Humanities, Arts and Social Sciences Studies

Research Article

Humanities, Arts and Social Sciences Studies
https://so02.tci-thaijo.org/index.php/hasss
ISSN (Online): 2630-0079

DEVELOPMENT OF THAI CULTURE ONTOLOGY-BASED METADATA: HEET SIB SONG (TWELVE MONTHS FESTIVAL)

Wirapong Chansanam and Kulthida Tuamsuk*

Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand

ABSTRACT

Semantic web technologies are rapidly changing digital humanities research. Over the last decades, an immense amount of new qualitative data have been accumulated and made available in interchangeable formats, in social sciences and humanities, opening new possibilities for solving old questions and posing new ones. Archiving, storing, linking, and reviewing data from several different sites, called semantical inclusion, is one of the biggest challenges for digital humanities research. Semantic integration is a mechanism by which the data, along with their associations and ontology, are conceptually interpreted to remove potential heterogeneity. For interface design, Ontologies may be used to map and merge metadata from various sources. This paper introduces a framework that eases scholars' access to historical and cultural data about the local wisdom and customs in the Northeast of Thailand that are distributed across different data sources. This paper aims to discuss metadata and ontology in semantical incorporation of Thai culture research from different datasets. "Heet Sib Song", a traditional twelve-month festival well-known in northeastern Thailand, is the regional Thai culture this study is concerned with. As a computational representation of the Heet Sib Song domain, we use Ontology Development 101 to provide semanticized metadata aggregation from the various databases. In further research, we will use the resulting ontology to create a semantic search system that can be used in the future. The digital content management method will be used in combination with the cultural heritage database system of the Digital Humanities Research Group, Khon Kaen University.

Keywords: Ontology; metadata; Thai culture; Heet Sib Song; semantic integration

Corresponding author:
Wirapong Chansanam
wirach@kku.ac.th

Received: *25 June 2021*
Revised: *15 November 2021*
Accepted: *16 November 2021*
Published: *23 December 2022*

Citation:
Chansanam, W. and Tuamsuk, K. (2022). Development of Thai culture ontology-based metadata: Heet Sib Song (Twelve Months Festival). Humanities, Arts and Social Sciences Studies 22(3): 725-737.

1. INTRODUCTION

People in the northeastern region of Thailand still follow the customs exclusively from the ancient days. Tradition has been a common practice for hundreds of years. The story behind Heet Sib Song-Klong Sib See derived from a religious belief that possibly originated in Indian culture (Pintong, 1991). In brief, the story was called "Perng Baan Perng Muang" or, locally, "Heet-Klong" which means "to be brought into society," and the story of Heet Sib Song-Klong Sib See can be understood differently by various elders. Nevertheless, they are all a bit strange. The variants were mainly memorized and remembered when the writing or recording methods

had a shortcoming. A few such stories have been published and widely distributed. 'Heet' meaning 'normal practice' is abbreviated from the Thai word 'Ja Reet', but people in the Northeast pronounce the 'H' sound instead of the Thai way of pronouncing it with the 'R' sound. Hence, Heet Sib Song ('Sib Song' meaning the number twelve) is a procedure generally done throughout the twelve months of a year. In ancient times, the first month, called "Duan Ai", was the start of the new year, and the twelfth month, called "Duan Jiang", finished the year. The different events or practices of each month were mostly based on Buddhism. However, if a practice was originally based on a religion or faith other than Buddhism, people deliberately sought to associate it with Buddhism so that they could participate in merit-making rituals in temples (Pintong, 1991).

In the past, not only was it a year-round festival that was considered as the way of life which emphasized collectivism and the importance of community spirit through participation in group activities, but it was also deemed to be a set of social standards that, when followed, led to peace in society. The idea behind this tradition is that everyone should respect and observe Thai culture in order to foster a peaceful and harmonious social environment. To elaborate further, Heet Sib Song and Klong Sib See are established and agreed-upon conventions or social rules for community members to abide by, and they serve to define one's duty in society with the assumption that every community that adheres to Heet Sib Song and Klong Sib See will ultimately bring about peace for all. Therefore, the storage of knowledge is important in preserving the Heet Sib Song tradition. According to Heet-Klong, safeguarding and transmitting the correct traditional practices is the identity of the northeastern region. It is a guideline for people in society to live by, maintain, and pass down to future generations.

Researchers in digital humanities sciences operate either independently or in study teams. Raw data obtained in field studies is usually processed in electronic formats to take advantage of the science computer's efficient computing capability. To carry out unbiased research or review experimental observations, it is also important to share helpful knowledge between various researchers or research groups. Increasingly, distributed research computing, including GRID-based architectures, is now being developed (Bramley et al., 2000). The Internet is an essential forum for the dissemination of information on cultural heritage. However, before smooth interoperability and sharing can take place, specific issues do have to be overcome. The main aspect can be established due to the diversity of operating systems and data storage environments. To carry out their study, separate research groups depend on numerous technological facilities. The best way to choose the correct answer to their particular sample is to make those improvements. However, this lack of continuity has also determined how various heterogeneous data sources can be combined.

The use of metadata and ontologies is a promising approach. Metadata is information that captures instance data characteristics from data sources (Vaduva and Vetterli, 2001), possibly including data format and structure, organization, and the conceptual context of the instance. Metadata is used to locate information, interpret information, and integrate/transform data. An ontology is an abstract definition of the conceptualization of a system (Fensel et al., 2001). Ontologies enable domain experts, creators of frameworks, and software developers/programmers to think about the nature of domains. An ontology can include standardized information models and electronic vocabulary (e.g., Getty vocabulary). Furthermore, ontologies allow knowledge-based applications to be developed. The benefits of an ontology include: facilitation of systems for knowledge sharing and reuse, facilitation of the acquisition of new knowledge, and improvement of confirmation and authentication of knowledge-based systems.

The Digital Humanities Research Group: DHRG (2020) has continuously focused on the convergence of different datasets containing information on cultural heritage obtained during ten years of research studies. The types of data differ from relational databases, flat text, and obsolete files. The group's digital humanities research focuses on the interpretations of concepts/conditions that form the basis of its study in information science, given by different academies, such as the Princess Maha Chakri Sirindhorn Anthropology Centre (SAC), Thailand's National Electronics and Computer Technology Center (NECTEC), and Getty vocabularies, etc. These fundamental concepts provided by different academies are used to assess study performance, promote the organizational quality enhancement, meet regulatory criteria, and supply as well as use national comparative data specifications on an individual basis. However, the data executing the Thai custom technology ("Heet Sib Song") by utilizing the standard World Wide Web technology explanatory frame may not include a management system to semantically scan for knowledge that displays the Thai custom in various dimensions.

Our study aimed to investigate how to facilitate access to these data sources through a centralized repository and analysis system. Thus, we have developed the ontology to identify and define accessible data sources in the digital humanities research context as a metadata management architecture. This makes it easier to map internally and standardize external data representation. Such a method will significantly benefit DHRG for processing data and the search/recovery of standard terms and meanings. We also developed an Ontology-based Thai Culture Data Integration Project (TCDIP) to promote the whole process, which is excellent for idea

recovery and data management. Semantic translation into the DHRG corpus is achieved through definitions from heterogeneous sources.

2. PROBLEM STATEMENTS

Most digital humanities scientists generally disregard the underlying data context and tend to collect and interpret information using common terms from the cultural heritage vocabulary. The knowledge access mechanism can convert user requests into single database queries described in the field-wide vocabulary of cultural heritage. This means that a metadata management function is necessary to organize the translation, retain mappings of vocabulary and data source representations, and keep the mapping information correct since the underlying data source will alter or be changed over a while.

Data integration is a process to ensure consistent and transparent access to multi-database information (Calvanese et al., 1998; Lee et al., 1999). In general, the metadata management component provides three types of data integration: physical, logical, and semantic. Physical integration transforms heterogeneous data sources (such as eXtended Markup Language) into a standard data format. The task of logical integration is to connect all data to standard process models. Finally, semantic integration enables data in a common vocabulary or ontology to be cross-referenced and sometimes inferred. In this research, we are interested in the conceptual and semantic dimensions of combining data from multiple datasets into a standard method model to assist the study of Thai culture.

The metadata criteria in a research setting for Thai culture include:

• Define content-independent information (for example, location of tangible documents), domain information (for example, arrangement or configuration of electronic records for cultural heritage), and domain-specific information (for example, local tradition). Often defined is the level of abstraction at which the knowledge is interpreted.

• Describe mapping and relation between source attributes and words in a particular domain language (e.g., "the data form of the word tradition in the field of cultural heritage in a database record has coded meaning in the structured cultural heritage lexicon").

• Define the rules for reading knowledge in a particular context ("traditional customs and mores generally have the same meaning").

• Capture/record/calculate machine-readable information.

• Enable fast retrieval through an index or hierarchical structure during processing.

The standard metadata managing of cultural heritage database systems such as DHRG has been performed by manually scanning and typing the conceptions needed in the database. Consequently, typographical mistakes, duplicate records, and nominal values have been saved in the data archive. In previous studies, such mistakes cause contradictions and incompleteness. Therefore, it is imperative to collect such words and significances of cultural heritage from various sources and store them locally. These are the specifications that must be taken into account.

Furthermore, to avoid inserting redundant information, the framework should be sufficiently scalable to recognize existing concepts. It can also encourage the end-user to incorporate different meanings for the exact meaning of cultural heritage, i.e., to combine terms from various contexts to be used for analysis. Finally, the code must be built so that the original DHRG data structure is not changed and that the meanings can be converted into a DHRG-locally based structure.

3. RELATED WORK

In recent years, Information Systems has been the field that has been most involved in the development and the usage of ontology (Chansanam et al., 2020). Several terminologies in various subdomains (e.g., beliefs, ethnics, artisanal craft, performing arts, palm leaves, textiles) are established for different purposes (literacy indexing and retrieval, tangible electronic record, statistic relationships reporting). Various organizations have developed these terms (UNESCO, Files in human relations: HRAF Yale, SAC, etc.). Each vocabulary has its world-representation, appropriate for its purposes.

Ontologies are considered a core component of interoperability, collaboration, and knowledge brokering terminology (Amann et al., 2002). Noy and McGuinness (2001) define knowledge engineering as a program of domain knowledge contents to create ontologies. Van Zyl and Corbett (2000) lay down a structure that enables several ontologies and their instruments to be compared in various applications. In implementing the schema for multiple purposes, including cultural patrimony knowledge systems, incorporation and reuse of ontologies can be analyzed. One example of continuing attempts in establishing a detailed structured

documentation glossary of the world heritage knowledge exchange terminology electronically is the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (1996). A broad range of technology systems has used the UNESCO information tools to address recovery problems created by terminological gaps and distribute specific data through several databases.

Current metadata research focuses on data merging through understanding the structures used by domain ontologies for the underlying data sources. However, the approach to cultural heritage information systems is relatively new to metadata. The latest developments in the convergence and the sharing of cultural heritage knowledge rely on federated database structures, for example, in the study on folktales (Tuamsuk et al., 2016; Tuamsuk et al., 2018), or on data management methods, for instance, in the study on folksongs (Kaewboonma and Tuamsuk, 2018). These methods rely on a fixed record or data storage system to make the source data a global system. The only downside is that it cannot be used with periodic complex scheme changes since the integration process must be replicated if schemes change.

Furthermore, autonomy is often compromised to settle procedural disagreement when databases have to expose or alter the scheme to converge all the knowledge in their logical schemas. Users' right to render queries depending on various domain-specific semantics is limited by a defined overall scheme as well (Halevy et al., 2003). Additionally, wall paintings (Hoaihongthong and Kwiecen, 2019), as well as Belief Knowledge-based system (Chansanam and Tuamsuk, 2015) are among other database inclusion initiatives focused on metadata.

Data are shared among organizations, for testing and evaluation, in the field of digital humanities research. Several metadata standardization programs were implemented to promote interoperability and encourage improved connectivity. The Getty Knowledge Institute acknowledged the role of systematic vocabulary in identifying and cataloging cultural heritage resources for an extended period. The Art & Architecture Thesaurus (AAT), the Union List of Artist Names (ULAN), and the Getty Thesaurus of Geographic Names (TGN) are effective cataloging tools and information navigators – semantic "tracks" that allow users to locate important roads across the wide array of resources now accessible on electronic networks. They can be used as instruments to build a conceptual bridge that connects various cultural heritage materials regardless of where they exist or how they are represented (Lanzi, 1999). In order to promote the norms of sharing information between cultural heritage and local territories, the Getty Information Institute proposed establishing an XML metadata model in 1998 (Gill, 2008). The Code List for Organizations with Cultural Heritage provides a short alphabetical code describing library names and other forms with organizations needed in a bibliographical context (Clark, 2010).

For those allocating with Machine-Readable Cataloging (MARC) records, systems are reporting libraries, numerous borrowing systems, and those involved in regional, national or international cooperative projects. This list of codes is an essential reference tool. The institutional identities are named for with some data elements in MARC format. The primary data elements are the agencies that assign the record number, the entity producing or modifying documents, and the entity that keeps the object. This collection guides codes for keeping bodies covered by the Library of Congress National Union Catalog (NUC) and other Union archive publications containing collections for reporting bodies. In 1932, as part of a community project, new codes were created for cultural heritage organizations on an ongoing basis. Over time, a small number of existing codes have been updated or become obsolete. For any of the above relevant codes, references are given. The common usage of standard identifiers by school libraries at home (especially for government projects) and globally should contribute a high number of codes to the global exchange of knowledge on the Internet. Their use is growing continuously.

While this list of organizations, with more than 30,000 identified institutions, focuses on US institutions, it also includes codes for institutions in other countries that requested them. However, unless the organization is a branch of the United States organization, MARC codes are allocated to Canada, Germany, and the United Kingdom. There are more than 41,000 entries in this article. The codes and information on their history and structure are stored in a database where they can be searched for on the MARC Code List for Organizations website (Library of Congress, 2020). In addition, the information is available in the following section (the research methodology) of this paper. An important challenge is, therefore, the integration of data from various systems.

Related to these activities, we add a metadata element to the role of identifying a domain ontology's relationship with plans for the underlying sources of data. Our main contribution is to apply the metadata model, the database, and vocabulary used in Thai culture, namely Heet Sib Song (Festival of the Twelve Months).

4. RESEARCH METHODOLOGY

Now we explain our approach using an ontology-based framework for research into Thai culture.

4.1 Semantic knowledge-based understanding

In the field of cultural heritage, knowledge processing is becoming more and more critical. This can include: (1) the usage to consolidate the workflows of content management relating to cultural heritage, (2) attempts to establish data standards from different fields of cultural heritage for culture communications and indexing purposes; and (3) expanded use of digital resources and information systems that archive and store data relating to the cultural heritage of interest.

The implementation of ontologies, i.e., the conceptualization of a subject, or formal schemas, was one approach suggested for handling the complexity of data on cultural heritage. Ontologies benefit from the probability that various entities represented in separate data sources have the same typical semantical value from some areas of interest. As a result, requests about coherent information through the multiple data sources are feasible by asking questions about common ontology.

The ontologies of users/practitioners who are not necessarily knowledge engineers or logicians themselves are expected to be practical and not too complicated. Data sources can then appear like bases of knowledge, allowing the computer to access its semantic structures. However, this also overlaps, as multiple ontologies have similar meanings or potentially include sub-sets of different ontologies within themselves. Therefore, data incorporation and possible database activity must be logically and semantically defined and reorganized for the different ontologies. The concept of metadata is thus helpful for the treatment of data sources and their simple ontologies. This suggests that the ontology-based data sources need a structured metadata management model.

4.2 TCDIP data elements

The information science group will identify and exchange the data sources by recordings. In separate organizations, the lists of cultural heritage vary. Some organizations primarily concentrate on cultural heritage data, while some focus on local data, etc. In order to carry out high-quality research publishable and nationally appropriate, education organizations shall use data sources accepted and supported by the worldwide community.

In this case, we see a resource as a logical map for a variety of affiliated organizations. Resource recognition is unique to properties. The metadata manager's function is to maintain the mapping of the named resources at all levels and ensure that they are accurate. The resource monitor at each layer displays the resource status. Retrievals are carried out using the appropriate description method. The model outlines the framework and an essential aspect of the interoperability of the information supplied. Finally, the access management data demonstrates how information requests are treated based on the connections in a database.

Under the framework of TCDIP, structured definitions and terminology for cultural heritage are preserved for more efficient contact and sharing of information with other organizations. The services used by TCDIP at the DHRG are presented as follows.

DHRG Database: DHRG uses a code-based approach to describe the words and principles of cultural heritage. DHRG consists of a documentation table which provides code definitions and categories of data and the active state, generated by a new definition introduced into the DHRG framework. The code table incorporates the code with its details, categories, and active state. A code can be correlated with several definitions (Code Definition) and, therefore, library IDs. This enables the user to identify the source of a particular definition quickly. Each code will have its list value (Code List Value), and each file may have list values as well (File List Value). Moreover, the category of 'Cultural Heritage' contains coding that form part of specific groups, and so the references or Community Code Xref is provided within this table. Next, the Table of Alias Code includes the numerous aliases which can have passwords. Finally, DHRG uses the SAC, HRAF, and UNESCO databases, or libraries, to define concepts. Each library has specific library IDs in this table.

A HRAF Database: The Human Relations Area Files (HRAF) contains cultural information for education and research (Levinson, 1989). As part of the attempt to create a standardized cultural heritage thesaurus, the HRAF has identified 142 meanings divided into many categories. HRAF membership covers over 27,000 words in the US and worldwide.

SAC Database: A list of 421 meanings in the Princess Maha Chakri Sirindhorn Anthropology Centre (Princess Maha Chakri Sirindhorn Anthropology Centre, 2020) is subdivided into separate groups. The purpose of the SAC meanings is to standardize information storing in databases related to traditional cultural heritage and, in effect, to enhance the quality and practice of Thai culture.

CIDOC CRM: The CIDOC relationship data model of more than 400 tables, first evaluated in 1996, could be reduced to approximately 50 groups and 60 properties of much greater applicability than the initial schema. On the other hand, the more schemes are studied, the fewer improvements in the model are required. This model consists of 80 groups and 132 objects, which represent hundreds of schemas. In partnership with ISO, the CIDOC began the standardization process in 2000, following the original Relational Model's successful reformulation. In September 2006, the code was approved as ISO21127:2006.

We have created a Thai custom ontology called the Heet Sib Song ontology in the DHRG Data Repository to explain the meta-structure of the code knowledge. The ontology Heet Sib Song precisely reflects a common interpretation of the core ideas in the field of cultural heritage. They define the conceptualization of cultural heritage through formalism for reflecting information, which can be used and exchanged between various entities (DHRG, 2020). The Ontology of Heet Sib Song is a repository for all ideas relevant to cultural heritage. It compares the local term meaning for wisdom to the related meanings of UNESCO, SAC, and HRAF. In a sense, local ontology acts as a contextual ontology that facilitates the retrieval from diverse sources of cultural heritage vocabulary and meaning. Any interface question, therefore, asks for the local ontology first. The sampling class shall alert any new concepts and accordingly upgrade the local ontology. This improves service efficiency (since ontology is locally stored) and prevents any request from being linked to external sources. The Heet Sib Song ontology is set up using Protégé-5.5.0 (Protégé, 2003) and stored in RDF format (Brickley et al., 2014) based on Heet Sib Song metadata elements as shown in Table 1.

Table 1: The Metadata Elements of Heet Sib Song

No.	Elements	Meanings
1	Tradition name	The name used for the tradition
2	Local tradition name	Other names used to refer to that custom
3	Organized month	International calendar month names for traditional organizing
4	Time according to the lunar calendar	The period specified according to the format in the Thai lunar calendar
5	Objective	What the organizers want to appear according to expectations in organizing traditions
6	Activity	The model of the method of staging performances
7	Ritual	Model of practice patterns with objects, things, words, and chants or text used in the performances
8	Literature	Stories, fairy tales, fables, jataka, allegories, books, prose works, poetry, or writing
9	Belief	A strong faith in supernatural powers that control human destiny / Comments that are consistent with what is involved in any subject
10	Place	An area or areas used for holding a tradition
11	Ceremony	A person who performs the primary duty of a leader in the ceremony
12	Participant	People who take part in the ceremony
13	Equipment	Objects or things used to organize a tradition
14	Buildings	Something built for the tradition or that occurred after the tradition was held

Ontology has been established in several steps (Chansanam et al., 2014; Chansanam and Tuamsuk, 2015; Chansanam and Tuamsuk, 2016). The first step is to gather as much knowledge as possible about the subject. The second stage entails defining all ontology groupings and subclasses. The third move is to decide a class's and category's attributes. Properties are divided into two types: the properties of the objects and the properties of the records. Subject properties usually describe class or entity relationships. The data properties describe the relation between instances and values of data. Any property includes a jurisdiction and a selection. The fourth step is to determine the jurisdiction of each property and its reach. Comments can also have the classes and properties for the domain description. The fifth stage is structured to set up class instances and specify data and objects' properties to explain the relationship between class and subclass instances. Ontop Reasoner can be used to assess the exactness of ontology. We also use Ontop Reasoner plugins to check that Ontology is continuously evolving. Finally, the ontology for querying must be stored in RDF or OWL data format and exported to the necessary interface.

4.3 Metadata integration TCDIP semantics

The semantical model for metadata incorporation used in the DHRG TCDIP framework is defined in this section. The laying stack contains (a) the information layer, (b) the vocabulary domain layer, (c) the mapping layer, and (d) the context layer, which is a layer that is a structured solution to data processing.

A resource is an information processing interface that provides the simple physical representation of data within the context of metadata management. Objects are officially stored in the resource layer in R = {O, P, T} form, in which O is an entity representing resource R, P is a parent entity of O, and T is the resource type or scheme. The relationship between the parent and child hierarchy enables the construction of the tree structure to represent data sources, for example, XML Data Source containing < pid>1669</pid > > < pid >, World Wide Web Consortium (2003) expression as an object, Xpath expression to the entry node < festival > as a parent, XML schema part to the < pid > unit as a resource type (e.g., in an XML data source comprising: < pid>>1669</pid > as an item). This tuple structure is often used intuitively to model relational resources and object-oriented services.

A domain vocabulary is a set of declarations (S) which define concepts and links for those concepts in the ontology. A statement is formally articulated in Domain Vocabulary (S) in the triple S form (S = {sub, prop, val}) in which the sub is an assertion word, the prop is a relationship (a word) that represents a property of the subject, and val is a phrase or a literal one representing the property value. The three formats are adequate to represent both definition and relationships between them because sub, prop, and Val can be represented by statements that claim each declaration is given a specific namespace as an identity identifier.

Mappings take advantage of how data source elements are linked to domain vocabulary concepts and how domain vocabulary concepts are linked to other vocabulary concepts. There are two major classes, (1) mapping from data-source to the domain (resource-to-concept) vocabulary and (2) mapping from domain vocabulary to another domain (concept-to-concept). There are four instances to remember when translating resources-to-concepts:

1. To study a single concept, a single resource is used, and the (semantically equivalent) term is contained in one or more vocabulary.
2. To instantiate a set of concepts, a method is used, and the concepts are contained in one or more vocabulary.
3. Using a collection of tools to consider a particular idea.
4. In a component that comprises more than one element, each element will reflect the exact definition of semantic disjointness.

This mapping category is formalized accordingly. A resource-to-concept mapping is denoted as Mapping RC= {type, ER / AR, EC / AC}, where the element represents one of the above cases; ER as a collection of analogous resources (one or more), AR as an aggregate set of resources; EC as a set of corresponding concepts (one or more) and AC as a set of aggregating concepts. In the case of a single resource for a single concept map, only one resource and a concept (as instantiated) in EC should be available. If the arities are the same, we can use more than one property in ER and more than one term in EC in the same mapping. The list of single mappings is viewed as an ordered set.

The second mapping group is more difficult since a process of ontological integration is required. Noy and McGuinness (2001) suggested a system of this nature. This type of mapping is best defined here as additional statements for (mapping) that are stored in the domain vocabulary layer.

The context layer holds the rules for deciding when a mapping exists. This can be formally expressed as context rule CX = {E, C, A} where E is one or more occurrences to be observed, C is the set of one and more conditions to search for context, and A is the set of one or more acts to be taken if condition(s) are determined to be valid. This is the case. The definition of the Active Rules should be applied here. The meaning is described as a declarative sentence in an explicit rule:

ON some event e
IF some condition c
DO some action a

In this case, a rule is understood to perform any allowed acts only if and when a series of criteria apply. The question is the main form of occurrence to be tracked. Notice that the Current Regulations can be nested to trigger the following rules by a case: (1) the context (e.g., identifying whether or not the source is active), (2) the semantical position (e.g., the local tradition is a subject of the query because it only considers the cultural heritage terminology, unless specified otherwise) and (3) the contextual situations (e.g., current sources of information). Actions in this field involve operations which create an overview of the global scheme by activating the mappings allowed under previously assessed conditions.

4.4 The TCDIP architecture

The ontology-based Thai Culture Data Integration Project (TCDIP) is developed to provide exact access to heterogeneous data sources such as Thai cultural books, cultural heritage information files, and qualitative research databases. The metadata management aspect incorporates and organizes data sources by maintaining conceptual and semantic knowledge of cultural heritage and sources, capturing connections between data source attributes through words from a cultural heritage language, and qualitative computing information from these experiences and other resource-relevant data. Contextual information helps the retrieval mechanism produce questions from heterogeneous sources and provide data that fit the researchers' needs. The architectural style is shown in Figure 1.

Figure 1: Steps for Ontology Building

The Content Provider shall decide which data sources should be consulted to fulfill a high-level user information requirement. This is determined by using the Information Index to search the metadata of data providers that match the criteria. The criteria are further developed based on qualitative awareness. The metadata is then sent to the database processor to deliver the low-level database to the specified data sources.

The Metadata Registry manages to maintain and handle the various types of metadata the service provider uses. The Service Provider calls the Metadata Registry when a customer demands information from the database processor. The Metadata Registry shall provide the Service Provider with the required services, services attributes for the request for the database, and any specific information about selecting and reviewing resource elements under the user's request conditions. The Metadata Registry also carries out surveys for the Terminology Information Base using optimization techniques.

The Resource Directory includes an accurate description of each database's contents. The data are contained in a regular meta-model. The attributes stored from various sources can be sorted using a list, making it more easily accessible. Information descriptions include relationships between parent/child attributes (object-oriented or XML data source) and table-column relation (ratio-data sources), a human-readable description attribute, and attribute-based query syntax awareness (possibly a reference to a stored SQL method or Xpath). The directory also keeps information separately from the text, including the host server address for the data source.

The Vocabulary Knowledge Base encompasses field-specific ontologies for semantics, which are essential for the practice of information technology meanings and terms. This encourages storage of various ontology models and uses the key to seek information from the user. It also benefits from gathering knowledge in that it allows for the use of a range of ontology models to specify the same domain definitions at multiple abstraction levels. Finally, for clear definitions, ontologies should be integrated from more than one domain model.

The Mappings Knowledge Base contains the maps of groups and properties in various ontologies and comparisons of attributes with the abstract meaning specified in the ontology in specific data sources. One property of a data source may typically be compared in one or more ontologies of properties. The manager can identify mapping relationships, but data mining techniques can be employed automatically to achieve mapping (Fowler and Martin 1997).

The Context Knowledgement Base stores information relating to the interpretation by specific request criteria of the data source attributes. This knowledge base contains rules on what resource directory constraints are required when compiling the query processor's metadata.

The references are based on the related ontologies' requests, mapping knowledge, and relationships between concepts. This allows parameters for the data source questions to be optimized to increase semantic precision. It also makes the system more adaptable because specific contextual rules can address various information requests.

We provide well-defined Application Programming Interfaces (APIs) in the metadata management section that allows other systems and applications to consume accessible metadata resources (Content Provider and Metadata Registry). This enables the TCDIP to be expanded beyond basic user query functionality.

5. THE TCDIP IMPLEMENTATION

5.1 Implementation environment

TCDIP is based on Linux Debian Ubuntu's operating systems. The primary technologies for TCDIP architecture are the PHP web interface distribution layer, the middle/ corporate Apache Jena Fuseki and MySQL layer, and the XML server layer. The RDF defines the Heet Sib Song ontology, and the Jena-1.4 approach is used to check for and update RDFs. The Web Server description embraces a TCDIP interface using HTTP apps to provide global application connectivity utilizing XML messages, such as web browsers. The TCDIP interface can be accessed globally via DHRG.

5.2 Performing queries on TCDIP

The TCDIP Query interface, as shown in Figure 2, enables the user to request a cultural heritage definition by presenting the query as an HTML type. The form components displayed on this page are as follows: Concept Text, which specifies user types for the definition, and Timeline, which shows that DHRG data repository definitions are allocated as well as the date they are triggered. Inactive meanings, along with the date of their inactivation, are still retained in the archive. The Timeline drop-down box enables the user to decide that only operational definitions are to be retrieved for the chosen year; definition libraries define the user as one or more libraries to find a description of a cultural heritage phrase. In the current implementation, DHRG slot libraries are DHRG Data Repository Definitions and Getty vocabulary for the Linked Open Data provider vocabulary. The user should use a particular data source for Getty vocabulary. The Relation to Metadata Management leads to adding, changing, and disabling code definitions (for information, see Section 5.3).

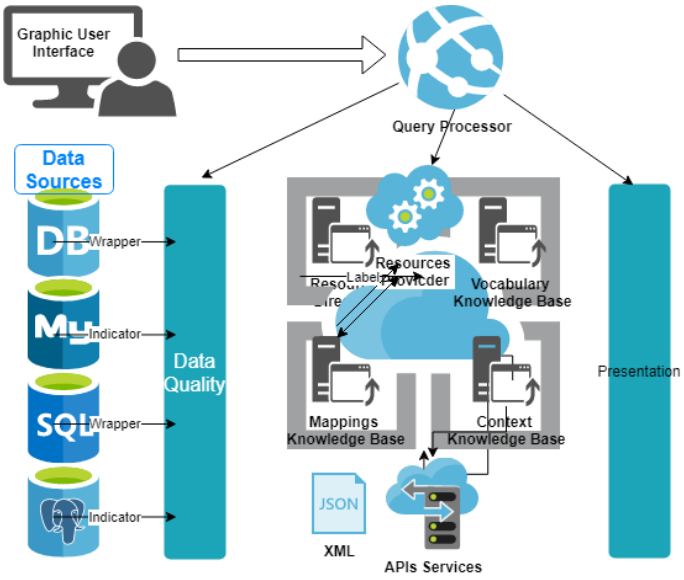

Figure 2: The TCDIP System Architecture

If the word entered by the user is not found in any of the specification repositories, it will be returned to the user's summary on the list of TCDIP results (Figure 2). The findings are usually given with the ID for the library description, which specifies the term's fundamental identity, followed by the classification of the source library of the description.

The name of the library description is the concept's name. The library's meaning is the interpretation or description of what the word is, how the word is applied correctly, and how it is defined. It provides specific details so that, for example, when the Month or Calendar mentions a category name, that name is considered part of the defined meaning. Various categories may have different names for the same meanings (i.e., synonyms) and may have different names for the terms. Getty vocabulary terms are linked via a semantic network. The words that serve a 'broader' or 'narrower than' partnership may be associated with each cultural heritage term. For instance, 'Festival' is a narrower term or specialization for the cultural heritage concept 'Tradition.' Similarly, 'Custom' is a broader term or generalization of 'Heritage'. The TCDIP gives the user a means of interpreting these associated words as shown in Figure 3. The outcomes of a question can also be presented, as seen in Figure 4 in XML format.

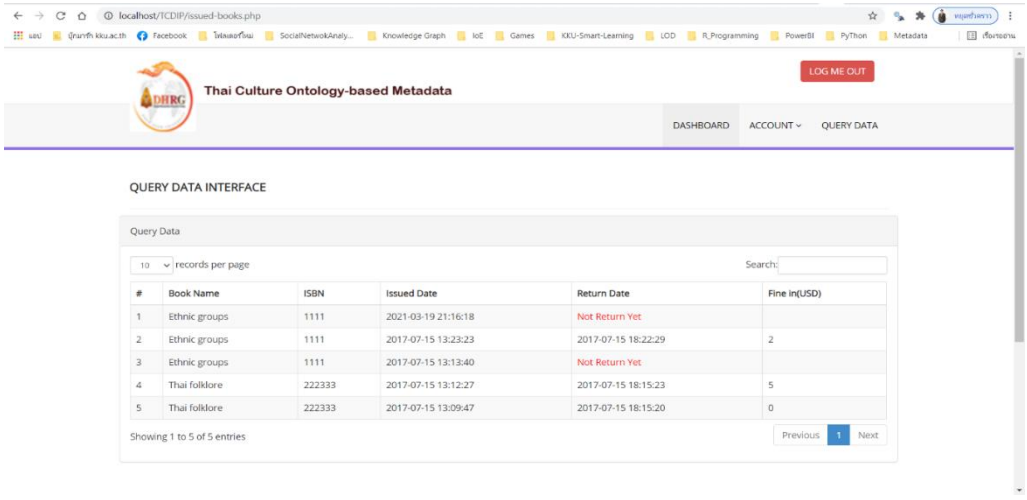

Figure 3: The TCDIP Query and Result Interface

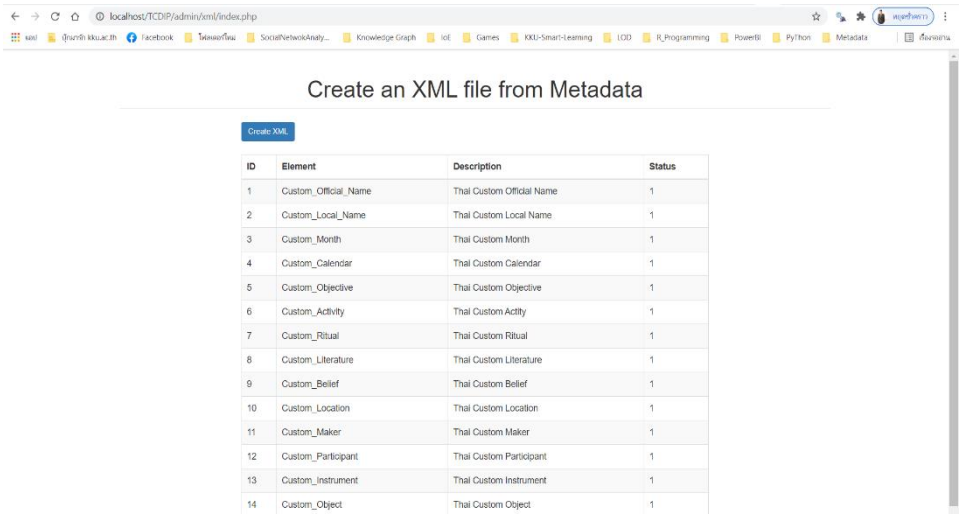

Figure 4: The Interface for the Creation of an XML File from Query

5.3 Metadata management

The metadata management framework (Figure 5) facilitates implementing a new concept to the DHRG data repository, the extension of a new meaning to an existing concept, and the introduction of list attributes to modify user interface types. In such instances, a word might not be included in the queried DHRG data repository. The user can use this concept in the DHRG data repository if a question returns a match from the UNESCO, SAC, and HRAF libraries. If the DHRG data repository already includes a term, the user can add additional concepts for UNESCO, SAC, and HRAF libraries.

The new coding is synonymous with introducing a new idea to the DHRG data repository. Upon showing a part of the Code Entry form, the user will enter a new design definition along with a brief description and allocate it to a DHRG code category. The definition can also refer to a code classification that requires the term or specification to be used by such attribute classes as a particular category. If the code is activated currently, it will appear with an active status value of 1, or 0 if it is inactive. The user can access the list of all codes currently disabled, modify or cancel changes, and cancel the code data. A collection is a set of values that can be viewed as user choices in a custom user interface. The user will add value to a predefined user interface list in the DHRG data repository using the metadata management module.

Figure 5: The TCDIP Query XML Result Interface

5.4 Use of the TCDIP system at DHRG

The TCDIP method aims to provide valuable cultural heritage services to DHRG anthropologists, historians, and statisticians. Specifically, the framework allows users to recover DHRG code meanings in their connection to the currently structured vocabulary of cultural heritage. Now, the TCDIP system is being implemented by the Digital Humanities Study Group (DHRG). The key stakeholders of the TCDIP system, as seen in Figure 6, are Information Science Scholar, Knowledge Engineer, System Administrator, and Cultural Heritage Information System. The following use cases support our statement that the TCDIP framework is broadly relevant for different stakeholders' digital humanities research community.

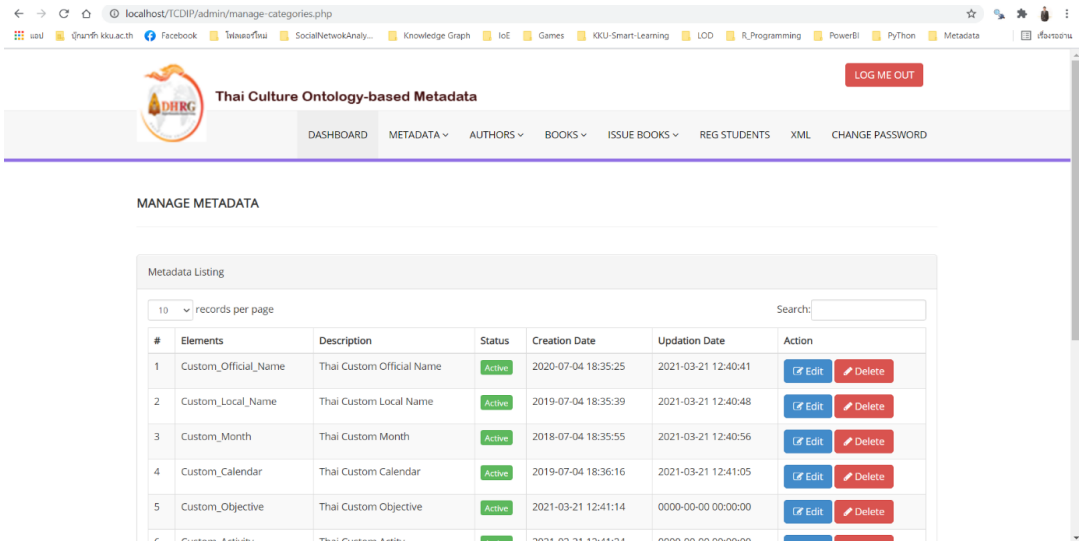

Figure 6: The TCDIP Metadata Management Interface

Information Science Researcher: The TCDIP system's key stakeholders are digital humanities researchers. Dr. Chansanam is writing a report on intangible cultural heritage studies for academic purposes. His thesis focuses on the paper produced by the DHRG database manager. The post, however, includes DHRG's particular definition for the word 'tradition'. He uses the TCDIP method to access the related ICD-10 code for the DHRG-specific definition of 'tradition' and explains DHRG's unique terms for the word 'tradition' which are referred to as 'I21' (International Disease) code. Corresponding meanings can be further derived from other outlets, such as SAC or Getty, through the TCDIP UI. This scenario reveals that the TCDIP system facilitates the retrieval by personalized mapping between internal representations and external representations of cultural heritage terminology.

Knowledge Engineer: Heet Sib Song ontology is developed and maintained by the Knowledge Engineer (KE). The KE also manages metadata via the TCDIP user interface. The KE defines routing rules for global schemes between local schemes and territorial network operations and network operations to another concept as part of metadata management. The understanding of metadata offers ample information on semantics and source characteristics. Moreover, the KE specifies guidelines for data conversion. For instance, the term 'local practice' taken from a data source in Getty vocabulary has a broader meaning, which in turn requires more than just the trading of words. It means that the data have a Getty vocabulary root and must be interpreted and preserved locally in the TRADITION hierarchy. The TCDIP framework helps them apply several meanings of the same medical term, i.e., combining terms from numerous sources semantically.

Database Administrator: The database administrator focuses primarily on building and modifying the TCDIP metadata database schema. In addition, the TCDIP method allows them to understand the meanings of data components and schemes underlying them.

Cultural Heritage Information Systems: Knowledge-based systems, cross-organizational, collaborative data science analysis, and cultural data storage are the main sets of cultural heritage applications systems that benefit from the TCDIP system. For example, as part of the Information Science Research Center (MEISEI University, 2020), the Digital Humanities Research Group works with different information science groups in Thailand. This partnership's main objective is to promote today's cultural heritage preservation and provide a language platform to facilitate and stimulate future academic research. In addition, the research aims to encourage seamless interoperability between various educational institutions with distinct semantical conceptions of the same area. For communication between these semantically diverse cultural organizations, the TCDIP system would play the role of language mediator.

6. CONCLUSION

A metadata management system based on ontology is introduced in this article, TCDIP, combining data from heterogeneous sources, such as the SAC, KKU Library, and Getty vocabulary. The ontology-based metadata management system enables the individual incorporation of various databases of cultural heritage. Furthermore, the system's automatic mechanisms for maintaining the DHRG database metadata are promising. The Digital Humanities Research Community introduced the TCDIP method, and the revised TCDIP model was tested. Furthermore, integrated presentation (user view) and data quality control (for data accuracy and integrity) are significant issues in the contemporary science testing environment. Therefore, the data quality monitoring and presentation elements are introduced in the TCDIP framework to improve information processing efficiency. This article indicated that organizing the Heet Sib Song traditional knowledge system can be used as a basic knowledge set to create an ontology.

ACKNOWLEDGEMENT

This research project was financial supported by the young researcher development project of Khon Kaen University.

REFERENCES

Amann, B., Beeri, C., Fundulaki, I. and Scholl, M. (2002). Ontology-Based Integration of XML Web Resources. In I. Horrocks and J. Hendler (Eds.), *The Semantic Web — ISWC 2002*, *Lecture Notes in Computer Science, Vol 2342*, pp. 117-131. Heidelberg, Berlin: Springer.

Bramley, R., Chiu, K., Diwan, S., Gannon, D., Govindaraju, M., Mukhi, N., Temko, B. and Yechuri, M. (2000). *A Component Based Services Architecture for Building Distributed Applications*. In I. Foster and P. Steenkiste (Chairs), *Proceedings the Ninth International Symposium on High-Performance Distributed Computing*, pp. 51-59. New York: IEEE.

Brickley, D., Guha, R. V., and McBride, B. (2014). *RDF Schema 1.1. W3C Recommendation*. [Online URL: http://www.w3.org/TR/rdf-schema/] accessed on January 30, 2021.

Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D. and Rosati, R. (1998). Description Logic Framework for Information Integration. In *Computational Logic: Logic Programming and Beyond. Lecture Notes in Computer Science*, edited by A. C. Kakas and F. Sadri, pp. 2-13. Heidelberg, Berlin: Springer.

Chansanam, W. and Tuamsuk, K. (2015). Development of an Imaginary Beings Knowledge Structure. *Lecture Notes in Computer Science* 9469: 291-293.

Chansanam, W. and Tuamsuk, K. (2016). Development of Imaginary Beings Ontology. In: A. Morishima, A. Rauber and C. Liew (Eds.), *Digital Libraries: Knowledge, Information, and Data in an Open Access Society. ICADL 2016, Lecture Notes in Computer Science, Vol 10075*, pp. 231-242. Cham: Springer.

Chansanam, W., Tuamsuk, K., Chaikhambung, J. and Sugimoto, S. (2020). Linked open data framework for ethnic groups in Thailand learning. *International Journal of Emerging Technologies in Learning* 15(10): 140-156.

Chansanam, W., Tuamsuk, K., Kwiecien, K., Ruangrajitpakorn, T. and Supnithi, T. (2014). Development of the Belief Culture Ontology and Its Application: Case Study of the GreaterMekong Subregion. In T. Supnithi, T. Yamaguchi, J. Pan, V. Wuwongse and M. Buranarach (Eds.), *Joint International Semantic Technology Conference*, pp. 297-310. Cham: Springer.

Clark, K. (2010). Values in Cultural Resource Management. In *Heritage Values in Contemporary Society*, edited by G. S. Smith, P. Messenger and H. Soderland, pp. 89-100. Walnut Creek CA: Left Coast Press.

Digital Humanities Research Group (DHRG). (2020). *About Us*. [Online URL: https://www.dhrg.huso-kku.org/?page_id=744] accessed on January 31, 2021. [in Thai]

Fensel, D., McGuiness, D. L., Schulten, E., Ng, W. K., Lim, G. P. and Yan, G. (2001). Ontologies and electronic commerce. *IEEE Intelligent Systems* 16(1): 8-14.

Fowler, J. and Martin, G. (1997). The Healthcare Administrator's Associate: An Experiment in Distributed Healthcare Information Systems. In D. R. Masys (Ed.), *Proceedings: A Conference of the American Medical Informatics Association. AMIA Fall Symposium*, pp. 548-552. Nashville, Tennessee: Hanley & Belfus, Inc.

Gill, T. (2008). Metadata and the Web. In *Introduction to Metadata,* edited by M. Baca, pp. 20-38. Los Angeles, CA: Getty Publications.

Halevy, A. Y., Ives, Z. G., Suciu, D. and Tatarinov, I. (2003). Schema Mediation in Peer Data Management Systems. In U. Dayal, K. Ramamritham, and T. M. Vijayaraman (Eds.), *Proceedings 19th International Conference on Data Engineering*, pp. 505-516. New York: IEEE.

Hoaihongthong, S. and Kwiecien, K. (2019). Defining metadata elements for murals. *Journal of Information Science* 37(2): 94-114.

Kaewboonma, N. and Tuamsuk, K. (2018). Ontology of folk songs in the Greater Mekong Subregion (GMS). *Knowledge Organization* 45(1): 33-42.

Lanzi, E. (1999). *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information*. Los Angeles, CA: Getty Publications.

Lee, M. L., Bressan, S., Goh, C. H. and Ramakrishnan, R. (1999). Integration of disparate information sources: A short survey. *ACM Multimedia*. 155-158.

Levinson, D. (1989). The human relations area files. *Reference Services Review* 17(3): 83-90.

Library of Congress. (2020). *Code List for Cultural Heritage Organizations*. [Online URL: https://id.loc.gov/vocabulary/organizations.html] accessed on April 15, 2021.

MEISEI University. (2020). *Information Science Research Center*. [Online URL: https://www.meisei-u.ac.jp/english/facilities/inf.html] accessed on January 31, 2021.

Noy, N. F. and McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. [Online URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf] accessed on January 31, 2021.

Pintong, P. (1991). *Traditional Thai Traditions*. Ubon Ratchathani: Offset printing. [in Thai]

Princess Maha Chakri Sirindhorn Anthropology Centre. (2020). *Database*. [Online URL: https://www.sac.or.th/portal/th/database] accessed on January 31, 2021.

Protégé. (2003): *The Protégé Project Website*. [Online URL: http://protege.stanford.edu/] accessed on January 30, 2003.

Tuamsuk, K., Chansanam, W. and Kaewboonma, N. (2018). Ontology of folktales in the Greater Mekong Subregion. *International Journal of Metadata, Semantics and Ontologies* 13(1): 57-67.

Tuamsuk, K., Kaewboonma, N., Chansanam, W. and Leopenwong, S. (2016). Taxonomy of folktales from the Greater Mekong Sub-region. *Knowledge Organization* 43(6): 431-439.

United Nations Educational Scientific and Cultural Organization (UNESCO). (1996). *Glossary of World Heritage Terms: A Glossary of Terms relating to the Implementation of the World Heritage Convention*. [Online URL: https://whc.unesco.org/archive/1996/whc-96-conf201-inf21e.pdf] accessed on January 31, 2021.

Vaduva, A. and Vetterli, T. (2001). Metadata management for data warehousing: An overview. *International Journal of Cooperative Information Systems* 10(3): 273-298.

Van Zyl, J. and Corbett, D. (2000). *A Framework for Comparing Methods for Using or Reusing Multiple Ontologies in an Application*. [Online URL: https://www.ukma.edu.ua/~gor/courses/AIUS2008/Ontologies/LingOntol.pdf] accessed on January 30, 2021.

World Wide Web Consortium. (2003). *Xpath Cover Page*. [Online URL: http://www.w3.org/TR/xpath/] accessed on January 30, 2021.