The Traditional Malay Textile (TMT) Knowledge Model: Transformation towards Automated Mapping

Syerina Azlin Md Nasir, Nor Laila Md Noor, Suriyati Razali

Abstract**—** The growing interest on national heritage preservation has led to intensive efforts on digital documentation of cultural heritage knowledge. Encapsulated within this effort is the focus on ontology development that will help facilitate the organization and retrieval of the knowledge. Ontologies surrounding cultural heritage domain are related to archives, museum and library information such as archaeology, artifacts, paintings, etc. The growth in number and size of ontologies indicates the well acceptance of its semantic enrichment in many emerging applications. Nowadays, there are many heritage information systems available for access. Among others is community-based e-museum designed to support the digital cultural heritage preservation. This work extends previous effort of developing the Traditional Malay Textile (TMT) Knowledge Model where the model is designed with the intention of auxiliary mapping with CIDOC CRM. Due to its internal constraints, the model needs to be transformed in advance. This paper addresses the issue by reviewing the previous harmonization works with CIDOC CRM as exemplars in refining the facets in the model particularly involving TMT-Artifact class. The result is an extensible model which could lead to a common view for automated mapping with CIDOC CRM. Hence, it promotes integration and exchange of textile information especially batik-related between communities in e-museum applications.

Keywords**—**automated mapping, cultural heritage, knowledge model, textile practice

I. INTRODUCTION

PRESERVING national heritage has become a key agenda in many countries in the world. In Malaysia, one of its preserved values is Traditional Malay Textiles (TMT) such as Songket, Batik, Pua Kumbu, Cindai and etc. [18]. At present, most of TMT collections have been exhibited in Malaysian museums using traditional curation methods with constraints of limited descriptions and articulations to portray the distinctiveness of TMT artifacts. There are also those individuals who posses and own priceless TMT artifacts and knowledge and those who produce these textiles and preserve the techniques of making them. Apart from that, attempts have been made by several institutions to initiate the digitizing of documents and artifacts. However, most of the works are isolate, unknown and for private usage. Therefore, work has been done to collect all these Malay textiles to strengthen the conservation for national cultural heritage preservation in museums and documented as TMT Knowledge Model [2]. To fully grasp the advances of web technology, the construction of this model will be extended by mapping the model with the standard ontology in cultural heritage domain, CIDOC CRM.

The growth in number and size of ontologies indicates the well acceptance of its semantic enrichment in many emerging applications. Each ontology being designed has been represented in different ways with different purposes in mind. The current scenario requires these available ontologies to be mapped together [9], [11] which raises an issue to find similar entities and semantic correspondences that exist between them. In this case, CIDOC CRM provides both a global and extensible model into which data originating from distinct sources can be mapped and integrated, and base concepts that future metadata initiatives could build on when developing domain specific vocabularies [29]. Mapping TMT Knowledge Model and CIDOC CRM raises several challenges that need further investigation. The first and primary challenge is the diversity of the metadata terms used to describe and structure the collections [3]. The second challenge is the rigid structure of TMT Knowledge Model which makes it unable to deliver very rich meaningful artifacts [20]. Finally, the result of the mapping will hopefully allow room for ontology growth in e-museum application and create a wealth of opportunities for interactions with collections of experts' knowledge from the community.

Numerous studies and a number of solutions have been proposed to solve the ontology mapping problem. There are reports on the success and difficulties of ontology building using either manual approaches or with tool supports. Each claims one is better than the other in terms of its performance and scalability. [24] states that a lot of research has been monopolized by heuristics and machine learning approach (automation). Various mapping techniques have been

M.N.Syerina Azlin is with the Faculty of Computer and Mathematical Science, Universiti Technologi MARA, 40000 Shah Alam, Malaysia (phone: 6013- 977-2706; fax: 603-5544 2000; e-mail: syerina@ kelantan.uitm.edu.my).

M.N.Nor Laila is with the Faculty of Computer and Mathematical Science, Universiti Technologi MARA, 40000 Shah Alam, Malaysia (e-mail: norlaila@tmsk.uitm.edu.my).

R.Suriyati is with the Faculty of Computer and Mathematical Science, Universiti Technologi MARA, 40000 Shah Alam,, Malaysia (e-mail: suriyati@tmsk.uitm.edu.my).

explored and categorized [15] to further understand the mapping process. At the same time, a lot of work has been reported on the development of heuristics and machine learning tools (e.g. AnchorPrompt [26]; COMA++ [10]; GLUE [4]; S-Match [16]). Most of these tools, methods or techniques are available. However, in reality, they are not freely available or still in development stage or in the form of algorithms that need clarifications before it can be further exploited. Therefore, this brings us back to the basic of mapping by reviewing the manual approach. Some of the works claim that manual mapping is laborious, time consuming, error prone, difficult to maintain and update [10], [19] and yet, [7] suggest the creation of benchmark ontologies by manually defining mapping between concepts. [25] elaborates on the concept of 'shared ontology' which is agreed upon by developers from various applications for future adoption and extension to their specific application. Thus, this study aims to transform the TMT Knowledge Model accordingly in which would enable the representation of TMT by modeling in an adequately consistent way the conceptualization of the reality behind textile practice.

The remainder of this paper is organized as follows. Section II summarizes the previous work concerning standard ontology, CIDOC CRM. Section III then describes the approach used in this work. The actual results are presented in Section IV. The paper concludes with Section V that discusses on the anticipated future work directions.

II. RELATED WORK

Of late, the number of ontologies created for various domains has been increasing. Each specific domain expects a well-developed ontology, where most of the vocabularies that are needed for meaningful interaction and communication are available between people and application systems, hence, is able to handle vast amounts of distributed and heterogeneous computer-based information. There are several domain ontologies which gain popularity among researchers like in the area of medicine [28], [6], food industry, tourism [22], [8], bioinformatics [2], [1] and cultural heritage [13]. Studies on creating a common ontology also rise in different domains. Among the earliest ontologies being built are Suggested Upper Merged Ontology (SUMO) and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). Over the years, several other reference ontologies have been constructed for specific domains of discourse.

The existence of upper level ontology in specific domains helps many ontology developers in creating ontology which conform to the standard. For example, in cultural heritage domain, ontology for cultural heritage information like CIDOC CRM is referred. Among works that adopt CIDOC CRM are the implementation of cultural heritage web information system on the Semantic Web such as SCULPTUER, AMA, CRM-EH and Digital Museums. SCULPTUER project handles museum multimedia collections by mapping the museum's partner legacy system with CIDOC-CRM for cross collection searching [30]. The Archive Mapper for Archaeology (AMA) project aims to create tools for semi-automated mapping from archaeological archive materials, reports, catalogues and databases to CIDOC-CRM [14]. The English Heritage Centre for Archaeology Ontological Model (CRM-EH) is an extension of the CIDOC-CRM which aims for effective search across multiple different databases and their associated controlled vocabularies [5]. The project on building global ontology for distributed digital museums employs CIDOC-CRM to identify and classify the semantics of data derived from local museums [23]. All these projects utilize CIDOC-CRM as a common standard either as global or extensible model.

TABLE I
SUMMARY OF CIDOC CRM HARMONIZATION PROJECTS

	ABC	**MPEG-7**	**FRBR**	**Dublin Core**
Domain area	Digital library	Multimedia in museum collection	Intellectual creation process, performing arts, recording and publication work, bibliographic practice	Digital library
Purpose	Provide a common conceptual model to facilitate interoperability between metadata vocabularies from different domains	Provide a single ontology for describing managing multimedia in museum	Capture & represent underlying semantics of bibliographic information	Preserve the semantics of the DC records that correspond to different material types
Source of Data	Museum metadata records with associated multimedia digital objects	Museum multimedia content	Museum and bibliographic information	Metadata schemas of digital libraries
Methods for ontology mapping	Metadata model with temporal semantics for the class of descriptions (Manual)	Merge both models and ***extend*** it through MPEG-7 specific sub-class and sub-properties (Manual)	Merge both models by mapping the semantic of bibliographic records (Manual)	Map DCMI Type vocabulary to CIDOC CRM (Manual)
Tool or system associated	None	None	None	None

On the other hand, harmonization works are also carried out between CIDOC CRM and other ontologies such as ABC Model [21], MPEG-7 [17], Functional Requirements for Bibliographic Records [4], [12] and Dublin Core [20]. Harmonization is defined as a *process of modifying two ontologies, preserving their intended functionality but integrating them into a coherent wider model* [13]. Apart from that, CRM also supports integration in a diverse range of different domains including e-science and biodiversity [12]. Cultural resources require metadata schema rich in structure and semantic to cover material heterogeneity and variety of memory institutions [20]. It is important for ontology developers to include semantics and structures as a way of transforming this unstructured information into a format that machines could understand. The harmonization projects are summarized as shown in Table 1. All these projects aim to create a single ontology that represents the conceptualization of reality in the domain area. The works are done manually which involve experts and individuals from related institutions before the common ontology is agreed upon. In relation to this study, the investigation on CIDOC CRM is made to provide an insight on the importance of an upper ontology in mapping between two ontologies, especially when related to TMT.

III. Methodology

A. Data Description

This study will employ Traditional Malay Textile (TMT) Knowledge Model which is designed by Suriyati et al. [32]. As defined by Siti Zainon, I [31], TMT means the handicrafts which relates to the process of weaving invented by human beings for daily clothing in Malay Archipelago with the understanding of historical factors, symbolic and aesthetic values. The research is solely focused on the historical factors and limited to the description of textile from the Malay Peninsula. Due to the current nature of the Malay Textile Knowledge and Artifacts that are scattered among domain experts and novices, this model was designed. The taxonomy aims to represent Malay Textile content in organized structures for retrieval by user preference and to allow association between system development and user interaction platform (Malay Textile Community and Museum Community that have different backgrounds).

B. Motivating Examples

Example 1. ABC ontology is a model for the exchange and integration of digital library information. It was designed to *model physical, digital and analogue objects of all media types; abstract concepts such as intellectual content and temporal entities; and describe other entities that occur across many domains.* It encompasses 14 classes and 25 properties that present the conceptual basis for generating domain, role or community specific vocabularies across different domains as guidance to interested communities. The primary concerns are on volatile digital objects and intellectual property rights [21]. The mapping was done in 2001 by using the OntoClean approach previously used in integrating WordNet with OCT ontology [27].

In this example, the study described a detailed specification of all ABC classes and properties and represented it in graphical form. This provides an example on how entities are modeled and relationships are built between them. As such, it guides on how to build descriptive ontology [21]. It also provides examples of the model implementation which allow for construction of metadata repository of RDF description and sophisticated queries through the search interface. It pays more attention on the conceptualization of object transformation over time in which in this case, on TEMPORALITY category.

Example 2. The Moving Pictures Expert Group (MPEG) released the “Multimedia Content Description Interface” standard for MPEG-7 for describing multimedia content. The combining effort between MPEG-7 and CIDOC CRM metadata models resulted in the creation of a standardized model for describing and managing museum multimedia content. The work merged both ontologies and further extended MPEG-7 with additional sub-classes and sub-properties to cater for multimedia concepts and descriptions. The vast diversity and variety of multimedia resources especially in the museum domain makes it a necessity to record these elements in the most eloquent manner for preservation and dissemination purposes [28].

Through this example, it is learned that certain categories covered by CIDOC CRM requires for extension. This gives an alternative on how to represent the detailed description on the making process of textiles which is not represented in the model. The study shows how extension is carried out to further describe the digital multimedia contents by extending the INFORMATION OBJECT category.

C. Purpose

The TMT Knowledge Model was originally designed to preserve the Malay textiles through semantic relationship between artifact and its precedent knowledge by mapping the model with CIDOC CRM. This study aims to create Batik Heritage Ontology as the result of the mapping between both models to embody the whole process of batik making. However, the current state of the TMT Knowledge Model hinders the mapping process and requires for transformation to take place beforehand. Hence, this study is motivated to reach a common view for automated mapping with CIDOC CRM and to capture and represent the underlying semantic of Malay textile information for ease of integration and exchange between communities in e-museum applications.

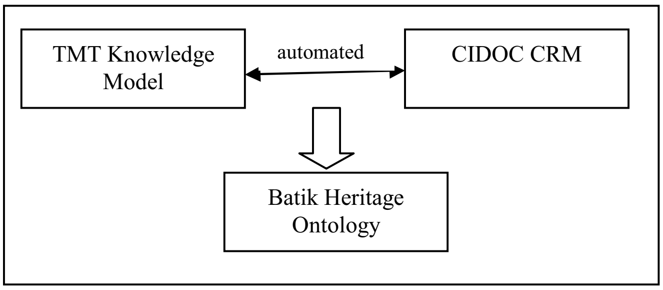

Fig 1. Motivation: Towards creation of Batik Heritage Ontology

D. Transformation Process

Initially, TMT Knowledge Model is developed using Protégé software. It consists of six concepts and thirteen subconcepts. Based on the examples of the harmonization works, this study will adapt both works. Consequently, TMT-Knowledge Model (see Fig. 2) will undergo a refinement process to capture its concepts and properties. In this work, all facets in the model are re-created by transforming concepts and subconcepts into RDFS classes and properties.

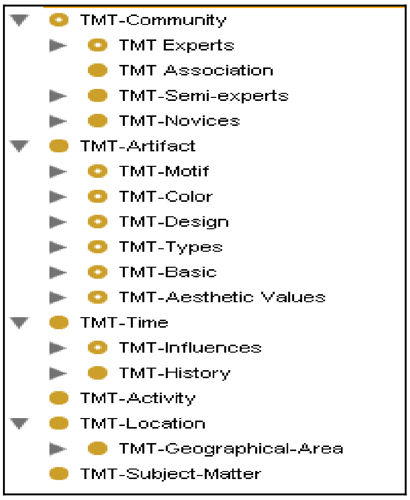

Fig. 2Snapshots of TMT concepts and subconcepts

Due to the current state of CIDOC CRM classes which is limited in describing TMT-Artifact category, therefore, this category will be further improved to capture the details of the textile making. As a result, the model is extended by classifying the artifact into three categories which are Aesthetics Value, Materialization and Production. The Aesthetics Value category defines any value that causes an object to be 'a work of art'. Materialization category is further divided into sub-categories describing Material (*e.g. brush, block, chanting, cloth, etc*), Motif (*e.g. traditionally-based on flora & fauna, dongson etc*) and Color (*e.g. chemical dye, natural dye, combination, etc*). Production category depicts the actual textile practice concerning Design, Technique and Making-Process. Design sub-category describes the different layout of patterns, pattern types, the design structure and the type of workmanship. Technique sub-category involves normal practices and skills of producing the artifacts. Lastly, Making-Process details the actual making which depends on the design and technique used.

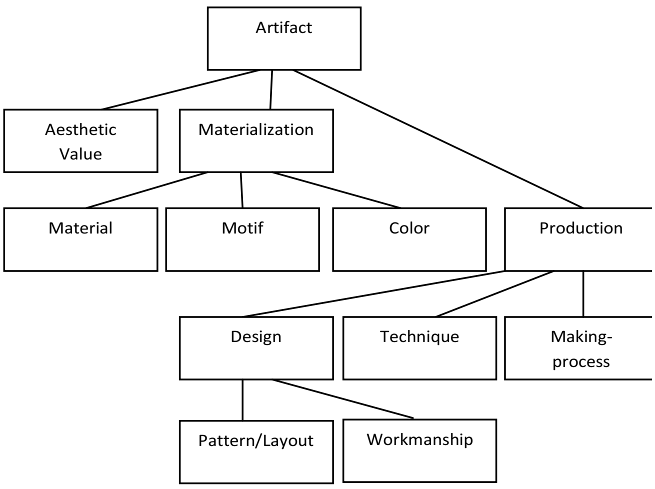

Fig.3 Classification of TMT:Artifact

IV. Results

This section presents the new version of TMT Knowledge Model where the primitive category is an *entity*. Fig. 4 shows the facets of the model which encompasses five main classes: *idea, existence, activity, time, location* and four subclasses: *handwork, community, artifact* and *subject* which are described below. Classes are shown in rectangles and sub-class relationships are indicated by solid lines. The modified TMT-Artifact categories as shown in Fig. 3 are also included in the model as an extension. All the classes in the model were given a name and an identifier consists of the word TMT constructed according to the conventions used in the CIDOC CRM model.

A. *Idea Class*

Subclass of: Entity
Superclass of: Handwork
Description:

The *Idea* category is created to enable the expression of concepts or ideas. It is used to express the notion of *Handwork* which exists through sensible way such as when it is told, demonstrated and shown in some was as a means to bind several *Materializations*. *Idea* is inspired by an instance of *Aesthetics Value* in which included in *Production*, which may be an instance of *Design*, *Technique* and *Making-Process*.

B. *Existence Class*

Subclass of: Entity
Superclass of: Community, Artifact, Subject
Description:

The *Existence* is similar to entity Persistent Item (E77) which stands in contrast to the *Idea* category. It expresses tangible cultural heritage which refers to something that remains intact to see, hold and is movable and normally preserved in its original form. TMT expresses this notion through the *inPlace* that associates *Existence* with *Location*.

C. *Activity Class*

Subclass of: Entity
Description:

The *Activity* is equivalent to entity Activity (E7) which involves *Community* in the context of *Production*. TMT expresses this notion through *hasInvolve* that associates *Activity* with *Community*.

D. *Time Class*

Subclass of: Entity
Description:

The *Time* is the same as entity Time-Span (E52) which represents a time span or point in time. The *influence* property binds the *Time* and the *Activity* in which activities are being influenced by some period of time.

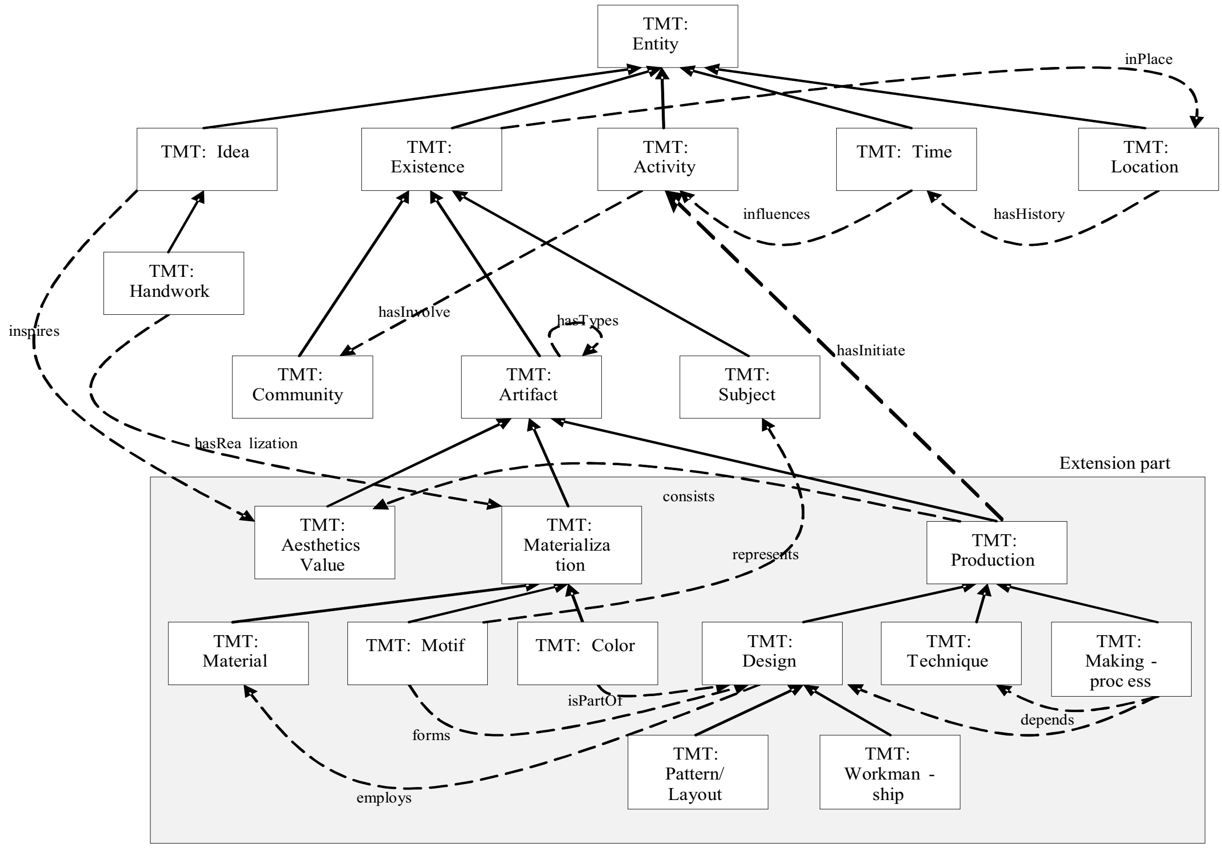

Fig. 4 The illustration of TMT Classes and Properties

E. Location Class

Subclass of: Entity

Description:

The *Location* is similar to entity Place (E53) which represents the origin or place where the artifacts can be found. The *hasHistory* property associates *Location* with *Time* to describe the actual existence of artifacts.

F. Handwork Subclass

Subclass of: Idea

Description:

The *Handwork* is a conceptual notion which only disclosed when it has been actualized through *hasRealization* property in some *Materialization.*

G. Community Subclass

Subclass of: Existence

Description:

The *Community* participates during an *Activity* which may be experts, semi-experts, novices, group from cultural institutions, etc.

H. Artifact Subclass

Subclass of: Existence

Superclass of: Aesthetics Value, Materialization, Production

Description:

The *Artifact* expresses logical entities that are tangible realization of concepts and can be manifested in many ways such as through *Aesthetics Value, Materialization* and *Production*.

I. Subject Subclass

Subclass of: Existence

Description:

The *Subject* refers to actual things that exist and symbolized by the *Motif* in the *Design* through the r*epresents* property.

V. Future Work and Conclusions

With the transformation of TMT Knowledge Model, the next step is to test, evaluate and refine it by using the real collections of cultural artifacts within museums. Digitizing the heterogeneity of cultural heritage collections is a major challenge for ontology developers in order to support the demand of online applications. For example, the nature and type of collections may consist of text, written on different materials, paintings, photographs, 3D objects, sound recording or even maps [7]. Similarly, the scope of artifacts covered in TMT may vary widely and the chance to discuss or mention the same objects is relatively small and sometimes limited. Therefore, mapping TMT with the upper ontology, CIDOC-CRM could resolve the problem of disagreements and misinterpretation of these artifacts. Hence, a future goal is to construct Batik Heritage Ontology through automated mapping between TMT Knowledge Model and CIDOC CRM in order to preserve batik-related information.

To conclude, this research intents to provide a platform for the museum community to enrich knowledge on cultural heritage artifacts and deploy an e-museum design pattern through community based e-museum. For that reason it could be implemented in any museum in Malaysia. Once the knowledge structure is constructed, the community could promptly access cultural heritage knowledge as well as sustain ongoing discussions on particular topics of interest. In addition, community based e-museum may also create opportunities to enhance user interest, new exciting experiences and to experiment with new strategies for user engagement that will help to build the community and attract new web audiences.

ACKNOWLEDGEMENT

Special thanks to Martin Doer for his suggestions and recommendations in this study. He is a senior researcher in the Institute of Computer Science, Foundation for Research and Technology – Hellas (ICS/FORTH), Greece. He has been involved in most of the harmonization projects involving CIDOC CRM.

REFERENCES

[1] Baclawski, K. and Niu, T., *Ontologies for bioinformatics*: MIT Press, 2006.

[2] Baker, P. G., Goble, C. A., Bechhofer, S., Paton, N. W., Stevens, R., and Brass, A., "An ontology for bioinformatics applications." vol. 15: Oxford Univ Press, 1999, pp. 510-520.

[3] Bao, H., Liu, H., Yu, J., and Xu, H., "An Ontology-Based Semantic Integration for Digital Museums," in *LNCS 3739*, Z. W. W. Fan, and J. Yang Ed.: © Springer-Verlag Berlin Heidelberg, 2005, pp. 626-631.

[4] Bekiari, C., Doerr, M., Le Boeuf, P., Aalberg, T., Barthélémy, J., Boutard, G., Görz, G., Iorizzo, D., Jacob, M., and Lamsfus, C., "International Working Group on FRBR and CIDOC CRM Harmonisation," *FRBR Object-Oriented Definition and Mapping to FRBR ER.(version 0.9 draft),* 2008.

[5] Binding, C., May, K., and Tudhope, D., "Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM," Springer, 2008, pp. 280-290.

[6] Bodenreider, O., "The unified medical language system (UMLS): integrating biomedical terminology." vol. 32: Oxford Univ Press, 2004, p. D267.

[7] Choi, N., Song, I. Y., and Han, H., "A survey on ontology mapping," *ACM SIGMOD Record,* vol. 35, pp. 34-41, 2006.

[8] Damljanovic, D. and Devedzic, V., "Applying semantic web to e-tourism," Information Science Publishing, 2008, p. 243.

[9] Ding, Y. and Foo, S., "Ontology research and development. Part 2-a review of ontology mapping and evolving," *Journal of Information Science,* vol. 28, p. 375, 2002b.

[10] Do, H. H. and Rahm, E., "Coma: A system for flexible combination of schema matching approaches," VLDB Endowment, 2002, pp. 610-621.

[11] Doan, A. H., Madhavan, J., Domingos, P., and Halevy, A., "Ontology matching: A machine learning approach," Springer, 2004, pp. 397-416.

[12] Doerr, M. and LeBoeuf, P., "Modelling intellectual processes: the FRBR-CRM harmonization," *LECTURE NOTES IN COMPUTER SCIENCE,* vol. 4877, p. 114, 2007.

[13] Doerr, M., Hunter, J., and Lagoze, C., "Towards a Core Ontology for Information Integration," *Journal of Digital Information,* vol. 4, p. 169, 2003.

[14] Eide, O., Felicetti, A., Ore, C. E., Andrea, A. D., and Holmen, J., "Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping," *Paper presented at the EPOCH Conference on Open Digital Cultural Heritage Systems.,* 2008.

[15] Euzenat, J. and Shvaiko, P., *Ontology Matching*: Springer-Verlag Berlin Heidelberg New York, 2007.

[16] Giunchiglia, F., Shvaiko, P., and Yatskevich, M., "S-match: an algorithm and an implementation of semantic matching," Springer, 2004, pp. 61-75.

[17] Hunter, J., "Combining the CIDOC CRM and MPEG-7 to Describe Multimedia in Museums," *Museums on the Web, Boston, April,* 2002.

[18] Ismail, S. Z., "Tekstil Melayu : Sejarah Fungsi Simbol dan Keindahan," in *Malaysia dari Segi Sejarah*. vol. 18, 1990, pp. 50-66.

[19] Jiayi, P., Cheng, C. P. J., Lau, G. T., and Law, K. H., "Utilizing Statistical Semantic Similarity Techniques for Ontology Mapping—with Applications to AEC Standard Models." vol. 13: Elsevier, 2008, pp. 217-222.

[20] Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., and Gergatsoulis, M., "Integrating Dublin Core metadata for cultural heritage collections using ontologies," 2007, pp. 27-31.

[21] Lagoze, C. and Hunter, J., "The ABC Ontology and Model," *Journal of Digital Information,* vol. 2, pp. 2001-11, 2001.

[22] Legrand, B., "Semantic Web Methodologies and Tools for Intra-european Sustainable Tourism," 2004.

[23] Liu, H.-Z., "Global Ontology Construction for Heterogeneous Digital Museums," *Paper presented at the Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong,* 2007.

[24] Marques, D., "A Survey of Recent Research on Ontology Mapping," Simon Fraser University, School of Interactive Arts & Technology, 2005.

[25] Noy, N. F., "Semantic integration: a survey of ontology-based approaches," *ACM SIGMOD Record,* vol. 33, pp. 65-70, 2004.

[26] Noy, N. F. and Musen, M. A., "Anchor-PROMPT: Using non-local context for semantic matching," 2001, pp. 63-70.

[27] Oltramari, A., Gangemi, A., Guarino, N., and Masolo, C., "Restructuring WordNet's top-level: The OntoClean approach," *LREC2002, Las Palmas, Spain,* 2002.

[28] Pisanelli, D. M., *Ontologies in medicine*: IOS Press, 2004.

[29] Signore, O., "Ontology Driven Access to Museum Information," *Paper presented at the Presented at the Annual Conference of CIDOC,* 2005.

[30] Sinclair, P., Goodall, S., Lewis, P. H., Martinez, K., and Addis, M. J., "Concept browsing for Multimedia Retrieval in the SCULPTUER Project," *Paper presented at the 2nd European Semantic Web Conference.,* 2005.

[31] Siti Zainon, I., *Malay Woven Textiles: The Beauty of a Classic Art Form*: Dewan Bahasa dan Pustaka, 1997.

[32] Suriyati, R., Norlaila, M. N., and Wan Adilah, W. A., "Towards Sustainable Heritage Information System: Conceptualization of Community based e-Museum.," in *International Conference on e-Commerce, e-Administration, e-Society and e-Communication (e-CASE)*, 2009.