

Survey and Modeling Metadata Schema Relationship in Agriculture Domain for Better Metadata Schema Service

Pitsanu Lousangfa, Naiyana Sahavechaphan, Marut Buranarach, Sornthep Vannrat, and Asanee Kawtrakul

National Electronics and Computer Technology Center, Thailand
112 Thailand Science Park, Pahon Yothin Rd
Klong 1, Klong Luang, Pathumthani 12120 Thailand
Pisanu.lou@nectec.or.th, naiyana.sah@nectec.or.th, marut.bur@nectec.or.th

Abstract

Information is typically described by a set of element values based on either *locally* or *globally* pre-defined element set. Essentially, the application of globally pre-defined elements set, termed as *standard* metadata schemas, enables an efficient information sharing and interoperability across repositories. According to this, many organizations and committees (i.e. FAO and FDGC) have thus developed various standard metadata schemas which are distributedly available via the internet. The discovery for metadata schema of choice through various web sites, however, can be time-consuming and error prone, making (i) similar information across repositories to be defined using in-house schemas or even different standard ones; and (ii) organizations or committees to define redundant metadata schemas for similar information. These in turn hinder information sharing and interoperability.

To address this, we have initiated *Metadata Schema Service* as the first portal for existing metadata schema available at different web sites. In this paper, we thus propose the classification and relationship model, namely CRM, as an underlying model of Metadata Schema Service. Particularly, CRM is a two-level model for classifying and relating metadata schemas. The upper level focuses on classification to facilitate the discovery for desired metadata schema, while the lower level defines relationship of metadata schema to support the development of new metadata schema. We have evaluated the applicability of CRM in agricultural and related domain as agricultural information sharing and interoperability is quite significant for agricultural countries. Specifically, we have (i) surveyed existing metadata schemas in agricultural and related domain; (ii) analyzed them in term of their classification and relationship; and (iii) instantiated CRM with the previous analysis result. In addition, the usefulness of CRM is shown with respect to metadata schema browsing and understanding.

Keywords: metadata schema, metadata schema service, information sharing, information interoperability

1. Introduction

Information is typically described by a set of element values based on either *locally* or *globally* pre-defined element set. Essentially, the application of globally pre-defined element set, termed as *standard* metadata schema, enables an efficient information sharing and interoperability across repositories. According to this, many organizations and committees have thus defined various standard metadata schemas with respect to their need and usage. For example, Food and Agriculture Organization (FAO) initiates “AgMES”¹ – to describe agricultural documents and “FiMES” – to describe fishery information, while Federal Data Geographic Committee (FDGC) defines “CSDGM” to describe digital geospatial data. These standard metadata schemas are made available on-line via the internet, allowing people to conveniently access, comprehend and utilize them. Their availability, however, is typically distributed throughout different Web sites. In addition, it is likely that there would be different standard metadata schemas (or elements) created for the same purpose but by different organizations or committees. As a consequence, people need to visit each individual Web site (see Figure 1 (a)) in order to discover for the desired schema for their repositories, or to ensure that the to-be developed schema (or element) has not yet been defined. Undoubtedly, this process is time-consuming and error prone which may cause (i) repository developers to use in-house metadata schema or even different standard metadata schemas to describe similar information; and (ii) organizations or committees to newly define metadata schemas (or elements) similar to existing ones. This would finally hinder information sharing and interoperability across repositories.

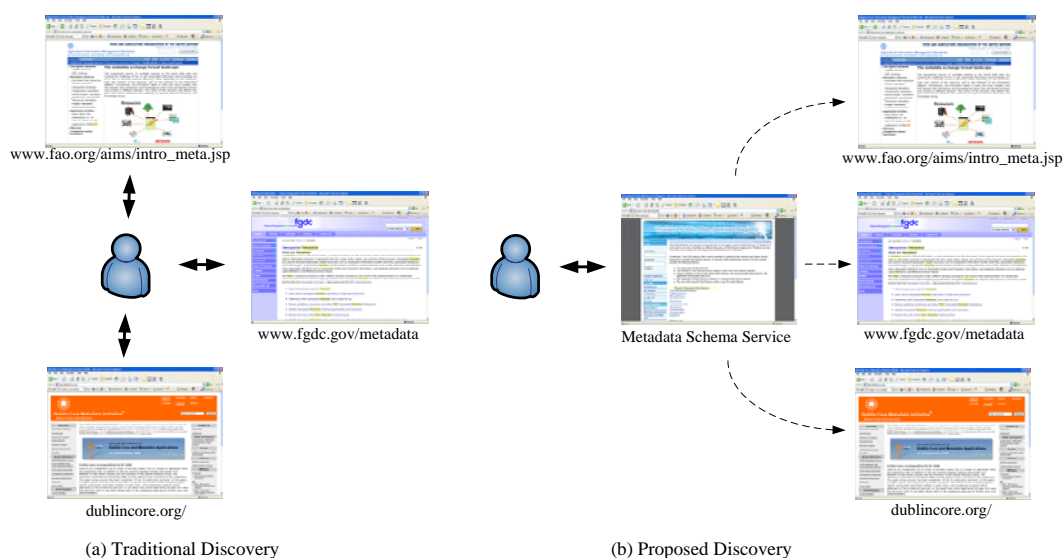


Figure 1: The Discovery for Metadata Schemas

To alleviate the above process and hence to promote information sharing and interoperability across repositories, we have initiated *Metadata Schema Service* as the first portal for existing metadata schema available at different Web sites. Like UDDI registry for Web Services, Metadata Schema Service provides the registration and discovery of metadata schemas. In this work, we focus on the metadata schema discovery. Its main goal is to support metadata schema

¹ AgMES is an extension of DCMES developed by Dublin Core Metadata Initiative (DCMI)

users to select the metadata schema of choice, and metadata schema creators to efficiently define new metadata schema by extending existing schemas or reusing elements already defined in any existing schemas. For simplicity, we refer *users* and *creators* as metadata schema users and creators respectively throughout the paper. To achieve this, in this paper, we thus propose the classification and relation model, namely CRM, as the underlying model for Metadata Schema Service. In particular, CRM is a two-level model for classifying and relating metadata schemas. The upper level focuses on classification to facilitate the discovery for desired metadata schema, while the lower level defines relationship of metadata schema to support the development of new metadata schema. We have evaluated the applicability of CRM in agricultural and related domain as agricultural information sharing and interoperability is quite significant for agricultural countries. Specifically, we have (i) surveyed existing metadata schemas in agricultural and related domain; (ii) analyzed them in term of their classification and relationship; and (iii) instantiated CRM with the previous analysis result. In addition, the usefulness of CRM is shown with respect to metadata schema browsing and understanding.

This paper is organized as follow. Section 2 defines CRM. Section 3 and 4 describes metadata schema classification and relationship, respectively. The application of CRM is illustrated in Section 5. Finally, Section 6 concludes the paper.

2. Classification and Relationship Model: CRM

Recall that Metadata Schema Service (MSS) is to assist users to select the metadata schema of choice, and creators to define new metadata schema. In this section, we thus define the overall concept of *Classification and Relationship Model* (CRM). Specifically, as shown in Figure 2, it consists of two levels: (i) the classification level – in that facets and their underlying terms are defined. Here, each metadata schema registered to MSS is then associated to each facet via one or more terms that classify it; and (ii) the relationship level – in that the dependency and similarity of metadata schemas are qualitatively and quantitatively evaluated. Here, each metadata schema to be registered to MSS is compared against existing metadata schemas in the repository of MSS to determine their dependencies and similarities. Details of metadata schema classification and relationship are described in Section 3 and 4, respectively.

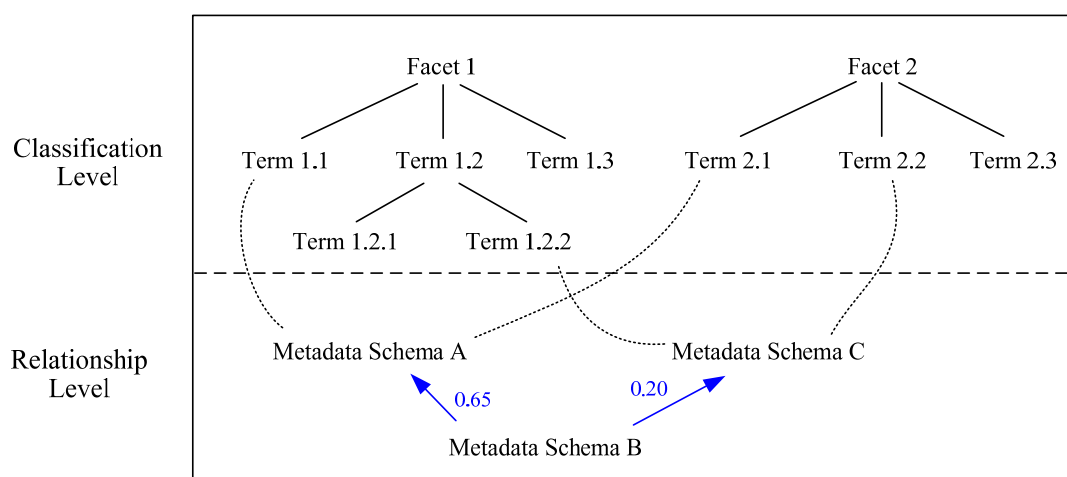


Figure 2: The Conceptual Classification and Relationship Model (CRM)

3. Metadata Schema Classification

We have applied facet-based classification technique (Vickery 1960) and hierarchical classification technique (Gordon 1987) to classify metadata schema. Specifically, different facets are defined along with their hierarchical terms. Based on a set of surveyed metadata schemas (see Appendix), we have analyzed and then defined 2 essential facets along with their terms that enable their classification as shown in Table 1. These facets include (i) *information* facet – to define types of information such as document, person and event in which a metadata schema is used to describe it; and (ii) *application* facet – to define **application domain** in which a metadata schema would be utilized or referred in its repository. In Table 1, an information facet consists of 8 terms wherein the term “rice” consists of 3 hierarchical terms: “variety”, “cultivation” and “protection”. The application facet consists of 2 terms, “library” and “agriculture”. Here, the term “library” has 3 hierarchical terms – “DSpace” (DSpace 2008), “GreenStone” (GreenStone 2008) and “Koha” (Koha 2008), while the term “agriculture” has “Fertilizer Expert System” (Kawtrakul et al. 2008) and “Disease/Pest Warning System”.

Information Object	Application
document-like object	Library.DSpace
GIS	Library.GreenStone
event	Library.Koha
person/Expert	agriculture.Fertilizer Expert System
project	agriculture.Disease/Pest Warning System
organization	
rice.variety	
rice.cultivation	
rice.protection	
soil	

Table 1: Facets and Terms for Metadata Classification in Agricultural Domain

According to Table 1, each surveyed metadata schema is associated with one or more terms in each facet as shown in Figure 3. As an example, DCMES is a metadata schema that is used to describe book, article, journal and any publications, and to implement Library application, namely DSpace. DCMES is thus connected to the term “document-like object” in the information facet and the term “library.DSpace” in the application facet.

In addition to classification, terms can be used to browse related metadata schemas, enabling the selection of the desired metadata schema. For instance, six metadata schemas DCMES, FiMES, AgMES, Marc21, AGLS and EDNA are resulted from browsing metadata schema using the term “document-like object”. This indicates that there are various metadata schemas for describing document information. With the term “agriculture.Fertilizer Expert System”, metadata schemas soil_char_*, rice_cul_*, rice_var_* and farmer are returned. This shows that several metadata schemas are utilized while implementing such application.

It should be noted that metadata schemas soil_char_*, rice_cul_*, rice_var_* and rice_pro_* are non-standard metadata schemas that are defined by different organizations. The involving organizations should be aware of these to promote the sharing and interoperability of information in agricultural and related domains.

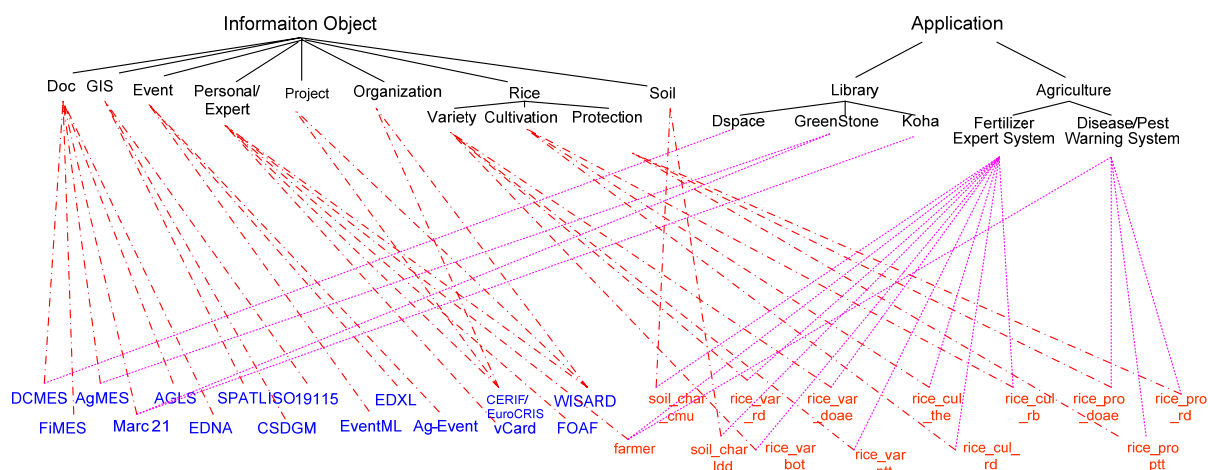


Figure 3: Metadata Schema Classification in Agricultural and Related Domain

4. Metadata Schema Relationship

We have defined relationship of metadata schemas in term of (i) *dependency* and (ii) *similarity*. The target metadata schema S_j is *dependent* on the source metadata schema S_i , if schema S_j either extends *all* elements defined in schema S_i , or reuses *some* elements defined in schema S_i . Here, the application of namespaces is used to determine whether an element e_j in schema S_j is defined as an element e_i in schema S_i . The target metadata schema S_j is *similar* to the source metadata schema S_i , if schema S_j has some elements in common with elements in schema S_i , with respect to element labels. Here, an element e_j in schema S_j is said to be in common with an element e_i in schema S_i only if there labels are the same.

4.1 Metadata Schema Dependency

Consider Table 2 as an example. Here, the schema DCMES consists of 15 elements. The schema AGLS extends DCMES and defines 4 new elements. Finally, the schema AgMES extends DCMES and reuses 2 existing AGLS elements and defines 1 additional element. The schema AGLS is thus dependent on the schema DCMES, while the schema AgMES is dependent on both schemas DCMES and AGLS. The dependency of these schemas can be illustrated in Figure 4a. Here, the arrowed line depicts the dependency of two metadata schemas wherein the head and tail of an arrow refer to source schema and target schema, respectively.

No.	Element>	DCMES	AGLS	AgMES
1	Contributor	√	√	√
2	Coverage	√	√	√
3	Creator	√	√	√
4	Date	√	√	√
5	Description	√	√	√
6	Format	√	√	√
7	Identifier	√	√	√
8	Language	√	√	√
9	Publisher	√	√	√
10	Relation	√	√	√
11	Rights	√	√	√
12	Source	√	√	√
13	Subject	√	√	√
14	Title	√	√	√
15	Type	√	√	√
16	Availability		√	√
17	Function		√	
18	Audience		√	√
19	Mandate		√	
20	Citation			√

Table 2: The Elements defined in schemas DCMES, AGLS and AgMES

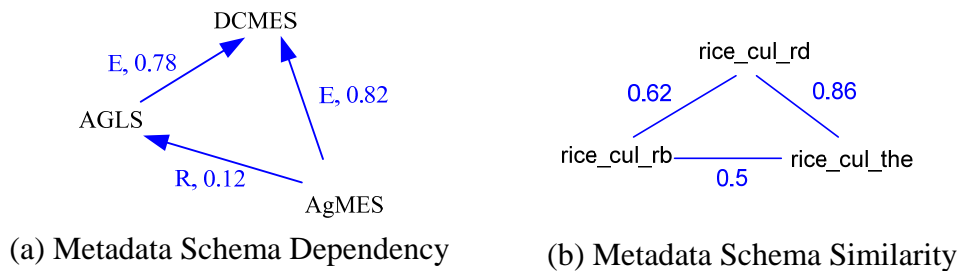


Figure 4: Metadata Schema Relationship

Moreover, the degree of dependency between two metadata schemas can be qualitatively and quantitatively defined. The qualitative degree is defined as either extension (E) or reuse (R). Here, an extension degree refers that *all* elements defined in the source schema are taken as part of the target schema, while a reuse degree indicates only *some* (not all) elements. The quantitative degree is computed with respect to Equation 1.

$$\text{Equation 1: Degree of Dependency} = \frac{|E_S \cap_n E_T|}{|E_T|}$$

where E_S denotes element set in source metadata schema, E_T denotes element set in target metadata schema, \cap_n denotes the match operator of two given element sets with respect to namespaces, and $| \cdot |$ denotes the number of elements in a given element set.

As shown in Figure 4a, the schema AGLS extends the schema DCMES. The dependency degree value of 0.78 indicates that 78% of elements in the AGLS are extended from the

DCMES, while 22% of its elements are newly defined. Similarly, the schema AgMES extends the DCMES schema and reuses the AGLS schema. Here, 82% of its elements are taken from DCMES, 12% from AGLS and 6% are newly defined elements.

4.2 Metadata Schema Similarity

A metadata schema S_j is considered to be similar to a metadata schema S_i only if some elements of schema S_j share the same labels with some elements of schema S_i . For example, “soil preparation” is an element label appeared in both the *rice_cul_rd* and *rice_cul_rb* schemas. Thus *rice_cul_rd* and *rice_cul_rb* are similar to each other. Moreover, the degree of similarity can be quantitatively defined with respect to Equation 2.

$$\text{Equation 2: Degree of Similarity} = \frac{2 \times |E_S \cap_1 E_T|}{|E_S| + |E_T|}$$

where E_S denotes element set in source metadata schema, E_T denotes element set in target metadata schema, \cap_1 denotes the match operator of two given element sets with respect to element labels, and $| \cdot |$ denotes the number of elements in a given element set.

As shown in Figure 4b, the similarity degree between two independent metadata schemas, *rice_cul_rd* and *rice_cul_rb*, is computed as 0.62. The similarity model is appropriate for applying with similar metadata schemas that are independent. Particularly, two schemas should consider interoperability, i.e. by reusing or extending existing elements, if they have high similarity degree.

5. Application of CRM

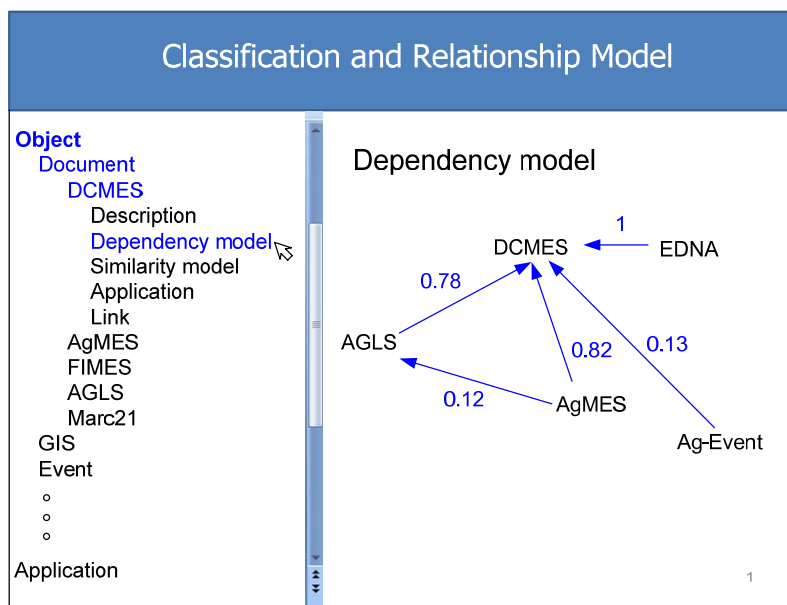


Figure 5: A Web Application of Metadata Schema CRM

According to Section 3-4, we will implement a CRM Web application as shown in Figure 5. Classification is presented at the left-hand side, while at the right-hand side the relationship is shown. In the classification aspect, a set of defined facets along with their terms are displayed. Each term is associated with a set of corresponding metadata schemas. As per a metadata schema, each user can browse for its description, dependency model, similarity model, application and the like to its home page. Consider Figure 5 once again. Here, DCMES is a metadata schema classified in facet “Object” and term “Document”. Its dependency model is actively shown when a user selects “Dependency model” under DCMES.

6. Conclusion

In this paper, we propose Classification and Relationship Model (CRM) for metadata schemas which can be applied for both standard and non-standard metadata schemas. Classification is based on facet-based classification and hierarchical classification techniques. Relationships between two metadata schemas are defined as schema dependency and schema similarity. Dependency indicates degree of extension and reuse of existing elements. Similarity indicates degree of common label sharing that may suggest possible schema interoperation. Some potential benefits of CRM include supporting (i) metadata users to reuse existing standards to describe similar information; and (ii) organizations or committees to not newly define metadata schemas similar to existing ones but to extend or reuse the existing ones. This would eventually encourage information sharing and interoperability across repositories.

7. References

- DSpace (2008) [Online]. Available at <http://www.dspace.org>.
- Gordon, A. D., (1987) A Review of Hierarchical Classification. Journal of the Royal Statistical Society, Series A (General), Vol. 150, No. 2: 119-137
- Greenstone digital library software (2008) [Online]. Available at <http://www.greenstone.org>
- Kawtrakul, A., (2008) A Smart Mobilized Fertilizing Expert System: 123 Personalized Fertilizer. Proceeding of World Conference on Agriculture Information and IT: IAALD AFITA and WCCA. Tokyo, Japan. August 24-27, 2008.
- Koha (2008) [Online]. Available at <http://www.koha.org>.
- Vickery, B. C. (1960) Faceted Classification: A Guide to Construction and Use of Special Schemes. Aslib, London.

Appendix

Metadata	Owner/Creator	Description	URL
Dublin Core Metadata Element Set (DCMES)	Dublin Core Metadata Initiative (DCMI)	A vocabulary of fifteen elements for use in resource description	http://www.dublincore.org/
Agricultural Metadata Element Set (AgMES)	Food and Agriculture Organization of the United Nations (FAO)	An extension of Dublin Core that promotes the use of metadata for information resources in the domain of agriculture	http://www.fao.org/aims/intro_meta.jsp
Fisheries Metadata Element Set (FiMES)	FAO	A proposed standard for the encoding of structured documents within the domain of fisheries research and policy	http://www.fao.org/fi/figis
Australian Government Locator Service (AGLS)	National Archives of Australia	A set of 19 descriptive elements to improve the visibility and accessibility of service and information over the internet	http://www.naa.gov.au/records-management/
AGRIS Application Profile (AGRIS AP)	FAO	A metadata standard created specifically to enhance the description and exchange of agricultural document	http://www.fao.org/agris/
Marc 21	Library of Congress	A standard for representation and communication of bibliographic information in machine-readable form	http://www.loc.gov/marc
Spatial Standard and Norms (SPATL)	FAO	An internal use standard in order to rationalize, harmonize and advance its GIS and cartographic activities	http://www.fao.org/spatl
Geographic information Metadata (ISO 19115)	International Organization for Standardization (ISO)	The schema required for describing geographic information and services	http://www.iso.org/iso/iso_catalogue.htm
Content Standard for Digital Geospatial Metadata (CSDGM)	Federal Geographic Data Committee	A standard to provide a common set of terminology and definitions for the documentation of digital geospatial data	http://www.fgdc.gov/metadata/csdgm/
Common European Research Information Format (CERIF/EuroCRIS)	European Commission (EC)	Standard created to exchange project and organization information to foster the diffusion of research information across Europe	http://cordis.europa.eu/cerif/
Web based Information Services for Agricultural Research for Development (WISARD)	Wageningen International	A web-based platform that search information on experts, organizations and projects in agricultural research	http://www.wisard.org
vCard	Internet Mail Consortium (IMC)	A file format standard for personal data interchange, specifically electronic business cards	http://www.imc.org/pdi/vcardoverview.html
Friend-of-a-friend (FOAF)	RDFWeb	A project to create a web of machine-readable homepages describing people	http://www.foaf-project.org/

Metadata	Owner/Creator	Description	URL
Education Network Australia (EDNA)	Education Network Australia	Support interoperability across all sectors of education and training in Australia	http://www.edna.edu.au/metadata
Emergency Data Exchange Language (EDXL)	OASIS	A standard message distribution framework for data sharing among emergency information systems	http://www.oasis-open.org/committees/emergency
EventsML-G2	International Press Telecommunications Council	A standard for conveying event information in a news industry environment	http://www.iptc.org/EventsML
Ag-Event AP	FAO	Metadata application profile used for exchanging event information within the agricultural community	http://www.fao.org/aims/ap_applied.jsp
rice_var_rd *	Rice department, Ministry of Agriculture and Cooperatives (MOAC), Thailand	Description about rice varieties by the Rice Department	http://www.ricethailand.go.th
rice_var_doea *	Department of Agricultural Extension, MOAC, Thailand	Description about rice varieties by the Department of Agricultural Extension	http://www.doea.go.th/library
rice_var_ptt *	Pathumthani Rice Research Center, Rice Department, MOAC, Thailand	Description about rice varieties by the Pathumthani Rice Research Center	http://ptt.ricethailand.go.th
rice_var_bat *	Bio Agro Thai Company, Thailand	Description about rice varieties by the Bio Agro Thai co.	http://www.bio-thai.com/rice
rice_cul_rd *	Rice department, MOAC, Thailand	Description about rice cultivation by the Rice Department	http://www.ricethailand.go.th
rice_cul_rb *	Rakbankerd Company, Thailand	Description about rice cultivation by the Rakbankerd co.	http://www.rakbankerd.com/agriculture
rice_cul_the *	Thai Encyclopedia	Description about rice cultivation by a Thai encyclopedia	
rice_pro_rd *	Rice department, MOAC, Thailand	Description about rice protection by the Rice Department	http://www.ricethailand.go.th
rice_pro_doea *	Department of Agricultural Extension, MOAC, Thailand	Description about rice protection by the Department of Agricultural Extension	http://www.doea.go.th/library
rice_pro_ptt *	Pathumthani Rice Research Center, MOAC, Thailand	Description about rice protection by the Pathumthani rice research center	http://ptt.ricethailand.go.th
soil_char_ldd *	Land Development Department, MOAC, Thailand	Description about soil characteristics by the Land Development Department	http://www.ddd.go.th/thaisoils_museum/INDEX.HTM
soil_char_cmu *	Faculty of Agriculture, Chaing Mai University, Thailand	Description about soil characteristics by Chaing Mai University	http://www.agri.cmu.ac.th/main/
farmer *	NAIST Lab, Kasetsart University, Thailand	Description about farmers used in the fertilizer expert system	http://vivaldi.cpe.ku.ac.th/~vasu/soil/main.php

Table 3: Survey of Metadata Standards in Agricultural Domain

Note: * indicates non-standard metadata