [23].

The Government of Denmark has joined the LOD movement with the development of GovAgriBus in 2015. The GovAgriBus published data from many domains to be accessible by a cross-domain query and available on the SW. The GovAgriBus has agriculture, business, and government services domain with corresponding locations attributes that convey the opportunities to spatiotemporal analytics on the SW [24]. Representation of LOD for cube could be represented with the QB4SOLAP vocabulary. The QB4SOLAP is an extension of QB4OLAP with the support of spatial query functions and operators. The operators also include spatial aggregation with spatial hierarchies[25]. The research on a spatial data warehouse with LOD has significant progress with the QB4SOLAP vocabulary findings [26], but this model is still limited only for the spatial domain. On the other hand, the spatiotemporal DW has excellent advancement with Mobility DW implementation in MobilityDB extension on PostGreSQL DBMS[22]. This Mobility DW has a different approach that becomes the main reference for this research. This research explored the mobility DW features from the MobilityDB to

support the decision making in agriculture production analysis.

Table 1 describes the comparison of the spatiotemporal DW implementation.

The mobility database is data storage for moving or spatiotemporal objects and manages the data for query retrieval. Spatiotemporal objects data have spatial and temporal attributes. Spatial attributes describe real-world phenomena consisting of descriptive components, which are presented with traditional data types such as integers, boolean, strings, and dates. Spatial attributes are also described by spatial extent components, which explain the real world in the form of geometric data types such as Point, Points, Line, Polygon, and Surface (MADS Model). Temporal attributes describe time, changing in time. Time types and temporal types could represent spatiotemporal attributes. Vaisman et al. (2019) define temporal types as moving types and time types as data types that represent the time, like timestamp, timestamp with time zone, period, timestamp set, and period set [22]. Temporal types are based on time types. The example of temporal types is a temporal integer, temporal float, temporal points, and temporal geometries type.

Temporal types are functions that map time instants to values from a domain. They are temporal(integer) is a continuous function

$$f : instant \rightarrow integer$$

The temporal integer type could represent the evolution of employee salary, the continuous production of farmer groups, the evolution of employee quantity of several departments, and so on. Temporal types also have operators that are associated with this type. This temporal types represent data in a natural and simple way. Figure. 2 shows the example of temporal integer types of farmer groups agricultural production per half month

Table 1. Comparison of spatiotemporal data warehouse implementation

| Paper	Open Data Index	Dimension Data	DW domain	Type of measures
[4]	3-stars	non-spatial	e-goverment service	base data types
[5]	3-stars	spatial, non-spatial	agriculture	base data types
[6]	3-stars	spatial, non-spatial	multi domain e-goverment	base data types
[7]	4-stars	spatial, non-spatial	development planning	base data types
[8]- [10]	5-stars	spatial, non-spatial	multi domain e-government	base data types
[11]	5-stars	spatial, non-spatial	e-goverment service	base data types
[12]	5-stars	spatial, non-spatial	environment	base data types
[17]	5-stars	spatial, non-spatial	health service	base data types
[19]	5-stars	temporal, spatial, non-spatial	multi domain e-goverment	base data types
[20], [21]	5-stars	temporal, spatial, non-spatial	multi domain e-goverment	segments of trajectories
[23]	4-stars	temporal, spatial, non-spatial	disaster analysis	base data types
[24],[25]	5-stars	spatial, non-spatial	multi domain e-goverment	base data types
This paper	5-stars	temporal, spatial, non-spatial	agriculture production	spatiotemporal data types
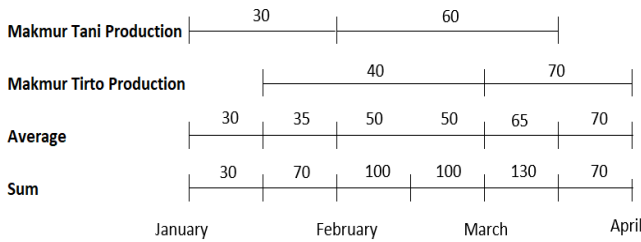
Figure. 2 The example of temporal aggregate operator

and aggregation operations of two agriculture production (average and sum).

3. Research methodology

This research is categorized as **constructive research**. This constructive research focused on the development of methods, models, modules, tools, and approaches that have applicable further than the realistic study that drove its creation. Figure. 3 shows the fishbone diagram of the research activities. This study is comprising of five activities. The detail of the activities is:

Activity 1: Strategic Planning for Villages and Rural Area Information System (SIDeKa) that develops with several steps, including (1) Literature Review (2) Data Collection that held by location observation, interview the village governmental stakeholders, survey, and business process analysis. The business processes analysis conducted by several tools, including Porter's Value Chain [27] and Business Process Analysis [28]. (3) Enterprise Architecture modeling, using phases in TOGAF ADM methodology[29]. This method is a complete and comprehensive approach for strategic information system planning [30]. Figure. 4 shows the Value Chain Diagram of the Village Services.

Activity 2: The Village Information System (SIDeKa) development for primary activity services such as public services for governance, asset management, and development planning is the second activity of the research. This system has been implemented in several pilot areas.

Activity 3: The third activity of this study is the development of SIDeKa extension for the agricultural module on the web and mobile platforms. This module has new functionality that could reduce the supply and demand chain between farmer and consumer and between farmer and supplier. The system also captures the crop production process data, including spatiotemporal attributes data. Figure. 5 depicts the screenshot of the agriculture module on SIDeKa.

Activity 4: SIDeKa's setup for data integration using API Web Service and Linked Open Data (LOD). This step is designed for providing data
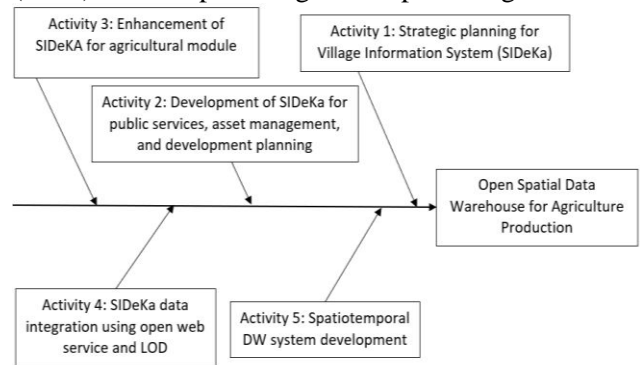

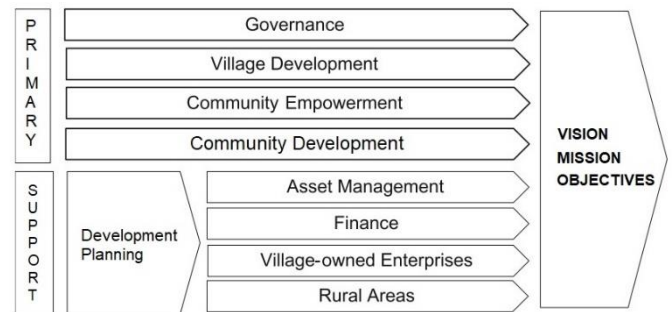Figure. 3 The fishbone diagram of research steps
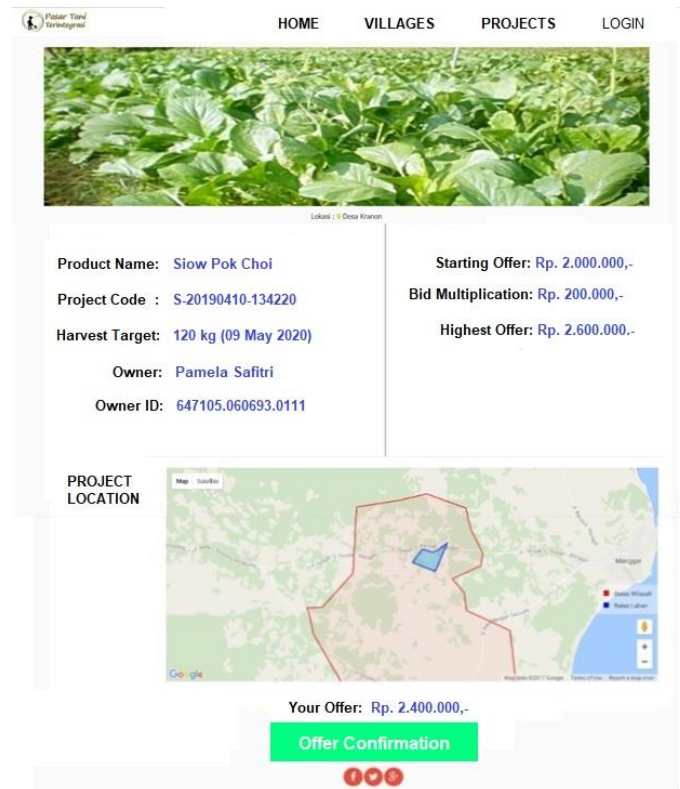

Figure. 4 The village services value chain

Figure. 5 Screenshots of agriculture module of SIDeKA

access for village data as a data source for Data Warehouse.

Activity 5: Development of Open Spatiotemporal Data Warehouse System with the following steps: (1) Identification of the user requirement, definition, refinement, and goal prioritization (2) Design a snowflake/star dimensional model that have fact tables, measure, and dimensions. The design process also introduces spatiotemporal measures and dimensions, (3) Define open integration model/architecture with Extract, Transform and Loading (ETL) process, and (4) Implementing the OLAP query for the DW.

4. Discussion and results

The Open Spatiotemporal Data Warehouse development is initialized with the requirement analysis. Requirement analysis is conducted with several steps:

(1) Identification of user group, identification of user requirement, definition, refinement, and goal prioritization. The user group that wants to use the DW system is:

(a) Executive: Agriculture Ministry, District head, Sub-district head

(b) Business Agency: National Logistics Agency

(c) Another stakeholder: Citizen, General Public

Goals: Strengthening the food availability and handling of food security.

The query that operationalizes the goals:

a.   Query#1: Find how much corn, soybeans, and rice production in the fasting month and Eid al-Fitr this year in the predetermined polygon of sub-districts? This query has spatiotemporal aggregation illustrated in Figure. 6

b.   Query#2: Find out how much average crop potential agricultural loss (currency) suffered by the Boyolali district because of the eruption of Mount Merapi on January 21, 2020. It is assumed that the affected area is at a radius of 5 km.

c.   Query#3: Determine how much credit prepared by the bank at per the quarter year 2019 in subdistrict 'Kalasan'?

d.   Query#4: Determine the nearest village that could fill the production gap of corn, soybeans, and rice of predetermined villages in the fasting month this year?

e.   Query#5: Compute the need for fertilizing and seeding material for corn, soybeans, and rice cultivation in Bantul district in the first quarter of the year 2020?

f.   Query#6: How much harvest product for each field of Karangtengah Village in June 2020?
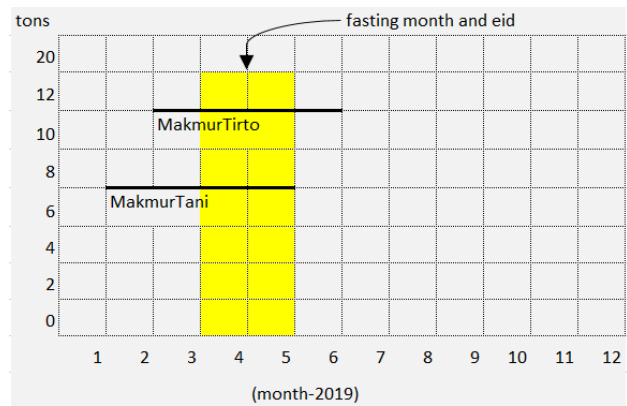

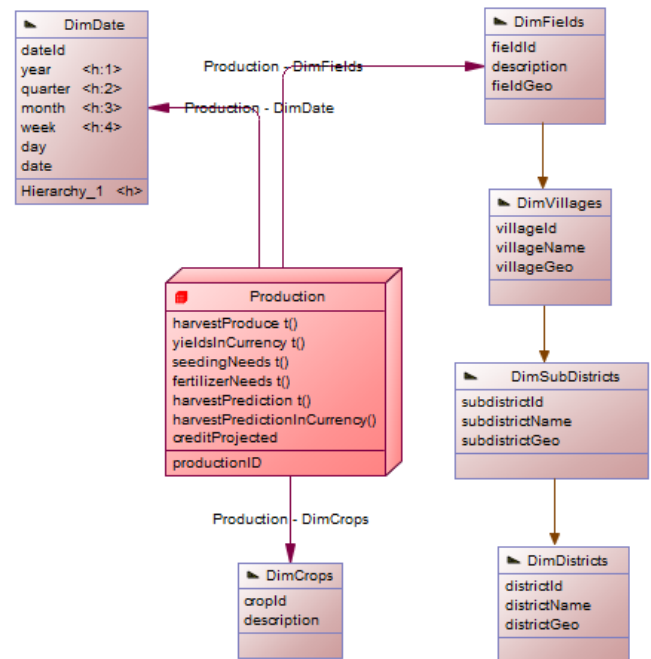Figure. 6 The soybeans production data example in 2019

Figure. 7 The agriprod data warehouse relational schema

(2) Designing a multidimensional model for the Spatiotemporal Data Warehouse, including facts, measure, and dimension. This model introduces new measures with temporal types and spatial dimensions that have a spatial hierarchy. The attribute with _Geo postfix is a spatial attribute that contained a set of location or geographics information. Figure. 7 depicts the relational representation of Agriculture Production DW.

(3) Implement Extract, Transform, and Loading (ETL). The ETL process is the way to integrate the various data source to load it into Enterprise DW. The SIDeKa's distributed and open operational data will be integrated into the enterprise DW. Integration employs REST API web services that run on the top of the SIDeKa server. The Web Service is built using PHP language with framework CodeIgniter 3.0.6. Each SIDeKa server has several different endpoints. Each endpoint will be responsible for one function,

Table 2. SIDeKa API endpoints

No	Endpoint	Source Table
1	../getAllMember	sideka.member
2	../getAllLelang	sideka.lelang
3	../getAllPenduduk	sideka.tbl_penduduk
4	../getAllTawaran	sideka.tawaran_lelang
5	../getAllProduct	sideka.produk_pertanian
6	../getAllProgress	sideka.progress_proyek
7	../getAllProyek	sideka.proyek_pertanian
8	../getAllTipeProduk	sideka.ref_tipe_produk

Table 3. Transactional data sample

Subject	Predicate	Object
B001	hasDistrict	Sleman
B001	inVillage	Purwomartani
B001	projectStart	2018-10-10
B001	projectEnd	2019-01-20
B001	plantedBy	Duryono
B001	crop	Duryono
B001	hasArea	2 (ha)
B001	hasGeometry	Polygon (109.2272791 -7.339392, 109.2274508 -7.3397644, …, 109.227279,-7.339392)
B001	landPreparation	2018-10-12
B001	plantingSeed	2018-10-15
B001	fertilization	2018-10-22
B001	harvesting1	2019-01-02
B001	harvesting1y	7.5
B001	harvesting2	2019-01-09
B001	harvesting2y	10.0
B001	harvesting3	2019-01-16
B001	harvesting3y	8.0

and each of these functions will return a value/values from a transactional database in SIDeKa using JSON formats.

Table 2 displays the list of API endpoints of the SIDeKa agriculture module.

The sample of data from the transactional system in triples format, as shown in Table 3.

The transactional data also represented in Linked Open Data (LOD) in RDF format. The example of the LOD representation of the sample data is:

```
@prefix ex: <http://example.com/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix gns: <https://dbpedia.org/resource/>  .

ex:B001
 ex:hasDistrict gns:Sleman_Regency;
 geo:hasGeometry "POLYGON((109.2272791 -7.339392,
109.2274508 -7.3397644.. ))"^^geo:wktLiteral .
ex:B001
 ex:landPreparation "2019-10-12"^^xsd:date;
 ex:plantingSeed "2019-10-15"^^xsd:date .
ex:B001
 ex:harvesting1 "2019-01-02"^^xsd:date;
 ex:harvesting1y "7.5" ^^xsd:float;
 ex:harvesting2 "2019-01-09"^^xsd:date;
```

```
 ex:harvesting2y "10.5" ^^xsd:float;
 ex:harvesting3 "2019-01-16"^^xsd:date;
 ex:harvesting2y "8.0" ^^xsd:float;
```

(4) Deploy the Open Spatiotemporal Data Warehouse. The architecture of the DW system is depicted in Figure. 8. The DW system deploys on PostgreSQL DBMS with PostGIS and MobilityDB extension. The query function of PostGIS extension is very useful for querying important information on diagnostic location analysis on agriculture production. Whereas, the mobilityDB extension can represent temporal and spatiotemporal data types and functions for spatiotemporal or Mobility analytics. The SQL Query for operationalizing the goal of this business intelligence listed below. The ST_ prefix functions are the PostGIS functions. The result of the query was visualized by PowerBI visualization tools. The data definition query for creation of the sample of the fact and dimension table is:

```
CREATE TABLE Production(
id SERIAL PRIMARY KEY,
cropId int,
startDate int,
endDate int,
fieldId int,
harvestProduct tfloat,
yieldsInCurrency tint,
seedingNeeds tfloat,
fertilizerNeeds tfloat,
harvestPrediction tfloat,
harvestPredictionInCurrency tint,
creditprojected float);

CREATE TABLE DimFields (
fieldId int PRIMARY KEY,
description char (15),
FieldGeo geography);
```

The tfloat and tint type is a temporal type with float and integer as a base type, respectively. The tfloat, tint instants in these tables declared as a category of instant set [31].
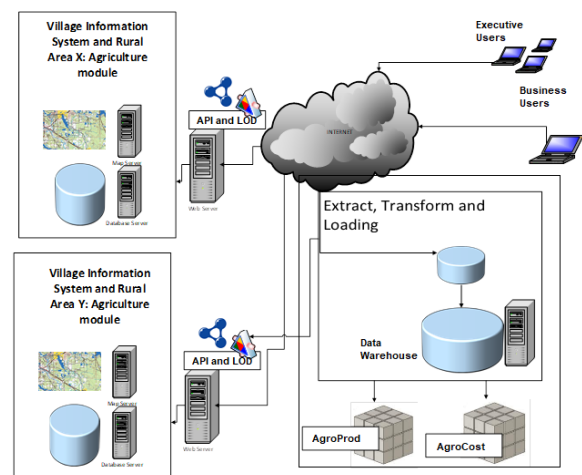

Figure. 8 Data warehouse system architecture

The prioritized queries of the business goals are described, and the visualized results are shown in respective figure.

Query#1: Find how much corn, soybeans and rice production in the fasting month and Eid al-Fitr this year in the predetermined polygon of sub-districts?

SQL Query (result shows in Figure. 9) :
```
SELECT dc.description, ds.subdistrictname,
(TSUM(ATPERIOD (p.harvestproduct,'[2019-05-
01 ,2020-06-01]' )))
FROM production p, dimfields df, dimcrops dc,
dimvillages dv, dimsubdistrict ds
WHERE p.fieldid = df.fieldid and p.cropid =
dc.cropid and df.villageid = dv.villageid
and dv.subdistrictid = ds.subdistrictid AND
p.description IN ('Rice','Corn','Soybeans')
ST_WITHIN( df.fieldgeo,'POLYGON((109.229998
-7.335044,   109.2303208   -7.3354481,...,
109.2299989 -7.3350438 ))')
GROUP BY dc.description, ds.subdistrictname
```

Query#2: Find out how much average crop potential agricultural loss (currency) suffered by the Boyolali district because of the eruption of Mount Merapi on January 21, 2020. It is assumed that the affected area is at a radius of 5 km.

SQL Query (result shows in Figure. 10):
```
SELECT                        d2.category,
TWAVG(TSUM(ATPERIOD( p2.harvestpredictionIn
Currency,'[2020-01-21, 2020-03-21]')))
FROM production p2 , dimfields d , dimcrops
d2, dimdistricts d3
WHERE p2.fieldid = d.fieldid AND p2.cropid
=   d2.cropid   AND   ST_DISTANCE('point
(110.445671 -7.540566) ', d.fieldgeo) < 5000
AND ST_WITHIN (d.fieldgeo, d3.districtgeo)
AND d3.districtName = 'Boyolali'
GROUP BY d2.category
```

Query#3: Determine how much credit prepared by the bank at per the quarter year 2019 in subdistrict 'Kalasan'?

SQL Query (result shows in Figure. 11):
```
SELECT d2.quartal, SUM (p.creditprojected)
FROM    production    p,    dimfields    d,
dimsubdistrict d3, dimdate d2
WHERE p.fieldid = d.fieldid and ST_WITHIN
(d.fieldgeo,    d3.subdistrictgeo)    AND
p.startdate = d2.dateid AND d2.year = '2019'
and d3.subdistrictName = 'Kalasan'
GROUP BY d2.quartal
```

Query#4: Determine the nearest village that could fill the production gap of corn, soybeans, and rice of predetermined villages in the fasting month this year?
SQL Query (result shows in Figure. 12) :
```
WITH  XY  as  (SELECT  d3.villageid  as
id,min(TWAVG(ATPERIOD   (p.harvestproduct,
```

```
'[2019-01-02,2020-03-10)')))   as  MinX  FROM
production  p,  dimcrops  d,  dimfields  d2,
dimvillages d3 WHERE p.cropid = d.cropid AND
```
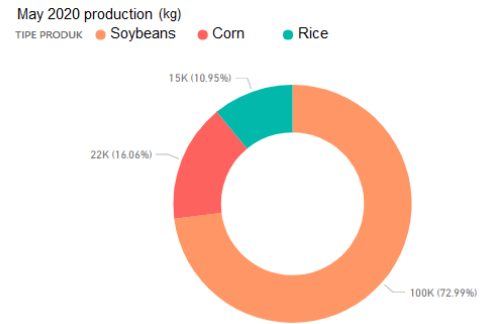

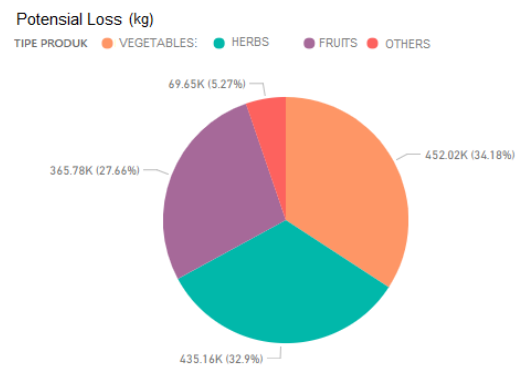Figure. 9 Query#1 result

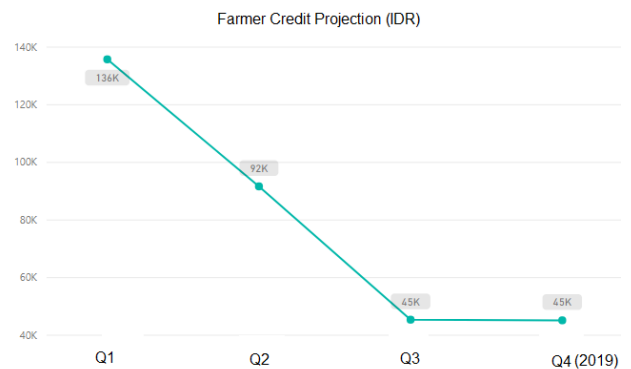Figure. 10 Query#2 result

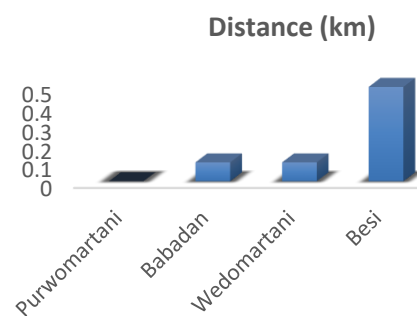Figure. 11 Query#3 result

Figure. 12 Query#4 result

```
p.fieldid = d2.fieldid AND d2.villageid =
d2.villageid GROUP BY d3.villageid)
SELECT       d2.villagename,      ST_DISTANCE
( d2.villagegeo,loc.geom)/1000 as dist_km
FROM dimvillages d2 INNER JOIN
(SELECT df.villageid as id,df.villagegeo as
geom FROM dimvillages df WHERE df.villageid
IN (SELECT XY.id FROM XY WHERE XY.minx =
(SELECT  MIN(Minx)  from  XY)))  as  loc  ON
ST_DWITHIN  (d2.villagegeo  ,  loc.geom,
10000)and d2.villageid <> loc.id
ORDER  BY  ST_DISTANCE  (d2.villagegeo  ,
loc.geom)
```

Query#5: Compute the need for fertilizing and seeding material for corn, soybeans, and rice cultivation in Bantul district in the first quarter of the year 2020?

SQL Query (result shows in Figure. 13):
```
SELECT dc.description, TWAVG(TSUM(ATPERIOD
(p.fertilizerneeds,  '[2020-01-06,2020-06-
06]'))),TWAVG(TSUM
(ATPERIOD(p.seeding,'[2020-01-06  ,2020-03-
06]')))
FROM production p, dimfields df, dimcrops dc,
dimdistrict d
WHERE p.fieldid = df.fieldid AND p.cropid =
dc.cropid   AND   ST_WITHIN   (df.fieldgeo,
d.districtgeo)  AND  d.districtname      =
'Bantul'  AND  dc.description  IN  ('Rice',
'Corn','Soybeans' )GROUP BY dc.description
```
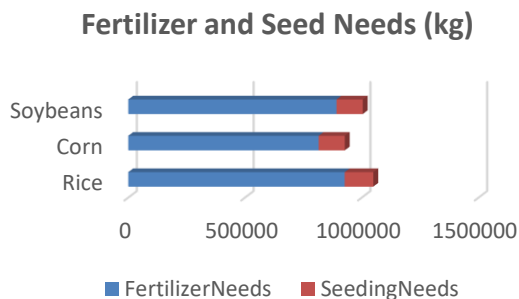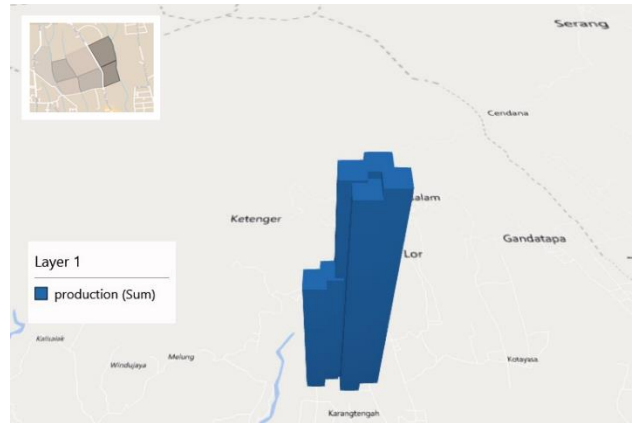

Figure. 13 Query#5 result

Figure. 14 Query#6 result

Query6: How much harvest product for each field of Karangtengah Village in June 2020?

SQL Query (result shows in Figure. 14):
```
SELECT        d.fieldgeo,        TWAVG(TSUM
(ATPERIOD(p2.harvestproduct,'[2020-06-
01,2020-06-30]')))FROM   production   p2   ,
dimfields d , dimcrops d2, dimvillages d3
WHERE p2.fieldid = d.fieldid AND p2.cropid
= d2.cropid AND d3.villageid = d.villageid
AND  d2.description  IN  ('Rice','Corn',
'Soybeans')     AND     d3.villagename    =
'Karangtengah' GROUP BY d.fieldgeo
```

5. Discussion on performance issues

The focus of the paper is the implementation of a spatiotemporal data warehouse with open integration with data sources using service-oriented architecture and linked open data. The data warehouse is deployed on mobilityDB, a PostgreSQL extension with rich spatiotemporal features [22] and PostGIS [32]. This implementation is the new method of data warehousing using temporal types measures and some spatial dimensions.

Since the paper is focusing on proposed the new concept of data warehousing and emphasizing the expressiveness of the spatiotemporal query on MobilityDB. Consequently, a detailed evaluation of

Table 4. Queries execution time

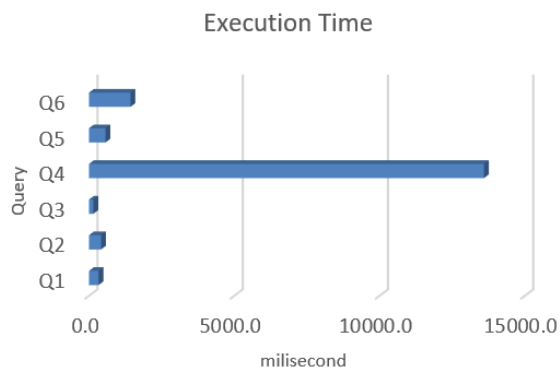| Query	Execution Time (ms)										
	1	2	3	4	5	6	7	8	9	10	Avg
Q1	327	337	329	330	331	331	335	311	333	333	329.7
Q2	433	426	421	429	414	399	406	418	406	405	415.7
Q3	157	154	147	154	149	142	154	161	148	142	150.8
Q4	14355	12947	14100	12891	13467	13898	13153	13634	13821	13682	13594.8
Q5	560	571	584	551	571	563	568	568	564	575	567.5
Q6	1129	1157	1124	1960	1161	1117	1660	1121	1870	1980	1427.9

Figure. 15 Query execution time

query performance is outside the scope of this work. But for showing the expressiveness of in this last section, this paper summarized the execution times of the six spatiotemporal queries for operationalizing the goal of the DW.

The DW fact table contains 324096 tuples with temporal integer/float for each tuple, 4495 tuples of field dimension with geographic data as polygons, 80 tuples of village dimension, dozens of tuples of the district, subdistrict, province dimensions. The DW time dimension contains 3653 tuples representing a date for ten years.

The six queries in section 4 were executed on Pentium Core i5 8th generation with a 12GB RAM computer. The results are not designed to be conclusive but only suggest the expressiveness of the query from the MobilityDB DW. The results of the execution test time of the six queries are shown in Table 4 and Figure. 15).

6.   Conclusion and future works

The implementation of this Open Spatiotemporal Data Warehouse for production analytics has been developed with a new concept of mobility data warehouse using the MobilityDB, an extension of the Postgresql. This system could answer the critical questions used for agriculture production analysis using mobility query for a roll-up, drill-down, slice, dice, spatial aggregation function, and temporal aggregation function with an impressive execution time. The query result from the query is beneficial for descriptive and diagnostic agriculture production analysis. The openness of the data source with the 5-stars index make an opportunity to link and enlarge the data source with the big data catalog from the internet and inference capability of the semantic web.

The experiments proved that the open spatiotemporal DW implementation has a convenient, simple model, and expressive performance (mostly less than 1s) for supporting the executive to make an

analysis based on spatiotemporal data, then make the best decision.

In the future, the challenge is how to extend all artifacts of data warehouse and multidimensional cube in the format of the LOD Semantic Web. Representation of DW in existing cube vocabulary, integration with another LOD will enhance the capability of the system with the bigger catalog and have standardization of terms and concepts of the application domain [33]. Using this LOD as a data source, the data warehouse could be enriched with the new semantic and mechanism to provide further knowledge to allow reasoning via logical axioms in the entire process of business intelligence and big data warehousing.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, writing—original draft preparation, writing—review and editing, Irya Wisnubhadra; review and supervision, Safiza Suhana Kamal Baharin and Nanna Suryana Herman;

Acknowledgments

References

[1]   S. Murti, "Pengaruh Resolusi Spasial pada Citra Penginderaan Jauh Terhadap Ketelitian Pemetaan Penggunaan Lahan Pertanian", *Geomatika*, Vol. 18, No. 1, pp. 84–94, 2012.

[2]   Deloitte, "eTransform Africa: Agriculture Sector Study Sector Assessment and Opportunities for ICT", 2012. Accessed: 2020. [Online]. Available: http://siteresources.worldbank.org/.

[3]   A. Vaisman and E. Zimányi, *Data Warehouse Systems: Design and Implementation*, 1st ed. Berlin: Springer Berlin Heidelberg, 2014.

[4]   X. Liu and X. Luo, "a Data Warehouse Solution for E-Government", *Int. J. Recent Res. Appl. Stud.*, Vol. 4, No. 1, pp. 101–105, 2010.

[5]   M. Maliappis and D. Kremmydas, "An online analytical processing (OLAP) database for agricultural policy data: A Greek case study", *CEUR Workshop Proc.*, Vol. 1498, No. September, pp. 214–225, 2015.

[6]   R. K. Arora and M. K. Gupta, "e-Governance using Data Warehousing and Data Mining", *Int.*

J. Comput. Appl., Vol. 169, No. 8, pp. 28–31, 2017, doi: 10.5120/ijca2017914785.

[7] I. Wisnubhadra and S. P. Adithama, "Design and build OLAP business intelligence for village sustainable development planning", in *IFIP Advances in Information and Communication Technology*, Vol. 504, pp. 559–569, 2017, doi: 10.1007/978-3-319-59111-7_46.

[8] P. E. R. Salas, M. Martin, F. M. Da Mota, S. Auer, K. Breitman, and M. A. Casanova, "Publishing statistical data on the web", In: *Proc. of IEEE 6th International Conf. on Semantic Computing, ICSC 2012*, pp. 285–292, 2012, doi: 10.1109/ICSC.2012.16.

[9] Y. Asano, S. Koide, M. Iwayama, F. Kato, I. Kobayashi, T. Mima, I. Ohmukai, and H. Takeda, "Constructing a Site for Publishing Open Data of the Ministry of Economy, Trade, and Industry: — A Practice for 5-Star Open Data —", *New Gener. Comput.*, Vol. 34, No. 4, pp. 341–366, 2016, doi: 10.1007/s00354-016-0403-y.

[10] J. Matsuda, A. Mizutani, Y. Asano, D. Yamamoto, H. Takeda, I. Ohmukai, F. Kato, S. Koide, H. Harada, and S. Nishimura, "Publication of Statistical Linked Open Data in Japan", In: *Proc. of Joint International Semantic Technology Conference*, Vol. 11341, pp. 307–319, 2018, doi: 10.1109/mdat.2014.2345111.

[11] A. Stasiewicz, M.A. Rezk, A. Ojo, E. Tambouris, E. Kalampokis, K. Tarabanis, T. Alcorn, and A. Leadbetter, "Using Linked Statistical Data to Improve Marine Search and Rescue Operations in Ireland", In: *Proc. of International Conf. on Theory and Practice of Electronic Governance*, pp. 412–418, 2018, doi: 10.1145/3209415.3209511.

[12] D. Gavin, T. Dhu, S. Sagar, N. Mueller, B. Dunn, A. Lewis, L. Lymburner, S. Minchin, S. Oliver, J. Ross, and M. Thankappan, "Digital Earth Australia - from Satellite Data to Better Decisions", in *International Geoscience and Remote Sensing Symposium*, pp. 8633–8635, 2018, doi: 10.1109/igarss.2018.8518160.

[13] J. Klímek, J. Kučera, M. Nečaský, and D. Chlapek, "Publication and usage of official Czech pension statistics Linked Open Data", *J. Web Semant.*, Vol. 48, pp. 1–21, 2018, doi: 10.1016/j.websem.2017.09.002.

[14] F. Cotton, R. Cyganiak, R. T. A. M. Grim, D. W. Gillman, Y. Jaques, and W. Thomas, "Xkos extension for representing statistical classifications", in *The 59th ISI World Statistics Congress 2014*, pp. 306–3011, 2014.

[15] L. Etcheverry and A. A. Vaisman, "Enhancing OLAP Analysis with Web Cubes", In: *Proc. of*

the 9th Extended Semantic Web Conf., The Semantic Web: Research and Applications, pp. 469–483, 2012, doi: 10.1007/978-3-642-10871-6.

[16] L. Etcheverry and A. A. Vaisman, "QB4OLAP : A New Vocabulary for OLAP Cubes on the Semantic Web", In: *Proc. of COLD'12 the Third International Conf. on Consuming Linked Data*, pp. 247–38, 2012.

[17] M. Bouza, B. Elliot, L. Etcheverry, and A. A. Vaisman, "Publishing and querying government multidimensional data using QB4OLAP", In: *Proc. of 9th Lat. Am. Web Congr. LA-WEB 2014*, pp. 82–90, 2014, doi: 10.1109/LAWeb.2014.11.

[18] L. Etcheverry and A. A. Vaisman, "Efficient Analytical Queries on Semantic Web Data Cubes", *J. Data Semant.*, Vol. 6, pp. 199–219, 2017, doi: 10.1007/s13740-017-0082-y.

[19] V. Mijović, V. Janev, D. Paunović, and S. Vraneš, "Exploratory spatio-temporal analysis of linked statistical data", *J. Web Semant.*, Vol. 41, pp. 1–8, 2016, doi: 10.1016/j.websem.2016.10.002.

[20] A. Vaisman, "Data Warehouses: Next Challenges", in *eBISS: European Business Intelligence Summer School*, pp. 1–26, 2012.

[21] R. Wagner, J. A. F. de Macedo, A. Raffaetà, C. Renso, A. Roncato, and R. Trasarti, "Mob-Warehouse: A Semantic Approach for Mobility Analysis with a Trajectory Data Warehouse," In: *Proc. of International Conf. on Conceptual Modeling*, pp. 127–136, 2018, doi: 10.1007/978-3-319-14139-8_15.

[22] A. Vaisman and E. Zimányi, "Mobility Data Warehouses", *ISPRS Int. J. Geo-Information*, Vol. 8(4), No. 170, pp. 1–22, 2019, doi: 10.3390/ijgi8040170.

[23] J. Zhao and P. Yue, "Spatiotemporal Data Cube Modeling for Integrated Analysis of Multi-Source Sensing Data", in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4791–4794, 2019, doi: 10.1109/IGARSS.2019.8900630.

[24] A. B. Andersen, N. Gür, K. Hose, K. A. Jakobsen, and T. B. Pedersen, "Publishing Danish agricultural government data as semantic web data", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 8943, No. Ld, pp. 178–186, 2015, doi: 10.1007/978-3-319-15615-6_13.

[25] N. Gur, K. Hose, T. B. Pedersen, and E. Zimányi, "Modeling and Querying Spatial Data Warehouse on the Semantic Web", In: *Proc. of 5th Joint International Conf., JIST 2015*, pp. 3–22, 2015, doi: 10.1007/978-3-319-31676-5_1.

[26] N. Gur, T. B. Pedersen, E. Zimányi, and K. Hose, "A foundation for spatial data warehouses on the Semantic Web", *Semant. Web*, Vol. 9, No. 5, pp. 557–587, 2018, doi: 10.3233/SW-170281.

[27] M. E. Porter, *Competitive Advantage Creating and Sustaining Superior Performance*. The Free Press, New York, 1985.

[28] F. N. Al-Aboud, "Strategic information systems planning: a brief review", *IJCSNS Int. J. Comput. Sci. Netw.*, Vol. 11, No. 5, pp. 179–183, 2011.

[29] T. O. Group, "Introduction to ADM", *TOGAF 8.1.1. Online*, 2006. .

[30] I. A. Alonso, J. C. Verdun, and E. T. Caro, "The IT implicated within the enterprise architecture model: analysis of architecture models and focus IT architecture domain", In: *Proc. of IEEE International Conf. on Service-Oriented Computing & Application (SOCA)*, pp. 165–170, 2010, doi: 10.1109/SOCA.2010.5707174.

[31] E. Zimányi, "MobilityDB 1.0-beta Manual", *https://docs.mobilitydb.com/MobilityDB/master/*, 2020. https://docs.mobilitydb.com/MobilityDB/master/ (accessed Jul. 04, 2020).

[32] R. O. Obe and L. S. Hsu, *PostGIS In action*. Greenwich: Manning Publication Co., 2011.

[33] I. Wisnubhadra, S. S. K. Baharin, and N. S. Herman, "Modeling and querying spatiotemporal multidimensional data on semantic web: A survey", *J. Theor. Appl. Inf. Technol.*, Vol. 97, No. 23, pp. 3608–3633, 2019.