The model of data interoperability in farm management information system

Henriyadi

Ph.D. student in Information System at Asian Institute of Technology, Thailand
stll8501@ait.ac.th

Abstract. The crop production process is not a single activity but is a series of activities, start for pre-planting, planting, harvesting, until marketing the product. Moreover, this process involves many resources also, both internal and external resources. To make all resources manageable effectively and efficiently, utilizing a Farm Management Information System (FMIS) is one of the tools that could be considered. Existing FMIS applications generally characterized as a closed application, focus only on few functionalities, utilize close data, and segment for the big scale farm due to some reasons. To make the FMIS application applicable to smallholder farmers, developing an FMIS application based on open source that utilizes open data is a necessity. However, utilizing open data in a data interoperability framework is not trivial, complicated, and deal with many heterogeneity problems. The four main heterogeneity problems that might arise are schema heterogeneity, the granularity of data, mismatch in entity naming and data unit, and inconsistency of data. Required a model of FMIS data interoperability which includes the algorithm to identify and tackle all those problems comprehensively. Functional requirement analysis will be employed for developing the model. To check the quality and validity of the proposed model, a series of the evaluation process will be performed. As proof of the model, this research will develop a mobile-based FMIS application. Finally, usability testing and impact analysis will be conducted to assess the benefits of using FMIS applications to the farm.

Keywords: Farm Management Information System, FMIS, semantic heterogeneity, data interoperability, conceptual model, algorithm.

1 Introduction

Agricultural sectors have a very important role in the economic growth of a country, especially in developing countries. The World Bank reported that the agricultural sector occupied 37.1% of the world’s surface in 2013 [1]. Additionally, the World Bank also reported which 70% of poor people in the world who occupants in the countryside are very dependent on agriculture sectors [2]. Regarding the number of workers in agriculture sectors, despite data from the International Labor Organization (ILO) shows declining trend percentage from 45% in 1991 to 31% in 2013 [3], ILO estimate 1.1 billion people depend on agriculture for their livelihoods.
However, the development of the agricultural sectors itself faces numerous issues, both internally and externally. A number of internal issues include (1) small size of agricultural land ownership, where more than 80% out of 570 million farmers in the world own farm smaller than 2 ha, and more than 87% are family farms [4], (2) availability productive labor, where working outside agricultural sectors is more appealing [5], and (3) limited arable land for plantation, where they should compete with non-agricultural sectors [5]. Whereas the external issues include (1) increasing the frequency of natural disasters, (2) biodiversity loss, (3) instability in food prices, and (4) the long supply chain of production [6].

Besides that, crop production, as a core activity in agricultural sectors itself, is not a single entity or single process, but a series of activities. Adapted from IRRI [7], and Ustriyana [8], generally, there are four groups of activities in crop production, namely (1) preparation for planting, (2) crop planting, (3) harvest and post-harvest handling, and (4) marketing the product. Moreover, crop production involves a lot of resources, both natural and artificial resources from internal and external as well. By seeing the complexities of the crop production process, indeed the farm cannot manage traditionally. Needed farm management to make all the crop production processes as well as utilization resources running effectively and efficiently. Farm management is a farm decision-making process through planning and controlling the allocation of limited resources of land, labor, and capital to achieve the farm objectives effectively and efficiently [6].

2 Rationale

Aligned with the significant growth of Information and Communication Technology (ICT) that have brought profound changes in human life, utilizing ICT for farm management is a necessity [5]. Farm Management Information System (FMIS) is a common term used to indicate utilizing ICT in farm management. FMIS is a system for the gathering, handling, putting away and disseminating data as information expected to convey [9]. Nowadays, there are a lot of FMIS applications already developed, both commercial and open source platforms. However, existing commercial FMIS application and academic research on FMIS mostly limits their objectives just to some of the individual components of FMIS [10]. The parts of the FMIS that appear to be popular in research are related to decision support, computational biological models, and estimates of yields [11]. Limitation focus of the FMIS research scope happens because the owner that utilize FMIS usually a big-farm that occupied big farm area and the functionalities provided solely focus on some features that satisfied the user necessities [12].

FMIS is a complex application and required supporting data both internal and external data sources. To handle many external data sources, an FMIS application required the ability to perform a function called data interoperability. Data interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged [13], [14].
3 Problem Statement

Utilizing a data interoperability framework in the development of FMIS application is not trivial, complicated, and deal with many problems [15]. The four main potential problems related to the use of open data through data interoperability framework are (1) heterogeneities of the schema, (2) granularity of data, (3) mismatch entity naming and data unit, and (4) inconsistency of data [16]. Heterogeneities of the schema are associated with the different schema between reference schemas and data source schema. Meanwhile, the granularity of data is related to how depth data representation for a particular entity. Moreover, mismatch of entities naming and data unit relates to the difference in defining entity naming or data unit for the same thing. Whereas inconsistency of data means that data that comes from several external data sources provide a different value for the same thing. The problem statement for this research is: How to tackle all of the four main potential problems in data interoperability framework, namely (1) heterogeneities of schema, (2) granularity of data, (3) mismatch entity naming and data unit, and (4) inconsistency of data comprehensively?

4 Relevancy

FMIS application promising farmers can manage their farm effectively and efficiently. However, existing FMIS application is relatively expensive for majority farmers in the world that could be categorized as smallholder farmers [4]. Providing FMIS application for free (except package data) for smallholders’ farmers is challenging. A solution that could be offered is developing a FMIS application that only provides functionally that smallholder farmers really needed and utilizing open data to support the application as much as possible. If this model and prototype successfully developed, we can help improve the welfare of smallholder’s farmers through implementation a more effective and efficient farming system. Government institutions and international organizations, as the largest source of external information, are encouraged to provide their data openly so that it can be utilized optimally as wide as possible.

5 Related Work

Until this time, have not been found articles related to constructing an ontology as well as utilizing the use of open data to support FMIS as a whole agribusiness process. There are two articles that may have similarity with this research namely FOODIE [17], [18] and OFIS (Open Farm Information System) by Kim et.al, [19]. However, we will try to present some of the previous research in related to (1) existing ontology in agricultural sectors, (2) the use of open data in agricultural, and (3) method or approach to handling the data interoperability problems.
5.1 Existing Ontology In Agricultural Sectors

There are some international organizations have been trying to propose a concept to utilize open data to support the farm activities such as FOODIE project, AGROXML, and AGROVOC. Moreover, there are two institutions as leading in term of an effort to map the existing ontology in agricultural sectors, namely (1) GODAN (Global Open Data for Agriculture and Nutrition) and (2) University of Montpellier & CNRS. GODAN present list all existing ontology into a map called Vest Agrisemantic (http://vest.agrisemantics.org) and until right now Vest Agrisemantic consist about 399 data standards in the field of food and agriculture. Whereas University of Montpellier & CNRS provides a list of ontology called Agroportal (http://agroportal.lirmm.fr/). Agroportal is an ontology portal which features ontology hosting, search, versioning, visualization, comment, and services for semantically annotating data with the ontologies [20]. Some well-known existing ontologies in agricultural sectors are AGROVOC, AGRO XML, CROP Ontology, FOODIE, OGC (Open Geospatial Consortium), and SSN (Semantic Sensor Network).

5.2 Utilizing data interoperability framework in Agricultural Sector

Utilizing data interoperability framework in Agricultural Sector is not new in research. There are some researcher and organization tried to do research in agricultural sectors by utilizing the advantages of the open data. Table 1 present the summary some previous works relate to utilizing data interoperability framework in agricultural sectors.

Table 1. Summary of some previous works related to utilizing data interoperability framework in agricultural sectors

<table>
<thead>
<tr>
<th>No.</th>
<th>Output</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Open FMIS that integrates into GEOSS and digital earth[21]</td>
<td>OGC, GEOSS</td>
</tr>
<tr>
<td>2</td>
<td>OFIS (Open Farm Information System): Basic farm information, Local weather information, Sensor information[19]</td>
<td>AgroXML, OGC, JSON, RESTFUL, Webservice</td>
</tr>
<tr>
<td>3</td>
<td>API-AGRO, a platform that offers unified access to a dataset[13]</td>
<td>DCAT, GeoJSON</td>
</tr>
<tr>
<td>4</td>
<td>Proposed (1) Precision viticulture, (2) open data for strategic planning, and (3) Integration of logistics through a service provider and farm management[17]</td>
<td>INSPIRE, GEOSS, SEIS</td>
</tr>
<tr>
<td>5</td>
<td>IFIMS (Integrated Farm Information Management System) [22]</td>
<td>SOAP</td>
</tr>
</tbody>
</table>

5.3 Previous work in handling data interoperability problems

As already explained previously, there is not available yet a scientific article that proposed a model, algorithm or approach to handling the four main problems in data interoperability comprehensively. However, there is some existing scientific paper that proposed a model, algorithm or approach to handling partially.
Schema heterogeneity. Ontology matching is the process to identify the degree of semantic relation between two objects. Ontology matching is an approach that many researchers used to overcome heterogeneities schema issues [23]. There are about 100 systems already proposed to overcome the heterogeneities of schema through schema matching method [24, p. 269]. However, the five systems that most use and discuss in a scientific article [25], namely AgreementMaker, COMA++, Cupid, Falcon-AO, and S-Match. Table 2 presents some previous research related to ontology matching.

Table 2. Some previous research related to ontology matching

<table>
<thead>
<tr>
<th>No.</th>
<th>Method/Approach</th>
<th>Evaluation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>1</td>
<td>Combination six basic matchers for string matcher and Autoweight++ structure matcher. [26]</td>
<td>0.95</td>
<td>0.82</td>
</tr>
<tr>
<td>2</td>
<td>Genetic Algorithm [27]</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>Lexical matching and structural matching [28]</td>
<td>0.84</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>Combination of (1) Syntactic and lexical comparison algorithm and wordnet, (2) DSI and SSC matchers, (3) LWC matcher [29]</td>
<td>0.93</td>
<td>0.62</td>
</tr>
<tr>
<td>5</td>
<td>Model Pool parses and matcher library [30]</td>
<td>0.918</td>
<td>0.872</td>
</tr>
</tbody>
</table>

Granularity of data. Granularity alignment is a method to align and overcome the granularity data problem. There are many approaches proposed to tackle the granularity data issues such as neutrality, adaptive mining technique, Clustering analysis, and Granular Computing. However, not found yet granularity alignment algorithm that handles granularity data between two ontologies. Almost all available method focuses on handling granularity data problem in term of data integration or data warehouse.

Mismatch entity naming and data unit. Ontology mapping is a method that many researchers used to overcome the mismatch with entity naming and data unit issues [31]–[35]. An ontology mapping is a specification that describes the data structured transformation process from one schema into a different schema. Some systems or approaches that most users use and discuss in a scientific article [36] as follows (1) Similarity-based: Similarity flooding, Anchor-PROMPT, (2) Terminology-based: COMA, OLA, (3) Instance-based: LSD, GLUE, (4) Semantic reasoning: CTXMATCH, S-match, and (5) Hybrid method: CODI, ServOMap, LogMap, YAM++, AROMA. Table 3 presents some previous research related to ontology matching.
Table 3. Some previous research related to ontology mapping

<table>
<thead>
<tr>
<th>No.</th>
<th>Approach</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1) The Ontology Concepts’ Profile, (2) Bipartite graph, (3) similarity measurement, and (4) Similarity based mapping reusing [37]</td>
<td>0.933</td>
<td>0.904</td>
</tr>
<tr>
<td>2</td>
<td>Classification with Word and CONtext Similarity (CWCONS), Longest Common Substring (LCS) and Tversky’s similarity model [38]</td>
<td>0.88</td>
<td>0.99</td>
</tr>
<tr>
<td>3</td>
<td>Check basic similarities based on some criteria and similarity measurement [39]</td>
<td>0.97</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Inconsistency of data. The ontology selection algorithm is a process to select the “best” external data source from two or more external data sources that have inconsistency data problems. There are some approaches used by researchers to tackle this problem, such as source preferences, linear and non-linear optimization, formal concept analysis, and the combination of Fuzzy Decision making and Gaussian Mixture Model. However, almost all proposed system focuses on handling inconsistency data regarding data integration, not selection the “best” external data source.

6 Research Questions

This research will discuss issues related to the development of a model of open data interoperability in FMIS. Hence, this study intends to answer the following questions:

1. How to develop a comprehensive FMIS ontology as reference ontology?
2. How to develop a model of data interoperability to overcome the heterogeneities of schemes, the granularity of data, mismatch entity naming and data unit, and inconsistency of data problems that may arise when utilizes many external open data sources to support FMIS application;
3. How to evaluate the model as well as an algorithm in term of handling the data interoperability problems?
4. How to develop an FMIS application as proof of a model of open data interoperability that proposed? How to evaluate application performance?
5. How to evaluate the usability and impact of using the FMIS application?

7 Research Framework

The research framework is the basic structure in a graphic presentation of how the research will be carried out. The research framework of this study includes the research problem, approaches to address the problems and technologies involved, output and the outcome expected from the research as shown in Fig. 1. The short explanation for this research framework as follows:
1. **Problem.** This research begins with the problem identified. The fact has shown that utilizing external data sources to support FMIS application face many data interoperability problems. To handle these problems, required a model to tackle the problem.

2. **The process to tackle the problem.** The process to tackle the problem will be formed in a model that consists of two components, namely FMIS ontology schema and algorithm. The FMIS ontology schema used as reference schema and developed based on farmers’ information needs. Whilst the algorithm has two functions, check whether the external data sources have the data interoperability problems and handling if the problems persist. In handling the problems, this algorithm uses “multiple algorithm matcher” that consists of a set of approaches and systems, namely ontology matching, granularity alignment, ontology mapping, and ontology selection;

3. **Output.** The output of this research is to open data interoperable with the FMIS application. It’s mean that external data sources ready to use in FMIS application;

4. **Outcome.** There are two outcomes expected from this research, namely: (1) farmers use this application, and (2) the ultimate outcome is farmers get benefit in utilizing FMIS application, whether economic and non-economic aspect.

### 8 Preliminary Result

I already finished working with farmers’ information needed assessment, mapping into FMIS ontology and construct an OntoFMIS, an FMIS ontology that focuses on handling the information that really needed by small scale farmers. Details for functionalities that needed by smallholders’ farmers as follows:

1. Farmland Identity: farm location and soil characteristics;
2. Farm knowledge: seed descriptions, cultivation technology, and map of areas that plant the same commodity
3. Farm monitoring: crop growing monitoring and cropping record
4. Farm recommendation: Pest and disease handling recommendation, and seed recommendation
5. Farm consultation: ask and consultation to the expert directly
6. Farm marketplace: information on market demand (product required, volume, and price), market price, and another region that growing the similar crop.

![Diagram](image)

**Fig. 2. Core class and subclass of OntoFMIS**

Whereas the OntoFMIS core class and subclass, with considering utilizing existing ontology, presented on **Fig. 2**

### 9 Approach

Block diagram as the initial conceptual model of FMIS data interoperability presented in **Fig. 3**. This initial conceptual model divided into four blocks to handle the required information from the user, namely: (1) identify sources of data, (2) handling the internal data source, (3) handling the external data source and (4) collecting and processing data. The short explanation of the process that will be performed in this conceptual model as follow:

- **Identify the source of data**
  - The process will be started to request information from the user. This request information will be handled in “identify the source of data” block.
  - FMIS application will select the module that matches the request and identifying the source of data required by the module. There are two possibilities source of data, (1) data from internal sources, (2) data from external sources.
    - For data that identified from internal, the process will be continued to handling the internal data source block.
    - For data that identified from the external, the process will be continued to handling the external data source block.

- **Handling internal data source**
  Process inside handles internal data source block as follows:
— Data that required from the user, the module will provide the e-form as an interface to input data;
— Data that required already available, the module will execute the interpreter and collect process

• **Handling external data source**
  Process inside handles external data source block as follows:
  — An algorithm to identify the data interoperability problems is employed.
  — An algorithm to handle the data interoperability problems that were identified

• **Collecting and Processing data**
  — The result of both processes will be pushed into temporary storage use ETL (Extract, Transform and Load) for data process and provide information as user required.

![Fig. 3. Conceptual Model to Handle Four Main Data Interoperability Problem](image)

10 **Evaluation Plan**

Two types of evaluation will be performed, evaluation of the algorithm performance and evaluation of the conceptual model. The precision and recall indicator will be used to evaluate the algorithm performance with formula as presented in Table 4.

<table>
<thead>
<tr>
<th>No</th>
<th>Indicator</th>
<th>Explanation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Precision</td>
<td>Precision is the ratio of correctly predicted positive observations to the total predicted positive observations</td>
<td>[ \text{precision} = \frac{TP}{TP + FP} ]</td>
</tr>
<tr>
<td>2</td>
<td>Recall</td>
<td>The recall is the ratio of correctly predicted positive observations to all observations in the actual class</td>
<td>[ \text{recall} = \frac{TP}{TP + FN} ]</td>
</tr>
</tbody>
</table>

Where: \( TP = \text{true positive; } TN = \text{true negative; } FP = \text{false positive; } FN = \text{false negative} \)

Whereas for evaluating the conceptual model will be conducted through experts evaluation with the close questionnaire regarding several indicators as shown in Table 5.
Table 5. Indicators to evaluate the conceptual model

<table>
<thead>
<tr>
<th>No</th>
<th>Indicators</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Completeness</td>
<td>Check the coverage of user requirements.</td>
</tr>
<tr>
<td>2</td>
<td>Reusability</td>
<td>Checks whether the model employs the previously developed models and whether this model can be reused in future</td>
</tr>
<tr>
<td>3</td>
<td>Relevancy</td>
<td>Find the relevancy between the concepts present in the model and the ones required by the users</td>
</tr>
<tr>
<td>4</td>
<td>Reliability</td>
<td>Check whether the model if it is prone to failure or no.</td>
</tr>
<tr>
<td>5</td>
<td>Practicability</td>
<td>Check whether the model employs the concepts or elements that are realistic and can be materialized.</td>
</tr>
</tbody>
</table>

11 Reflection

A conceptual model that focuses on addressing four major data interoperability problems comprehensively to support FMIS is the novelty of this research. I cannot say another researcher failed because, in fact, that is no researcher has focused on research similar to the topic of my research. I believe I will succeed in working on the topic of this research because some of my research has actually been done. Choose the most promising approach, do it with a little adjustment or improvement for each problem, and integrate all selected approaches as a comprehensive process is my strategy. Last but not least, the support of many experts in the fields of agricultural cultivation, land use, and socio-economics at the Agricultural Research and Development Agency (IAARD) was a factor that made me believe that I could succeed.

12 Acknowledgment

This research was fully supported by SMARTD Project (Loan IBRD No.8188), IAARD, Ministry of Agriculture. We thank our colleagues from IAARD who provided insight and expertise that greatly assisted the preparing this research proposal. We would like to show our gratitude to the Dr. Vatcharaporn Esichaikul as Chairperson and Dr. Chutiporn Anutariya as co-chair for my dissertation for your intensive discussion as well as your valuable correction, comment, and suggestion, therefore, I can finish my research proposal. We are also immensely grateful to Prof. Dr. Matthew Dailey and Dr. John K.M Kuwornu as committee members for any comments and suggestion during the preparation of this research proposal.

References

[3] KILM-ILo, Key indicators of the labor market. 2014.


