A Sustainable Multi-layered Open Data Processing Model for Agriculture: IoT Based Case Study Using Semantic Web for Hazelnut Fields

Şahin Aydin*, Mehmet Nafiz Aydin

Kadir Has University, Department of Management Systems, 34083, Turkey

A R T I C L E   I N F O

A B S T R A C T

In recent years, several projects which are supported by information and communications technologies (ICT) have been developed in the agricultural domain to promote more precise agricultural activities. These projects account for different kinds of key ICT terms such as internet of things (IoT), wireless sensors networks (WSN), cloud computing (CC). These projects are used for different agricultural products; and it is a well-known fact that they can be essential to perform precise agricultural activities for the relevant agricultural products. The implementation of these projects successfully depends on the extent to which various stakeholders provide support by leveraging relevant data, gathered from heterogenous data sources. Agriculture domain has a great number of stakeholders. These stakeholders need sophisticated data and appropriate intelligence to get benefits in order to perform precise agricultural activities. Authors agreed with scholars that "Open Data" idea, which means accessing data published on the web and available in a machine-readable format is an appropriate way to get benefits for precise agriculture by relevant stakeholders. In this paper, authors shall investigate the open data term in an agricultural context, create an open data processing model, and develop an IoT-based solution to gather environmental data from agricultural fields. Authors also show viability of the proposed model by developing an ICT-based solution. Considering the socioeconomic importance of hazelnut for Turkey, the stakeholders of hazelnut domain still have problems such as availability, meaningful, accuracy of the hazelnut related data. Therefore, authors shall focus on hazelnut within the scope of this paper.

1. Introduction

This paper is an extension of work originally presented in the 6th International Conference on Control Engineering & Information Technology (CEIT), in 2018 [1].

Agriculture increases employment that covers more than third of the world population via exportation and domestic income which makes it one of the leading sectors in the world. [2]. Turkey which is the leader country in terms of hazelnut production in the world, has an important role for contributing more information with respect to hazelnut production cycle from this aspect. As well as Turkey is the leader of hazelnut production, it is leading producer of dried apricots, raisins, and dried figs [3]. Since having such a convenient weather quality which only a few countries could have for hazelnut production, Turkey, has 75% of overall

production and 70-75% of the exportation. Furthermore, considering that there have been 4 million people correspond with almost 5% of the overall population of Turkey, who are concerned with hazelnut produced over an area of 550-600 thousand hectares, this makes hazelnut quite important in terms of socio-economic aspect [4]. In Turkey, statistics concerning agricultural products, including hazelnut are presented as only hypertext markup language (HTML) pages, spreadsheets, and comma-separated files (CSV) by Turkish Statistical Institute (TUIK). The context of hazelnut data presented by TUIK contains just a certain type of statistical information. However, the domain stakeholders such as farmers, experts, researchers, and analysts might generate more detailed and sophisticated data regarding hazelnut so that gaining knowledge from it must be straightforward and obvious. In order to achieve that, the data must include basic parameters used for the general management of the accession and environmental and field-specific parameters. This kind of data, freely available to use and

---
*Şahin AYDIN, Email: aydinoglu.sahin@gmail.com

republish and in machine-readable format, might give a chance to stakeholders for removing the obstacles of implementation of sustainable hazelnut production. Examples of issues regarding hazelnut, which might be commonly encountered by domain stakeholders, might be indicated as follows: limited information or uncertainty about supply and demand side. In addition, there are several problems regarding hazelnut production such as aging hazelnut gardens, land element and position, soil texture, blights, fertilizing, treatment, less pollination, less pruning, agricultural pests, price fluctuation, increase of hazelnut production areas uncoordinatedly, and lack of persistent exportation policies [5]. Furthermore, there exist positive or negative correlations between environmental factors and hazelnut production. For instance, a sudden drop in the temperature negatively affects hazelnut production. It is not generally possible to access this type of data for such a particular agricultural product. However, it might be a solution to leverage the reference body of knowledge and analyze the reports of scientific experiments by accessing this data. The data obtained from scientific experiments are not usually allowed for using by any stakeholders out of the relevant communities [6]. However, considering the presence of barriers for accessing such kind of data, most of organizations belonging to government or private sector are in charge of removing them and making the data freely accessible. Nevertheless, aforementioned permission barriers should be removed for hazelnut stakeholders to meet requirement accessing, reusing, and publishing sophisticated data regarding hazelnut as well.

In this paper, we shall look in detail into processing open data gathered from different data sources and a designed model, considering very exclusive layers from stakeholders' perspectives.

This paper begins with an introduction section in which the research problem is explained. Second section is the research background section that mentions some key terms regarding the model's layers. Then, open data processing model is disclosed in the model, followed by the findings section. Lastly, a case study, which is developed as an exclusive system to demonstrate the model viability, was explained in the implementation section.

2. Research Background

In Turkey, hazelnut production is being dominantly performed in the zone of Black Sea which comprises of two different areas named as First Standard District (FSD) and Second Standard District (SSD) respectively. While SSD is composed of the west and the middle areas of Black Sea zone, FSD is located in the east of the Black Sea zone. Although there are similarities between two areas, it should be noted that considerably differences such as environmental characteristics and climate conditions are exist between them as well. The aforementioned dissimilarities might generate an obvious discrepancy in hazelnut production year by year. When one looks at the statistical data of last thirty years in agricultural production, one sees that on average 520.198 tons hazelnut were produced in Turkey, and such data demonstrates that the position of Turkey as a leader in hazelnut production in the world. Turkey exports significant amount of hazelnut every year. The statistical data regarding agricultural exportation in the year 2019 illustrate that Turkey exported at about 320 thousand tons of hazelnut and it received 2 billion 28 million dollars of revenue via this. It has been identified that there are not adequate researches and studies about how to utilize contemporary ICT for such valuable agricultural product. Therefore, a multi-layered model

based on open standards for demonstrating how to process agricultural data is proposed considering the hazelnut which is a dominant agricultural product produced in a wide range area in Turkey in this study. Multi-layered model enables us to distinguish the essential layers such as data sources, data processing, semantic annotation, data storage, services, applications and users functions physically and logically. In addition, it enables interoperability for each layer to accomplish performing sustainable agricultural activities. There are a number of important benefits of the multi-layered approach. For instance, the multi-layered approach is secure, and its management is easy. It is also scalable, flexible and provides more efficient development abilities. Adding new features to the developed system is easy as well. That is why, a multi-layered model is presented and detailed within this study.

Although extensive research has been carried out on agricultural ICT technologies, not many studies exist which adequately focus on hazelnut. Furthermore, there is not a model which meets the requirements of hazelnut agricultural activities as well. So, a multi-layered open data processing model which aims to conduct sustainable agricultural activities of hazelnut is proposed as a part of this study.

The model proposed is essentially based on two pillars: the very idea of "Open Data", and "Semantic Web" approach to data modeling. Authors shall first provide a basic understanding of these pillars so that conceptual foundation of the proposed model is established. Publishing the data in machine-readable format using favorable open standards and open formats by enabling stakeholders to access, read, reuse, and share it with no license limit or permission barriers might be described as "Open Data" [7]. Furthermore, ICT's role is to facilitate distribution of data to users from various stakeholders in the relevant domain. Eliminating heterogeneity of data sources in the agricultural domain might be possible by using the strength of semantic web technologies. It is necessary to publish data using open standards and in machine-readable format to meet requirements of providing processable data by software agents. Semantic Web (SW) aims transforming data into suitable format which might be processed by software agents and human via software tools [8]. As well as SW eases retrieving the data by human or software agents via making interrelations between the objects. The linked data, which might be usable in varied formats on the web and published under open license, must be in structured and machine-readable format [9]. Resource Description Framework (RDF) is used to display information on the web as guided and identified data. It usually serves data including, but not limited to, personal information, social networks, metadata about digital artifacts. Furthermore, it contributes as a way of combination over diverse sources of information. The structure of SPARQL which is being used for querying RDF data, is described semantically considering this specification [10].

Utilizing semantic web technologies is quite efficient way to present domain-specific data on the web. It might ensure removing difficulties in terms of presenting data in appropriate formats, which can be processed by both software agents and humans. RDF, which is widely used and known as common standard for sharing data semantically on the web, has strong characteristics to ease merging data obtained from heterogeneous sources. Merging data obtained from varied sources might be performed using ontologies. RDF, which plays a significant role in SW's world in terms of

presenting data, is main pillar in terms of generating ontology regarding varied domains [11]. RDF also enables improving schemas in time with no requirement for changing by data consumers [12].

3. Model and Findings

As mentioned in the previous sections, authors created a multi-layered open data processing model, and focused on the hazelnut agricultural product. Even though this model focuses on a sustainable agricultural production lifecycle of hazelnut in this study, it is convenient for all agricultural products as well. This model is created by considering data sources, data processing, semantic annotations, data storage, services, applications, and users. The proposed open data model which demonstrates how to process data obtained from heterogeneous agricultural data sources and designed as multi-layered includes varied layers such as types of data sources (IoT devices, statistical data from government, market data, farmers, and other data sources); storing and processing raw data; semantic annotation layer (agricultural

product ontology, other ontologies and data interchange engine); data storage layer (graph databases, relational databases, cloud databases, RDF and XML file-based storage options and database service); services layer (XML web services, REST web services, web APIs, mobile services, analysis services, reporting services and SPARQL query services); applications layer (unified web based data platform based on open formats, mobile applications and desktop applications); and lastly, end users layer. The overall model and its each layer, proposed in this study, are being demonstrated in Figure 1.

The data are generally gathered from heterogeneous sources such as farmers, sensors, government statistical data, and market data in the agriculture domain. There is no doubt that the most important one of these heterogeneous data sources is "farmers". As a significant data source in terms of the model, farmers, play a vital role for performing agricultural production life cycle. Considering the hazelnut production cycle, there is much varied information regarding hazelnut might be collected by consulting of farmers.
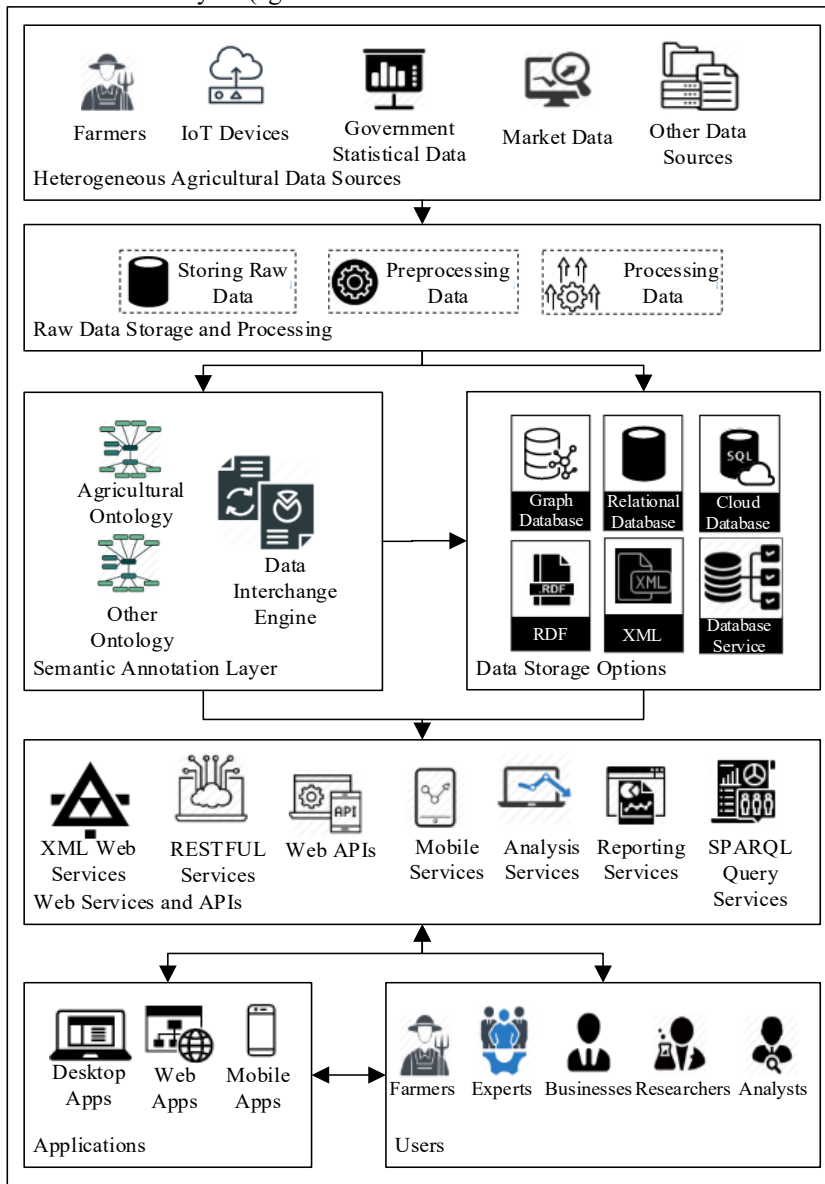

Figure 1: Overall Architecture of Open Data Model

For instance, age productivity and morbidity of trees, observation for soil fertility, the number of trees in orchards, location of orchards (southern slopes or northern slopes), pruning hazelnut trees, fertilization, agricultural spraying, cropping system, propagation method, irrigation method, overall vegetation surrounding the collecting site, stoniness, rockiness etc.

Today, IoT devices have been frequently used for varied domains such as smart cities, industries, education, health, and agriculture. These devices generally consist of a wide range of sensors which are used to collect specific data pertaining to environmental characteristics of site. Collecting and processing data with respect to these characteristics facilitates performing precise agricultural activities. Sensors might provide sensitive measurement values concerning slope, weather temperature, velocity of wind, rainfall, soil moisture, soil Ph, frost (date of most recent frost, minimum temperature, duration of temperature below 0°C), relative humidity (diurnal and seasonal range), light intensity, leaf anatomy, etc.

Agricultural production stats, which are a sort of official statistics data, are recorded by government agencies. The open data model takes into consideration the importance of statistical data concerning agricultural production so, it adopts them as an element of overall architecture. The statistical records regarding many varied domains in Turkey are compiled, evaluated, and published by TUIK [13]. The data with respect to agriculture domain are generally published by TUIK as spreadsheet, comma-separated values, and HTML files, which might be accessed through TUIK web platform.

Marketing operations in agriculture domain include several procedures such as production plan, burgeoning, reaping, classifying, packaging, shipping, managing storehouse, transforming agricultural products into food, delivery, promote, and sale, required for movement of a product from farmers to buyers [14]. It is a very crucial cycle for any agricultural product, so the proposed model contains it as a data source element. The experts, researchers, and analysts who are interested in relevant agricultural products might contribute to the system by loading some exclusive data. So, these kinds of data sources are defined as other data sources in the open data model.

According to NIST (National Institute of Standards and Technology), "cloud computing (CC) is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage, applications, and services, that can be rapidly provisioned and released with minimal management effort or service provider interaction" [15]. Considering the strength and robust characteristics of CC, it has been recommended that data storages, services and applications of the proposed model should serve on the cloud. Raw Data Storage and Processing Layer (RDSandPL) comprises component of raw data storing (SRD) , component of data preprocessing (PPD), and component of data processing (PD). Data obtained from heterogeneous agricultural data sources are stored into varied data storage options such as relational databases, graph databases, cloud databases, RDF files, and XML files via SRD. It is worth bearing in mind that any stakeholder might need to handle data obtained from heterogeneous sources in raw form. That's why, it is significant storing data in raw format using recommended data storage options according to open data model point of view.

Considering the data gathered from heterogeneous agricultural data sources might be incomplete, incorrect, or irrelevant, it is important to emphasize the requirement of a mechanism to detect and correct them. According to open data model, this mechanism is being corresponded with PPD component. Preprocessing the data ensures removing incompleteness and inconsistency with respect to the raw data and transforming such data into meaningful format to analyze effectively [16]. As well as, PPD decomposes raw data, it aims to apply exclusive techniques concerning preprocessing to such data. Applying preprocessing techniques such as data integration, data enhancements, data enrichment, data transformation, data reduction, data discretization and data cleansing boosts quality of the data [17]. Some of the traditional data preprocessing methods such as K-neighbors Classifier, Logistic Regression, Gaussian Naïve Bayes, Decision Tree, and Support Vector Machine might be used for sensor data, but it should be noted that the major purpose of development of these methods are not to apply them to sensor data [18]. This component is also responsible for semantic annotation of streaming data obtained from sensors, using existing or new ontologies. By this means, semantic annotation of streaming sensor data facilitates data integration processes between varied data sources and open data platforms. In addition, this component solves noisy and redundant problems of sensor data streams, solves missing value problems, detects anomalies related to data, and manages data heterogeneity using the strength of semantic web technologies [19].

The aim of carrying out the preprocessing techniques and methods to the sensor stream data is to build a robust technical infrastructure for data processing component of the proposed model. Data processing means that acquiring meaningful information from preprocessed raw data. Acquiring meaningful information enables users to make estimations for the future via implementing varied analyze techniques and allows creating analytical solutions. So, it is important to bear in mind that having such a component is essential in terms of open data model. The design principles proposed by Begoli and Horey, which allows using varied analytical techniques and substantially interested in maximizing the controllable factors, provide easily exploring and analyzing the data by researchers [20]. The practical analytical solutions are ensured corresponding with these design principles for stakeholders by component of data processing of open data model.

When implementation of relevant methods and techniques is completed by the components of RDSandPL, it is required transforming data into RDF datasets. The data interchange component of Semantic Annotation Layer (SAL) is in charge of transforming data into RDF datasets. The proposed model recommends using agricultural product ontologies to transform data into RDF datasets.

Ontologies distinctly specify conceptualization by means of constructing abstract and simplified view of a particular domain which requires to be represented [21]. Combining a number of objects and concepts, which are related to knowledge that is supposed to be expressed and making interrelations between them might be described as conceptualization [22]. The aim of generating ontologies is not merely to identify the conceptualization or point out the widespread savvy of a particular domain. In addition, they are utilized as conceptual models which provide common understanding to facilitate communicating

between varied types of software application systems and among people [23]. Nowadays, there exist several ontologies which are still being in use, concerning the agriculture domain. AGROVOC, one of the commonly known agricultural ontologies, was developed by Food and Agriculture Organization (FAO) of the United Nations in the early of the 1980s. Furthermore, it is the most popular and well-known agricultural thesaurus all over the world. AGROVOC, which is formed as linked dataset, is based on SKOS-XL (Simple Knowledge Organization System eXtension for Labels) concept scheme, and it comprised of more than thirty-five concepts which are defined in up to twenty-nine different languages. [24]. Institute for Learning & Research Technology created SKOS(Simple Knowledge Organization System) with the aim of providing linked and distributed knowledge organization systems such as controlled vocabularies, thesauri, taxonomies, and folksonomies [25]. SKOS-XL, as an extension version of SKOS, facilitates identifying lexical entities and making interrelationships among them, with the additional support [26]. The focus of this study is on the development of open data processing model. Therefore, authors intend to develop a comprehensive ontology using descriptors of hazelnut as future work [27]. Both RDF (Resource Description Framework) and OWL (Web Ontology Language) have vital role in the world of semantic web and are popular technologies which might be utilized to accomplish the aforementioned objective. RDF and OWL are the most known standards of semantic web technologies for describing data on the web. OWL is developed based on RDF and extends it including exclusive components such as axioms, classes, object properties, data properties, datatypes, individuals etc. [28]. OWL is the common, well-known, and standard language to create ontologies. Moreover, it has three sublanguages to meet different requirements such as OWL Full, OWL DL, and OWL Lite.

Data gathered from different data sources and processed by RDSandPL might be stored in graph databases, relational databases, cloud databases, XML files, and RDF files, which are recommended as data storage options in the open data model. The proposed open data model is noted the importance of converting raw data into RDF datasets. Therefore, transforming data gathered from varied sources into triple store which means subject, predicate, and object, has been built as a component of semantic annotation layer. Due to RDF datasets have robust features in terms of storage and indexing [29], they have been considered as a storage option in the proposed open data model. Relational databases component of data storage layer (DSL) is in charge of storing relational data. Graph databases component of DSL is in charge of storing data as graphs. Cloud databases might store data as both graphs and relational. Database services, which is a component of DSL, provide a secure connection, session management, user mapping, user authentication, and authorization between components of services layer (SL).

Another essential layer of proposed model is the services layer, which consists of seven different types of services: XML web services, REST web services, web APIs, mobile services, analysis services, reporting services, and SPARQL query services. According to W3Schools, "web services are defined as self-contained and self-describing application components that can be discovered using UDDI, be used by other applications, and they communicate using open protocols" [30]. Web Services and APIs Layer (WSandAL), which is in charge of providing interoperability between mobile, desktop, and embedded software applications and databases, is a significant part of the proposed

model. While providing interoperability by WSandAL, the RESTful Services component of this layer plays an important role. Software agents, tools and applications, which are being developed using different or similar development technologies, utilize agricultural data by consuming RESTful Services. Another important feature of the proposed model is mobility. To meet this requirement, it strongly recommends developing mobile applications. It is, therefore, needed to generate mobile services to develop an efficient infrastructure for mobile devices. Mobile services, which provide connection to different kinds of mobile devices must take part in the proposed model. Stakeholders might be limited by lack of knowledge to apply data analysis methods to agricultural data. The proposed model recommends building analysis services to solve this problem.

Designing and publishing detailed reports are required to develop usable software systems which aims to process data. Reporting Services component of open data model aims to meet these requirements. It enables domain stakeholders create dynamic, well-designed and detailed reports. As mentioned in previous section of this study SPARQL is the language to query RDF datasets which are stored in files or graph databases. In the proposed open data model, creating queries using this particular language and compiling them are provided and facilitated by service component named SPARQL Query Service.

Applications Layer (AL) allows end users to access agricultural data obtained from heterogeneous data sources. AL consists of web, mobile and desktop applications which enable us to import, export, report, analyze, and query data.

There are a number of varied end users belonging to proposed open data model, called domain stakeholders such as farmers, researchers, businesses, analysts, and domain experts. Domain stakeholders are provided to utilize data which are gathered from heterogeneous agricultural data sources and might be consumed through the WSandAL and processed by applications layer. Considering the importance of accessing, reusing, and processing such data, it allows domain stakeholders of hazelnut agricultural product construct common knowledge with respect to performing agricultural activities in an efficient way. Every year, several exploratory investigations are performed to estimate total amount of hazelnut production by different institutions of Turkey. However, these current exploratory investigations to find out the hazelnut production of a further year are limited because of unforeseen circumstances with respect to hazelnut harvest. Unfortunately, some discrepancies are seen among the estimations which are carried out by different government agencies every year. Due to the hazelnut price per kilogram is being affected from these discrepancies negatively, rapid price fluctuations are being seen for hazelnut in Turkey. When one takes into consideration the problems for hazelnut market, each user in other words domain stakeholder, play a vital role in terms of the proposed open data model.

The proposed open data model reveals that how each of stakeholders might exploit the processed and published agricultural data for removing the obstacles and problems concerning any particular agricultural product. The proposed open data model by its data processing perspective might enable farmers who have a key role in agriculture domain plan their production cycle effectively. The domain experts and researchers might take preventive measures to prevent further spread of the agricultural

diseases by producing the meaningful information using the applications and services developed based on the proposed open data model. Varied kinds of statistical analysis with respect to market, investment, and employment might be undertaken by analysts utilizing the conferred data by proposed open data model. Statistical analysis presented by analysts might be used to support making effective decisions concerning market share of agricultural businesses for the future.

4. Implementation

The proposed multi-layered open data processing model is introduced within the previous sections. In this part of the study, a prototype of the electronic system is developed by establishing a wireless sensor network (WSN), which consists of varied sensors and other electronic devices in order to test the viability of the proposed model. This implementation is only focused on streaming data from IoT devices to the databases through web services. Other aspects of the proposed model shall be elaborated in different studies in future studies. There are three sensor nodes named router 1, router 2, and router 3 in the developed system, respectively. These nodes transmit measured data to the coordinator. This system structure is an implementation of ordinary WSN. Figure 2 presents the architecture of the developed wireless sensor network.

Weather temperature, weather humidity, weather pressure, carbon monoxide, nitrogen, oxygen, and ultraviolet detection sensors have been plugged on router 1; digital light intensity, soil moisture, and precipitation detection sensors have been plugged on router 2; and accelerometer and gyroscope sensor have been plugged on router 3. Table 1 illustrates which sensor measures what type of measurements, and which sensor is plugged on which router. While some of these sensors measure different types of environmental events or changes at the same time, some of them measure only one event or change. Weather pressure, temperature, humidity changes have been detected by only one sensor named BME280. Likewise, rates of carbon monoxide, hydrocarbons and oxidizing gases such as NO_2 have been measured by only one sensor named MICS-4514. Arduino Uno microcontroller boards based on ATmega328P have been used to plug the sensors and

measure their values in the system. XBee is one of the most popular modules which allows devices connect to each other through wireless. So, XBee Pro S2C modules have been seen an appropriate embedded solution for established wireless sensor network to transmit data from routers to the coordinator. The coordinator consists of a Raspberry Pi 3 small single-board computer, an Arduino Uno, and an XBee Pro S2C module. Raspberry Pi has been included in WSN to run a universal windows application, which is in charge of reading, visualizing, and storing real-time data stream obtained from sensors. This application communicates with the coordinator device to read the data stream through serial port. The real-time data stream, which is visualized in favorable visualization objects by universal application, are stored in databases through web services. Figure 3 illustrates how sensor measurements are visualized by universal windows application on Raspberry Pi.

ZigBee is an international open standard and communication protocol, which depends on IEEE 802.15.4 and ensures low-power, low-cost, low-data-rate, and wireless mesh networking. [31]. Zigbee has three kinds of node types, such as coordinator, router, and end device. However, the developed WSN does not include any end device. Each of the XBee modules belonging to WSN were configured to API operating mode, which provides data transmission to multiple destinations with no requirement, using command mode. Receiving success and failure status of each transmitted RF packet is facilitated in this mode. In addition, the operation of identifying the source address of each received packet is readily provided by this mode. The routers transmit data, which is gathered via sensors, in an exclusive format named frame, represented as a ZigBee packet to the coordinator. The structure of the frame data; in other words, a ZigBee packet, which is used by the routers to transmit sensor data stream towards the coordinator, is demonstrated in Figure 4. The structure of frame data differs in compliance with the objectives of the API frame. The frame data structure of the established wireless sensor network, represented in Figure 4, consists of the fields, such as start delimiter, length, frame type, frame id, 64-bit destination address, options, RF data, and checksum.
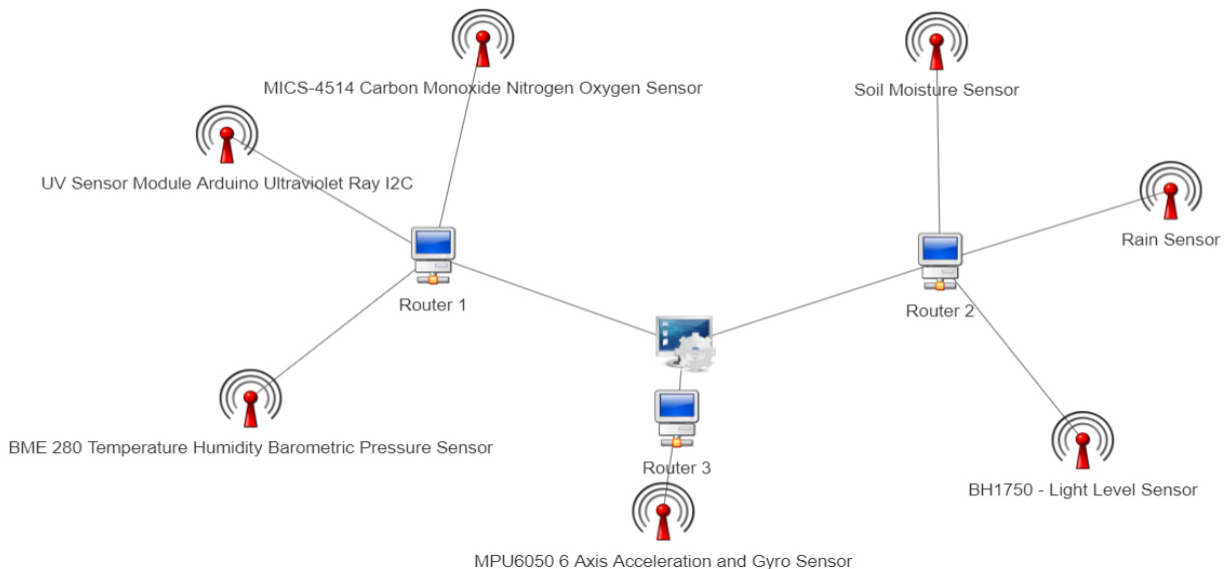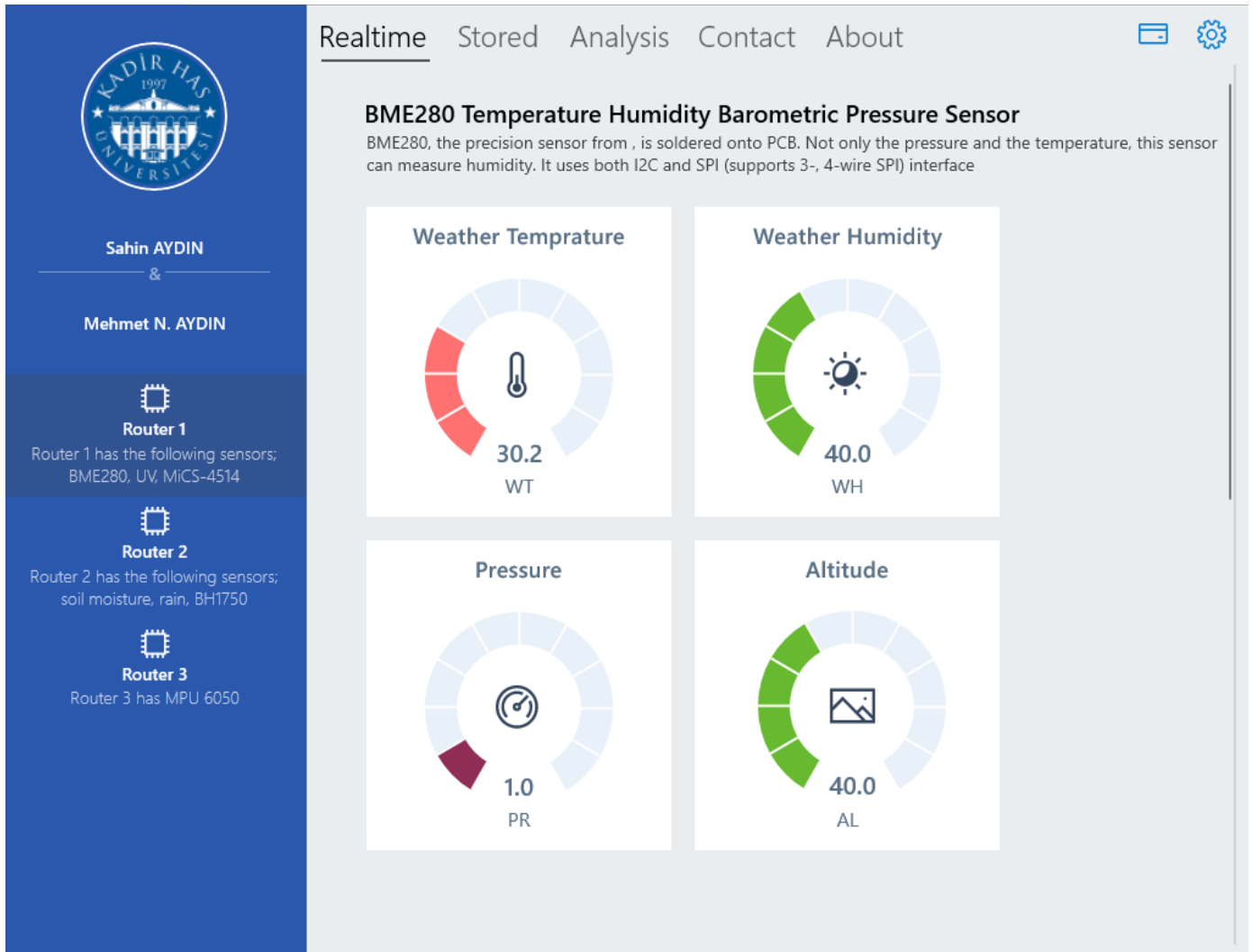

Figure 2: Developed WSN Architecture

Table 1: Sensors plugged on routers

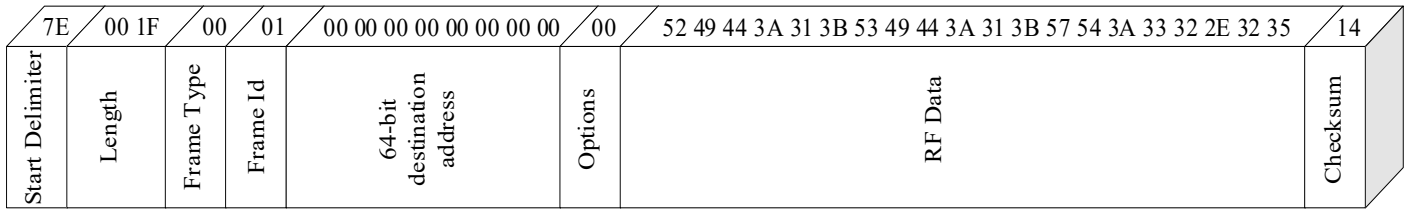| Router Name	Sensor Name	Measurements
Router 1	BME280 Temperature Humidity Barometric Pressure Sensor	Temperature
		Humidity
		Pressure
	GUVA-S12SD UV Sensor Module Arduino Ultraviolet Ray I2C	UV radiation in sunlight
	MICS-4514 Sensor	CO concentration
		Hydrocarbons
		Oxidizing gases
Router 2	Soil Moisture Sensor	Soil Moisture
	BH1750 - Light Level Sensor	Light Intensity
	Rain Sensor	Rain detection and rainfall intensity measurement
Router 3	MPU6050 (Gyroscope + Accelerometer + Temperature) Sensor Module	Accelerometer Gyroscope Temperature

7E	00 1F	00	01	00 00 00 00 00 00 00 00	00	52 49 44 3A 31 3B 53 49 44 3A 31 3B 57 54 3A 33 32 2E 32 35	14
Start Delimiter	Length	Frame Type	Frame Id	64-bit destination address	Options	RF Data	Checksum

Figure 4: Frame Data Format

The sensor data are transmitted from the routers to the coordinator using key-value pair identifiers. For instance, the frame data, represented in Figure 4, refers to "RID:1;SID:1;WT:32.25". The key-value pairs of this frame are RID:1, SID:1, and WT:32.25, respectively. Likewise, the values are 1, 1, and 32.25 as well. The keys, RID, SID, and WT stand for router id, sensor id, and temperature, respectively.

When the data are transmitted to the coordinator and transformed into a meaningful format by universal application, storage operations are performed by DPL. Data are stored in two different types of database management systems, such as Microsoft MS SQL Server and Neo4j graph database system, which are a component of data storage layer. Neo4j, which is a graph database management system, increases the system performance, provides flexibility and agility. Figure 6 presents the data stored in Neo4j. RDF file-based data storage option is used to store and export sensor data for this case study. Table 3, Table 4, Table 5 represent RDF/XML, Notation 3, and N-Triples datasets, respectively. These datasets can be created using the sensor listing UI of the open data platform represented in Figure 5.

The data gathered from sensors and stored on databases must be published in a suitable format for applications running on different platforms. Therefore, REST web services were developed using Windows Communication Foundation (WCF), which is a framework for building service-oriented applications. Furthermore, web services for Windows platforms were developed in the same way. Table 2 illustrates the formatted data in JSON (JavaScript Object Notation) format. Developed web services are accessible to stakeholders.

Table 2: Example of JSON data

```
[{
    "Id": 2,
    "SensorDataAddedDatetime": "\/Date(1580635128430+0300)\/",
    "SensorDataType": 2,
    "SensorDataValue": "32.45",
    "SensorId": 1000
}]
```

Table 3: Example of RDF/XML dataset

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE rdf:RDF [
        <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
        <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
        <!ENTITY xsd 'http://www.w3.org/2001/XMLSchema#'>
        <!ENTITY sensorData 'http://www.opendatainagriculture.com/sensors#'>
]>
<rdf:RDF xml:base="http://www.opendatainagriculture.com/sensors" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:sensorData="http://www.opendatainagriculture.com/sensors#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature">
    <sensorData:DataTypeId>2</sensorData:DataTypeId>
    <sensorData:Datetime>2/2/2020 10:04:07 PM</sensorData:Datetime>
    <sensorData:SensorTypeId>1000</sensorData:SensorTypeId>
    <sensorData:SensorTypeName>BME 280 Temperature Humidity Barometric Pressure Sensor</sensorData:SensorTypeName>
    <sensorData:TypeDef>WT</sensorData:TypeDef>
    <sensorData:TypeName>Weather Temprature</sensorData:TypeName>
    <sensorData:Value>32.434</sensorData:Value>
  </rdf:Description>
</rdf:RDF>
```

Table 4:Example of Notation 3 dataset

```
@base <http://www.opendatainagriculture.com/sensors>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix sensorData: <http://www.opendatainagriculture.com/sensors#>.

<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature> sensorData:DataTypeId "2";
                                    sensorData:Datetime "2/2/2020 10:04:07 PM";
                                    sensorData:SensorTypeId "1000";
                                    sensorData:SensorTypeName "BME 280 Temperature Humidity Barometric Pressure Sensor";
                                    sensorData:TypeDef "WT";
                                    sensorData:TypeName "Weather Temprature";
                                    sensorData:Value "32.434".
```

Table 5: Example of N-Triples dataset

```
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#DataTypeId> "2".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#SensorTypeId> "1000".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#SensorTypeName> "BME 280 Temperature Humidity Barometric Pressure
Sensor".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#TypeName> "Weather Temperature".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#TypeDef> "WT".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#Value> "32.434".
<http://www.opendatainagriculture.com/sensors/datatype#WeatherTemperature>
<http://www.opendatainagriculture.com/sensors#Datetime> "2/2/2020 10:04:07 PM".
```
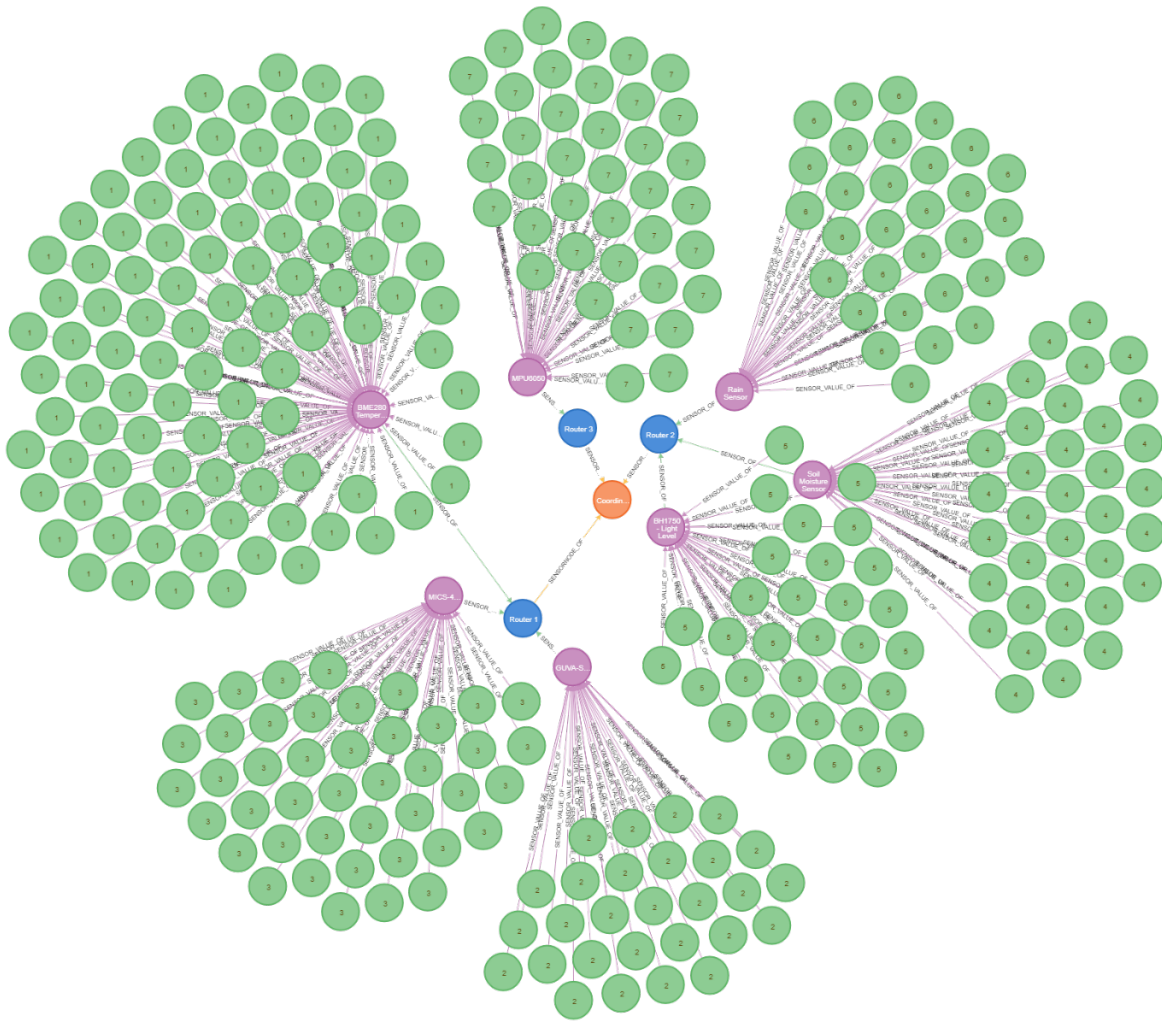
Figure 6: Storing sensor data in Neo4j Graph Database

Figure 7: SPARQL query service

Created RDF datasets might be manipulated using SPARQL query language through SPARQL query service, which is represented in Figure 7. This service filters files in compliance with the ".RDF" extension and creates buttons for each of them. Stakeholders should click to load the relevant dataset on SPARQL query service engine. After the loading process is completed, it should be written to execute the query into the text area.

5. Conclusion and Future Work

In this paper, authors proposed a multi-layered open data processing system that consists of different layers and components. All layers and its components were explained; and an implementation regarding sensors was developed and summarized in the implementation section. Authors examined only one aspect

in this implementation. However, there are four other aspects in terms of heterogeneous data sources. These other aspects are worth discussing and examining in the future. Also, authors tried to justify the suitability of the data life cycle in terms of the proposed multi-layered model by implementing the wireless sensor network system. Authors feel that this study enhances academic understanding of creating a multi-layered open data processing model. However, the life cycle of other data from other data sources should be examined and justified for viability. In addition, authors mentioned creating hazelnut ontology, but did not its development and description. As a future work, an ontology for hazelnut shall be created and adapted to the model. In addition, authors are planning to develop an open data platform based on the proposed model. Stakeholders will be able to create data acquisition forms using agricultural trait dictionaries to gather structured data regarding the relevant agricultural product without requiring strong technical skills. The platform will allow obtaining environmental data from heterogenous sensors and will be able to map these sensor measurements with relevant classes of agricultural trait dictionaries. By this means, domain-specific data will be able to be published in varied open formats using the developed platform. A number of restrictions of this study and areas for future research should be mentioned; and the following ones might be given as examples: difficulty of gathering data from farmers (lack of knowledge of technology usage), defining the boundaries of gathering market data, determining the potential data sources, which is defined as "*Other Data Sources*" in heterogeneous data sources layer, and spreading the usage of applications mentioned in "*Applications Layer*".

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

End notes

Web-based Open Data Platform can be accessed using the following URL: http://www.opendatainagriculture.com/

References

[1]. Ş. Aydın, U. Ünal and M. N. Aydın, "Open Data in Agriculture: Sustainable Model Development for Hazelnut farms using semantics," 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 2018, pp. 1-6.doi: 10.1109/CEIT.2018.8751875

[2]. Melissa J. Perry, Agricultural Health and Safety, Editor(s): Stella R. Quah, International Encyclopedia of Public Health (Second Edition), Academic Press, 2017, Pages 38-44, ISBN 9780128037089

[3]. Agriculture and Food. [online] Available at: http://www.invest.gov.tr/en-US/sectors/Pages/Agriculture.aspx [Accessed 21.07.2018].

[4]. Turkey's Hazelnut. [online] Available at: http://www.ftg.org.tr/en/turkish-hazelnutturkeys- hazelnut.html [Accessed 21.07.2018].

[5]. Fındık Veriminde ve Üretiminde Temel Sorunlar. [online] Available at: http://www.findikihracati.com/findik/findik-veriminde-ve-uretiminde-temel-sorunlar.html [Accessed 21.07.2018].

[6]. Peter Murray-Rust, Open Data in Science, Serials Review, 34:1, 52-64, 2008, https://doi.org/10.1080/00987913.2008.10765152.

[7]. Carolan, L., Smithy, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J., Gerry, W., 2015, How can we improve agriculture, food and nutrition with open data? , Open Data Institute.

[8]. I. Szilagyi and P. Wira, "Ontologies and Semantic Web for the Internet of Things - a survey," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, 2016, pp. 6949-6954.https://doi.org/10.1109/IECON.2016.7793744.

[9]. Berners-Lee, T., 2006, Linked Data, https://www.w3.org/DesignIssues/LinkedData.html [Accessed 23.07.2018].

[10]. SPARQL 1.1 Query Language. [online] Available at: https://www.w3.org/TR/2013/REC-sparql11-query-20130321/ [Accessed 23.07.2018].

[11]. Lata, Swaran & Sinha, Bhaskar & Kumar, Ela & Chandra, Somnath & Arora, Raghu, Semantic Web Query on e-Governance Data and Designing Ontology for Agriculture Domain. International Journal of Web & Semantic Technology. 4., 2013 http://doi.org/10.5121/ijwest.2013.4307.

[12]. Resource Description Framework. [online] Available at: https://www.w3.org/RDF/ [Accessed 23.07.2018]

[13]. Available at: http://www.turkstat.gov.tr/UstMenu.do?metod=gorevYetki [Accessed 24.07.2018].

[14]. Agricultural Marketing. [online] Available at: https://en.wikipedia.org/wiki/Agricultural_marketing [Accessed 24.07.2018]

[15]. Mell, P., G., Tim, 2018, National Institute of Standards and Technology, Information Technology Laboratory – a nonregulatory federal agency within the U.S. Department of Commerce – http://csrc.nist.gov/groups/SNS/cloud-computing.

[16]. Aisha Siddiqa, Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Mohsen Marjani, Shahabuddin Shamshirband, Abdullah Gani, Fariza Nasaruddin, A survey of big data management: Taxonomy and state-of-the-art, Journal of Network and Computer Applications, Volume 71, 2016, Pages 151-166, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2016.04.008.

[17]. I. Taleb, R. Dssouli and M. A. Serhani, "Big Data Pre-processing: A Quality Framework," 2015 IEEE International Congress on Big Data, New York, NY, 2015, pp. 191-198. https://doi.org/10.1109/BigDataCongress.2015.35.

[18]. Zhong Y, Fong S, Hu S, Wong R, Lin W. A Novel Sensor Data Pre-Processing Methodology for the Internet of Things Using Anomaly Detection and Transfer-By-Subspace-Similarity Transformation. Sensors (Basel). 2019;19(20):4536. Published 2019 Oct 18. https://doi.org/10.3390/s19204536.

[19]. Jason J. Jung, Semantic preprocessing for mining sensor streams from heterogeneous environments, Expert Systems with Applications, Volume 38, Issue 5, 2011, Pages 6107-6111, ISSN 0957-4174 https://doi.org/10.1016/j.eswa.2010.11.017.

[20]. E. Begoli and J. Horey, "Design Principles for Effective Knowledge Discovery from Big Data," 2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture, Helsinki, 2012, pp. 215-218, https://doi.org/10.1109/WICSA-ECSA.212.32.

[21]. Thomas R. Gruber, Toward principles for the design of ontologies used for knowledge sharing?, International Journal of Human-Computer Studies, Volume 43, Issues 5–6, 1995, Pages 907-928, ISSN 1071-5819, https://doi.org/10.1006/ijhc.1995.1081.

[22]. N. F. Noy and C. D. Hafner, "The State of the Art in Ontology Design: A Survey and Comparative Review", AIMag, vol. 18, no. 3, p. 53, Sep. 1997, https://doi.org/10.1609/aimag.v18i3.1306 .

[23]. Cristani, Matteo & Cuel, Roberta, A Survey on Ontology Creation Methodologies. Int. J. Semantic Web Inf. Syst.. 1.,2005, 49-69. 10.4018/jswis.2005040103.

[24]. LNCS About AGROVOC Page. [online] Available at: http://aims.fao.org/standards/agrovoc/concept-scheme, last [Accessed 2019/01/15].

[25]. Dean Allemang, Jim Hendler, Chapter 10 - SKOS—managing vocabularies with RDFS-Plus, Editor(s): Dean Allemang, Jim Hendler, Semantic Web for the Working Ontologist (Second Edition), Morgan Kaufmann, 2011, Pages 207-219, ISBN 9780123859655, https://doi.org/10.1016/B978-0-12-385965-5.10010-X.

[26]. LNCS SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL). [online] Available at: https://www.w3.org/TR/skos-reference/skos-xl.html, [Accessed 2019/01/15].

[27]. Koksal, A.,I., Gunes, N., Descriptors for Hazelnut, Bioversity International and FAO, 2008.

[28]. Pollock, J., T.: Semantic Web for Dummies. 1st edn. Wiley Publishing, Inc., Canada ,2009, ISBN: 978-0-470-39679-7 .

[29]. Luo Y., Picalausa F., Fletcher G.H.L., Hidders J., Vansummeren S., Storing and Indexing Massive RDF Datasets. In: De Virgilio R., Guerra F., Velegrakis Y. (eds) Semantic Search over the Web. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-25008-8_2.

[30]. XML Web Services. [online] Available at: https://www.w3schools.com/xml/xml_services.asp [Accessed 25.07.2018].

[31]. DIGI International Inc., XBee/XBee-PRO S2C Zigbee RF Module User Guide, 2018