

Open Data Integration for Lanna Cultural Heritage e-Museums

Watchira Buranasing^(✉), Marut Buranarach, and Chai Wutiwivatchai

National Electronics and Computer Technology Center, National Science and
Technology Development Agency, Thailand Science Park, Klong Luang,
Pathum Thani 12120, Thailand

{watchira.buranasing,marut.buranarach,chai.
wutiwivatchai}@nectec.or.th

Abstract. Cultural heritage is the way of life. Using digital technology to manage the information of cultural heritage has become an important issue on the perseverance of cultural heritage. Several cultural heritage institutions have their own large databases with different metadata schemas. Linked Open Data is a mechanism of publishing structured data that allows metadata to be connected. This paper presents a methodology for integrating, converting Lanna cultural heritage metadata and linking data to external data sources. Our framework uses OAI-PMH (Open Archives Initiative—Protocol for Metadata Harvesting) by converting the various metadata formats to Dublin Core metadata standard. The Dublin Core/XML was used for the integrated metadata to create a central repository of data sources. The OAI harvester was developed to extract data from the various repositories and gather data into the repository. A framework for publishing the repository data as a Linked Data source in the RDF format using the OAM Framework is described. In this framework, ontology is used as a common schema for publishing the RDF data from the database sources by mapping database schema to ontology. Finally, we demonstrate Linked Data consumption by creating a cultural heritage portal application prototype from the published RDF data sources.

Keywords: Open Data · Cultural heritage · Ontology · e-Museum
Metadata crosswalk

1 Introduction

Cultural Heritage [1, 2] is the way of life of group of people. It can be seen in ways of behaving, beliefs, values, customs, languages and traditions that are passed on from one generation to the next. Cultural heritage includes tangible culture such as buildings, monuments, inscription, manuscript, landscapes, books, works of art and artifacts, and intangible culture such as folklore, traditions, language, and knowledge and natural heritage. Using digital technology to manage the information of cultural heritage has become an important issue on the perseverance of cultural heritage. Besides written

documents, drawings and paintings, cultural heritage also includes various media such as video, photo, object virtual reality, panoramic photo and audio. Several cultural heritage institutions have their own large databases with difference metadata schemas. Some e-museum projects that focused on linking metadata from various sites include Europeana, Museum Finland, Amsterdam Museum, Smithsonian Museum, Bangladesh Museum and LODAC Museum. These projects focused on designing a common framework using ontology or other common metadata models for integrating various types of media formats, subjects, and metadata standards. The Resource Description Framework (RDF) standard [3] is often used as the metadata interchange format.

This paper proposes a new approach for interoperability between the datasets from different institutions using OAI-PMH and Linked Data framework. The methodology covers data management, crosswalk metadata, metadata harvesting, conversion and linked open data publishing. The framework was conducted over e-museum systems related to Lanna (Northern Thailand) cultural heritage. The project was conducted in collaboration with Cultural Heritage institutions in north of Thailand, including museums, temples, local agents and Cultural Information center to publish their own data to Linked Open Data. The remainder of the paper is organized as follows. Section 2 gives an overview of the methodology for integration cultural database and cultural ontology. Section 3 shows an application prototype of linked Lanna e-museums. Section 4 concludes and discusses some future directions.

2 Related Work

Linked Open Data [4] is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources. In previous research, “Linked Open Data for Cultural Heritage: Evolution of an Information Technology”, Edelstein et al. [5] have made a survey on the landscape of linked open data projects in cultural heritage, examining the work of groups from around the world. Traditionally, linked open data has been ranked using the five star method proposed by Tim Berners-Lee. This research developed a six-stage life cycle based on the five-star method, describing both dataset development and dataset usage. It uses this framework to describe and evaluate fifteen linked open data projects in the realm of cultural heritage.

In [6], Constantia Kakali et al. have presented a method of ontology-based metadata integration using CIDOC/CRM ontology and a methodology for mapping Dublin Core elements to CIDOC/CRM ontology. In addition, “Supporting Linked Data Production for Cultural Heritage institutes: The Amsterdam Museum Case Study” [7] developed by Victor de Boer et al. present the methodology to convert the small cultural heritage as Amsterdam Museum metadata to a Linked Data version. “data.europeana.eu The Europeana Linked Open Data Pilot” [8] developed by Bernhard Haslhofer and Antoine Isaac produce the Linked Data version of Europeana and published the resulting datasets on the Web. “MUSEUM FINLAND-Finnish Museums

Table 1 A comparison of strategies to create linked open data

	Type	Format	Model/Ontology	Metadata	Tool	Interlinking
Amsterdam	Tangible	RDF			ClioPatria, Amalgamc	
Europeana	Tangible/Intangible	RDF	EDM	ESE		
Bangladesh	Tangible	RDF	EDM			
Finland	Tangible	RDF	CIDOC-CRM			
Smithsonian	Tangible	RDF	EDM		Karma	DBPedia, Wikipedia, Newyork time
LODAC	Tangible	RDF	CIDOC/CRM/LODAC	Dublin Cure		DBPedia
MAC-USP	Tangible	RDF	CIDOC-CRM			
NHMs	Tangible	RDF	EDM		MMAT,OAI- PMH	
Russian	Tangible		CIDOC-CRM			DBPedia, British Museum
Gothenburg	Tangible		proton, CIDOC-CRM			GeoNames, Dbpedia, CoL, UniproL
Thailanna.	Tangible/Intangible	RDF	Thailand Cultural experts	ISAD++, CDWA+, specific metadata	QAM, e-museum, OAI-PMH	Cultural Information center

on the Semantic Web” [9] developed by Eero Hyvönen et al. produce the semantic portal MUSEUM FINLAND for publishing heterogeneous museum collections on the semantic web. “Connecting the Smithsonian American Art Museum to the Linked Data Cloud” [10] developed by Pedro Szekely et al., present the process of publishing data from the Smithsonian American Art Museum (SAAM) and linking dataset to DBpedia and the Getty Vocabularies. “Linked Open Data Representation of Historical Heritage of Bangladesh” [11] developed by Shima Chakraborty et al. use semantic web technology for data management of historical heritage in Bangladesh, linking data to Geo-Bangladesh and using SPARQL to retrieve and inference specific information. “Sharing Cultural Heritage Information using Linked Open Data at a Museum of Contemporary Art” [12] developed by Erika Guetti Suca and Flávio Soares Corrêa da Silva present the architecture for sharing cultural heritage data based linked data for the Museum of Contemporary Art at the University of São Paulo (MAC-USP). Using RDF as a schema language for creating logical relationships among cultural heritage items. “Elevating Natural History Museum’s Cultural Collections to the Linked Data Cloud” [13] developed by Giannis Skevakis et al. present the architecture for transition the Natural History Museum repositories (Cultural heritage content is collected from six Natural History Museums around Europe) to the semantic web and methodology for converting metadata into Linked Data. “Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing” [14] developed by Dmitry Mouromtsev et al. present the architecture and methodology for publishing open linked data from Russian Museum by using CIDOC-CRM ontology and linking data to Dbpedia and British Museum. “Reasonable View of Linked Data for Cultural Heritage” [15] developed by Mariana Damova and Dana Dannells present an application for data integration based on semantic web technologies from the Gothenburg City Museum by using PROTON and CIDOC-CRM ontology and linking data with GeoNames, DBpedia, Catalogue of Life (CoL) and Uniprot.

The above-mentioned works proposed various approaches to create linked open data from data-base with the same metadata schema. However, they can be complementary to each other. As another research work, Matsumura et al. [16] collected information of museums from their websites and generated Linked Open Data of museum information in Japan. Table 1 compares different approaches of existing e-museum projects based on the LOD framework.

3 Our Framework for Integrating Lanna Cultural Datas for Linked Open Data Publishing

3.1 System Architecture Overview

One of the challenges of integrating cultural heritage data is system and metadata heterogeneity. Every cultural heritage institution has its own collection management system such as e-museum management system, manuscript management system, mural management system. Furthermore, they usually occupy their own metadata schedule.

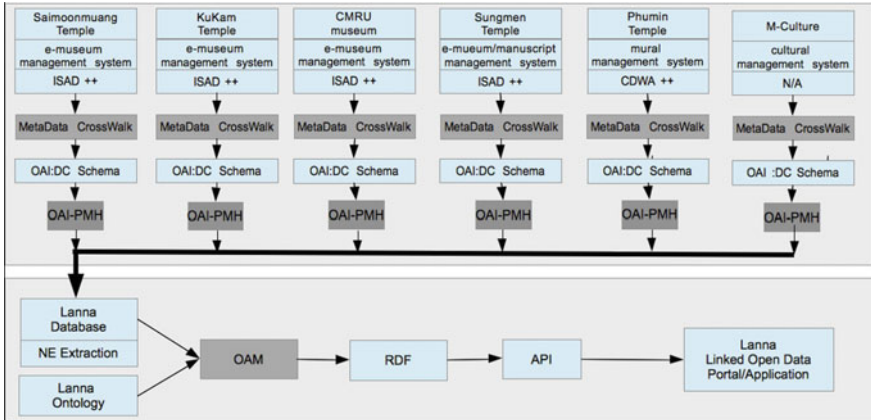


Fig. 1 Architecture and workflow for harvesting and converting to linked open data

Figure 1 shows our system architecture for integrating various e-museum systems of the Lanna culture. Our architecture is primarily based on OAI-PH and ontology frameworks. The interoperability and exchange of metadata is further facilitated by metadata crosswalks. This methodology crosswalks metadata from ISAD++, CDWA++ and specific metadata (cultural information center) dataset to Dublin Core metadata schema and export to a simple dublin core xml format. (OAI: DC schema). Using OAI-PMH to harvest data from xml. To extract information from the result and insert into the database, we map the extracted information with the cultural heritage ontology under the ontology-based application management (OAM) framework [17], then export them in the RDF format, and finally create a cultural heritage portal application by using such RDF data sources.

3.2 Lanna Cultural Data Management Systems

This section describes existing systems utilized in various Lanna cultural heritage sites. These systems are e-Museum, Manuscript and Mural management systems.

3.2.1 e-Museum Management System

Our e-Museum Management System is designed to aid users to archive and to catalog cultural heritage items in museums, temples, and cultural agents for digital exhibition and digital asset management. e-Museum Management System is based on ISAD (G) metadata standard and added some elements from cultural experts. This application allows for the upload, description, management and access of digital collections and also converts and exports OAI-DC xml format.

3.2.2 Manuscript Management System

Manuscript Management System is a web-based software application for the description, management, and dissemination of manuscript collections information and translation from local language to native language and also translation from native

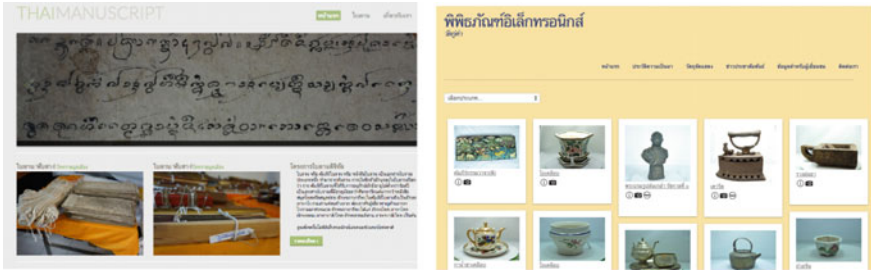


Fig. 2 e-Museum management system and manuscript management system

language to an interpretive approach. This application is based on ISAD (G) metadata standard and added some elements from manuscript expert interpreter. This application allows for the upload, description, management and access of manuscript digital collections (Fig. 2).

3.2.3 Mural Management System

Mural Management System is an application to administrate wall painting (mural) photo galleries. This application is based on CDWA metadata standard and added some elements from cultural expert. This application allows for the upload, description, management and access of manuscript digital collections.

3.2.4 Thailand Cultural Information center

Thailand Cultural Information center (<http://www.m-culture.in.th>) is a national cultural archive, which is one of an important database for education, economy and society. This archive was developed under the collaboration between Ministry of Culture and National Electronics and Computer Center. The content database is associated with person, organization, place and artifact. This archive is based on special metadata elements from cultural experts from Ministry of Culture and the content is approved by cultural agents from 76 provinces in Thailand (Fig. 3).



Fig. 3 Mural management system and Thailand cultural information center

3.3 Metadata Crosswalk and Open Archives Initiative Protocol for Metadata Harvesting

3.3.1 Metadata Crosswalk

A metadata crosswalk is a specification for mapping one metadata standard to another. A complete specified crosswalk consists of both a semantic mapping and a metadata conversion specification and all implementations of the crosswalk on a specific source content result in the same target content. Each archive has its own database with difference metadata schemas and the Dublin Core metadata schemes is a standard for cross-domain information resource description. The approach is to convert its own metadata schemas to Dublin core metadata schemas.

The following steps are mapping instructions for a *complete specified* crosswalk.

Step 1: extract terminology and properties. A lack of common terminology currently exists among the different metadata standards. Some of terminology has the same or nearly the same meaning. For example, ISAD metadata is identified using <identifier>, whereas Dublin Core metadata is identified using <reference code>. Besides, some of the metadata standards use similar properties in the definition of their metadata. The similarities need to be extracted and the concepts generalized and used in a common way across all metadata standards.

Step 2: element to element mapping including one-to-one and one-to-many transformation. Each metadata standard was built for a different propose and some elements were built from expert, so some of elements could match in Dublin Core, but there are some elements could not match in the other (Dublin Core) metadata standard. The importance of mapping method is extra elements to element mapping. For example, manuscript has <Pariwat> element (translate from the old northern *Thai language* to native Thai language), the solution to handle this problem is take <Pariwat> element maps to <description> element in Dublin Core metadata standard.

Step 3: content conversion and combination. Each metadata standard restricts the content format of each metadata element such as data type or range of values. It is important to convert between text and numeric values or text and date values. Some specific elements was built by cultural experts for example, Chula Sakarat (Minor Era) and Rattanakosin era is *necessary to convert to date format*, Besides, some of source element values is free text, that must be converted to same format such as *Anno Domini (A.D.) is necessary to convert to MM-DD-YY in Buddhist Era (B.E.)*. A general content conversion could map element to element with one value, but some element must be combinations of value for example, values of <source> elements could be a source of location and URL (Watkukam: www.emuseum.in.th/watkukam).

After the crosswalk change, the result of conversion is complete metadata elements and values, then export to OAI-DC XML file format for sharing. Table 2 shows the mapping of existing metadata schemes of the Lanna cultural systems to OAI-DC.

Table 2 Example of metadata crosswalk

	OAI-DC	Antique	Manuscript	Mural	Cultural Information center
	DC	ISAD++	ISAD++	CDWA++	N/A
1	Identifier	Reference code(s)	Reference code(s)	Unique id	id
2	Title	Title	Title	Title	Title
3	Date	Date	Date	Creation period	
4	Format	Extent	Extent	Work type	
				Classification	
5	Creator	Name of creator(s)	Name of creator(s)	Creator	
6	Description	Administrative/biographical history	Administrative/biographical history	Description	Description
		Archive history	Archive history	Shape	
		Finding aids	Finding aids	Dimension	
		Note	Note	Material	
		Archivist'sNote	Archivist'sNote	Technique	
		Date(s) of descriptions	Date(s) of descriptions	Culture	
		System of arrangement	System of arrangement	Inspiration	
		Appraisal, destruction and scheduling information	Appraisal, destruction and scheduling information	Inspiration	
			Burana phasa	Detail	
			Pariwat		

3.3.2 Open Archives Initiative Protocol for Metadata Harvesting

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a protocol developed by the Open Archives Initiative. It is used to harvest the metadata descriptions of the records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH representations metadata in Dublin Core. Based on OAI-PMH using Metadata Integration application applied from the PKP Metadata Harvester for centralized repository. Metadata can be harvested at any time, and as frequently as required. After the centralized repository is completed, then extract data from the result of repository using tag extraction. The method to extract the data find Dublin Core element tag patterns from the xml and insert into the MySQL Database. Figure 4 shows an example of the exported XML data of OAI-DC for OAI-PMH harvesting.

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<?xml-stylesheet href="http://www.openarchives.org/OAI/2.0/oai_dc.xsd" type="text/xsl" ?>
<?xml-namespace prefix="dc" uri="http://purl.org/dc/elements/1.1/" ?>
<?xml-namespace prefix="xsi" uri="http://www.w3.org/2001/XMLSchema-instance" ?>
<?xml-namespace prefix="oai" uri="http://www.openarchives.org/OAI/2.0/oai_dc/" ?>
<?xml-namespace prefix="xsi" uri="http://www.openarchives.org/OAI/2.0/oai_dc.xsd" ?>
<oai:metadata>
  <dc:title>ภาพวาด</dc:title>
  <dc:creator></dc:creator>
  <dc:subject></dc:subject>
  <dc:description>
    ภาพวาดที่ระลึกถวายงาน ที่งานพิธีเจ้าภาพถวายโคม (รูปคำ) ไม่ปรากฏปีสร้าง เป็นพิธี
    พิธีงาน พิธีหลายสมัย เหลือง เขียว เป็นภาพทาบ โคนกลางเป็นรูปโคม คงเป็นนักทาบ:เจ้าพิธี:
    ระทาบพิธี: เจ้า เป็นไม้ค้ำ
  </dc:description>
  <dc:description></dc:description>
  <dc:description></dc:description>
  <dc:publisher></dc:publisher>
  <dc:contributor></dc:contributor>
  <dc:date></dc:date>
  <dc:type></dc:type>
  <dc:format></dc:format>
  <dc:identifier>
    ภัณฑวัตถุคำ : www.emuseum.in.th/watkukan ; museum :
    http://emuseum.in.th/watkukan/emuseum/projects/objects/indexroom.php?
    objref=ภค-ภ-2557-0018
  </dc:identifier>
  <dc:source></dc:source>
  <dc:language></dc:language>
  <dc:relation></dc:relation>
  <dc:relation></dc:relation>
  <dc:relation>
    www.emuseum.in.th/watkukan/object/pic/thumb/thumb-DSC_1519.jpg
  </dc:relation>
  <dc:coverage></dc:coverage>
  <dc:rights></dc:rights>
</oai:dc:do>
</oai:metadata>

```

Fig. 4 Example of exported XML data of OAI-DC for OAI-PMH harvesting

3.4 Publishing Lanna Cultural Data as Linked Open Data

Although OAI-DC provides a common metadata representation using the Dublin Core standard, it has limited expressivity. Specifically, the DC elements are general and can not easily represent domain-specific metadata, e.g. paintings, museum objects, traditional archive, etc. Thus, domain-specific metadata representation is needed in addition to DC. In addition, Linked Open Data (LOD) is now accepted as an effective mechanism of data sharing to enable data integration based on RDF standard. Thus, we design a LOD data publishing framework that extends OAI-DC. In this framework,

data from DC elements can be further extracted for key concepts and relations. The extracted results can be mapped with the Lanna Cultural Heritage ontology and published as RDF data as a LOD data source. The ontology design approach and prototype system that utilizes the RDF data are described as follows.

3.4.1 Cultural Heritage Ontology Design and Mapping

Cultural heritage involves knowledge coherence and knowledge variety. Thus, it can be difficult to classify some knowledge into domains and design a complete cultural heritage ontology. Some of knowledge such as NOK-Hat-Sa-Dee-Ling (a legendary bird of Himavanta) can be classified into three domains: Antiques, Tradition and Ritual Instruments and tool. Therefore, it is important to limit the scope of the construction of domain ontology. Our initial ontology design focuses on three domains: antiques, manuscript and mural. The ontology model of cultural heritage was designed and evaluated with the involvement of cultural experts from Ministry of Culture, Thailand. The ontology was created in the OWL format (Web Ontology Language) using Hozo ontology editor. Figure 5 shows an initial design of Lanna Cultural Heritage ontology.

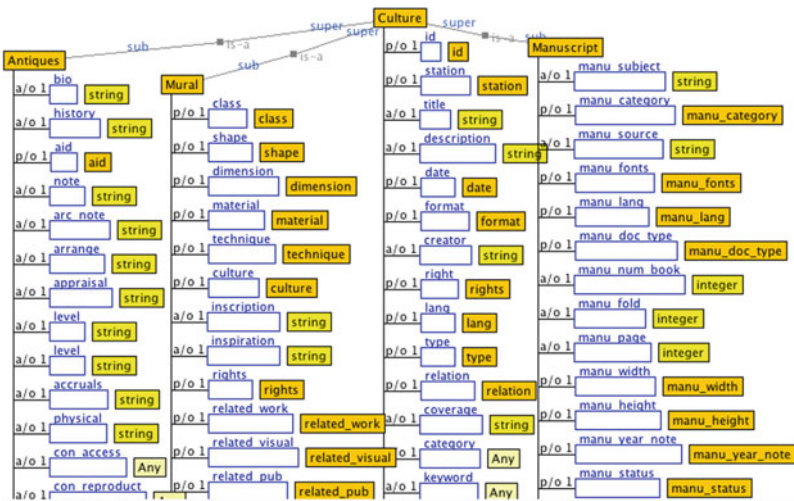


Fig. 5 Cultural heritage ontology (antiques, manuscript and mural)

The RDF data publishing used the Ontology Application Management (OAM) framework [18] for schema mapping and vocabulary mapping between OWL format and a database source. OAM facilitated defining mapping between ontology classes and database tables by class-table mapping and property-column mapping, then exports to RDF format.

4 Lanna Cultural Heritage Portal Application using RDF Data Sources

After the data sources are transformed to RDF data based on the schema defined in the ontology, the OAM framework provides the RESTful API that allows applications to utilize the RDF data sources in a uniform fashion. The “Lanna Cultural Heritage” portal is a web portal that provides a semantic view-based search engine where the knowledge base consisting of ontologies and data. The portal combines data from several data sources of museums and temples in the northern part of Thailand, for example, cultural objects from WatKukam, manuscripts and documents from Watsungmen, murals from Watphumin and cultural information center (www.m-culture.in.th) and so on (Fig. 6).

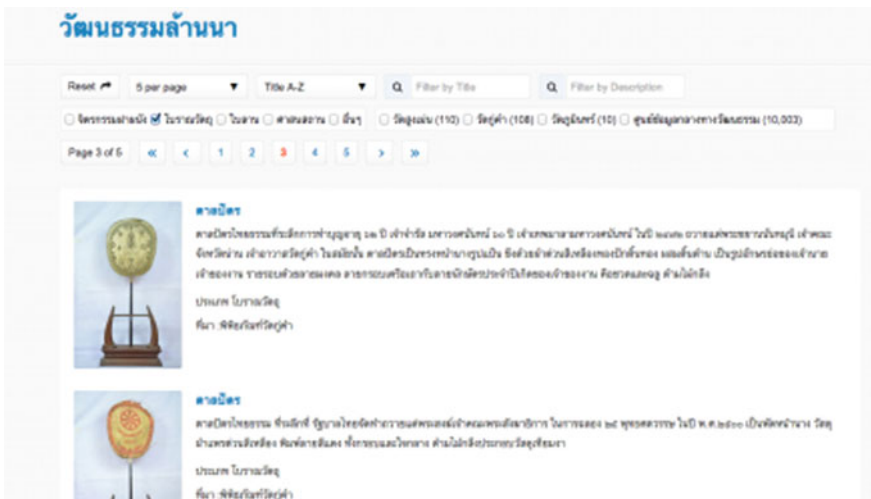


Fig. 6 Cultural heritage portal application using RDF data sources

5 Conclusion

This paper presents a methodology for integrating, converting Lanna cultural heritage metadata and sharing data to external data sources as linked open data. Case studies for this methodology are museum archives in north of Thailand and cultural information center from the Ministry of Culture. Our methodology is designed to convert different metadata formats to Dublin Core metadata standard as OAI-DC XML format, and to incorporate metadata to form a central repository of data sources. The extraction of key information from the result of repository is made by tag extraction and mapping of database schema using ontology under the Ontology-based Application Management (OAM) framework. The result is formulated in the form of RDF format. The data were

integrated to create cultural heritage portal application by using RDF data sources. Our directions for future work include applying the methodology to extract key concepts from the textual data and more complex ontology design for the Lanna cultural heritage domains.

References

1. Wikipedia: Culture (2015). <https://simple.wikipedia.org/wiki/Culture>
2. Culture. <https://www.tamu.edu/faculty/choudhury/culture.html>
3. Selçuk Candan, K., Liu, H., Suvarna, R.: Resource description framework: metadata and its applications. *SIGKDD Explor.* 3(1), 6 (2001)
4. The Europeana Foundation: Europeana Linked Open Data Culture. <http://labs.europeana.eu>
5. Edelstein, J., Galla, L., Li-Madeo, C., Marden, J., Rhonemus, A., Whysel, N. : Linked open data for cultural heritage: evolution of an information technology. In: *The 31st ACM International Conference on Design of Communication* (2013)
6. Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., Gergatsoulis, M.: Integrating Dublin Core metadata for cultural heritage collections using ontologies. In: *DCMI International Conference on Dublin Core and Metadata Applications, 27–31 August 2007*
7. de Boer, V., Wielemaker, J., van Gent, J., Hildebrand, M., Isaac, A., van Ossenbruggen, J., Schreiber, G.: Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study, *ESWC*, vol. 7295 of *Lecture Notes in Computer Science*, pp 733–747 (2012)
8. Haslhofer, B., Isaac, A.: data.europeana.eu The Europeana Linked Open Data Pilot. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications* (2011)
9. Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: Museum Finland—Finnish Museums on the Semantic Web, *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 3, issue 2–3, October, pp. 224–241 (2005)
10. Szekely, P., Knoblock, C.A., Yang, F., Zhu, X., Fink, E.E., Allen, R., Goodl, G.: Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In: *The 10th Extended Semantic Web Conference (ESWC)* (2013)
11. Chakraborty, S., Rahman, Md.H.H., Seddiqui, Md.H.: Linked Open Data Representation of Historical Heritage of Bangladesh, *Computer and Information Technology (ICCIT)*, 8–10 March 2014, pp. 242–248 (2014)
12. Suca, E.G., da Silva, F.S.C.: Sharing Cultural Heritage Information Using Linked Open Data at a Museum of Contemporary Art, *CIDOC*, 6–11 September 2014 (2014)
13. Skevakis, G., Makris, K., Arapi, P., Christodoulakis, S.: Elevating natural history museum’s cultural collections to the linked data cloud. In: *3rd International Workshop on Semantic Digital Archives* (2013)
14. Mouromtsev, D., Haase, P., Cherny, E., Pavlov, D., Andreev, A., Spiridonova, A.: Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing, *The Semantic Web Latest Advances and New Domains*, 21 May 2015, pp. 637–651 (2015)
15. Damova, M., Dannells, D.: Reasonable view of linked data for cultural heritage. In: *Third International Conference on Software, Services and Semantic Technologies* (2011)

16. Matsumura, F., Kamura, T., Kato, F., Kobayashi, I., Takahashi, T., Ueda, H., Ohmukai, I., Takeda, H.: LODAC Museum: integration and utilization of linked open data in museums. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence (2012)
17. Buranarach, M., Thein, Y.M., Supnithi, T.: A community-driven approach to development of an ontology-based application management framework. In: Semantic Technology Lecture Notes in Computer Science, vol. 7774, pp. 306–312 (2013)
18. The ontology-based application management (OAM) framework. <http://lst.nectec.or.th/oam/>