

# Cultural Heritage Information Aggregation in Networked Information Environments

Chiranthi Wijesundara

Graduate School of Library, Information and Media Studies,  
University of Tsukuba, Japan

chiranthis@gmail.com

**Abstract.** Cultural Heritage Information (CHI) is vital for understanding heritage assets and it is available in various formats, standards, qualities and quantities on the web. Primarily CHI is created, organized and delivered by memory institutions through delivery portals, referred to as Digital Archives in this research. Apart from carefully crafted CHI there are many other third parties who provide information through the web. This research deals with aggregating this diverse CHI on the web to help enrich poorly made information related to South and Southeast Asian region. Many drawbacks related to CHI on the web were realized after investigating several portals and websites, resulting in the proposal of a comprehensive metadata model known as Cultural Heritage in Digital Environment (CHDE). CHDE explicitly identifies instances to curate tangible and intangible cultural heritage resources into digital archives and to help identify metadata structures for resource aggregation. Further, the author investigated the metadata description identification based on their objectives, with the help of One-to-One Principle of Metadata through Description Module mapping aligned with the Dublin Core Application Profiles (DCAP).

**Keywords:** Cultural Heritage Information (CHI), Digital Archives, Dublin Core Application Profiles (DCAP), Metadata Aggregation, One-to-One Principle, South and Southeast Asia.

## 1 Introduction

Cultural heritage is the legacy of a society which showcases the values and qualities of its lifestyle, technology, religion etc. Memory institutions such as Libraries, Museums and Archives (LAM) play an important role in protecting these cultural heritages and delivering them to the public. Memory institutions try to preserve heritage's physical form and they also try to record related information using different mediums. A common goal of memory institutions is to create a collection of Cultural Heritage Information and archive them for delivery to the users on the web. In this research, the author is mainly dealing with web-based CHI on digital archives and websites delivered by different institutions and individuals.

The focus of this research is South and Southeast (S/SE) Asia and the author utilizes CHI datasets related to these regions. This research was initiated as an investigation related to Sri Lankan museum CHI aggregation and expanded further as a regional effort. During this investigation, many pitfalls in the CHI domain of the S/SE Asian regions were identified [7]. Nevertheless, it is a known fact that well-organized memory institutions in developed regions such as Europe and North America possess and deliver S/SE Asian CHI on the web. Therefore, the author assumed that it will be beneficial to aggregate this well-developed CHI in the western regions to enrich CHI in the S/SE Asian region, which is not very rich yet.

With these objectives, a model was developed, which is known as the Cultural Heritage in Digital Environment (CHDE) [6]. The CHDE enables one to differentiate between physical and digital environments of CHI and how they can be aggregated into a single unit known as a *Curated Digital Instance*. Currently, the author is investigating a real-time metadata aggregation process referred to as *Description Modules*. *Description Modules* tries to identify different components in a single CHI record, which is based on the DCMI's (Dublin Core Metadata Initiatives) One-to-One Principle of Metadata [9], and the Dublin Core Application Profiles (DCAP) [2].

## **2 Background**

### **2.1 Cultural Heritage Information (CHI) on the Web: General Issues**

Heritage related information is mostly considered as diverse and complex. This heterogeneous nature creates multiple formats, contexts, standards etc. [3]. Therefore, handling CHI becomes complicated in networked information environments.

Generally, museum objects are described as individual objects. This item-centric perspective is useful for knowledge organization and it provides user interfaces to find and access museum objects. Sometimes, a record displaying a single item may be unable to fulfill this user need, so that users need to find richer information. In addition, these metadata records have a mixture of information. For instance, a museum record may contain both the original object information and its digital representation information. This creates confusion because the objectives of the metadata are not clear [5].

Another major CHI issue is the scarcity of the Linked Open Data (LOD). Most of the web resources and digital archives in question are operated as isolated entities and they have no links to connect with other resources. This might create problems on the web in the long run. Problems associated with the longevity of digital resources and identifying their rights metadata also become a burden in the CHI domain.

Apart from these issues, S/SE Asia has its own CHI issues such as a lack of web-based digital archives, metadata standards issues and so on [7].

### **2.2 Model Based Metadata Aggregation**

Metadata aggregation can be performed using various techniques and, in this study, the metadata aggregation is done using models. Metadata models can be used as a key to unite dispersed information on the web. Some renowned metadata aggregation models

are The Open Archives Initiative- Object Reuse and Exchange (OAI-ORE)<sup>1</sup> and European Data Model (EDM) [4]. Europeana collects digital cultural heritage provided by memory institutions across Europe and this aggregation is done using the EDM. The aggregated CHI is further enriched and delivered to users through Europeana's portal<sup>2</sup>.

Since CHI are complex and use different standards, mapping metadata to each other for metadata aggregation is not a simple process. Therefore, this research proposes a generalized metadata model which can be used to aggregate metadata from diverse resources.

### 2.3 Objectives and Novelty of the Research

The primary goal of this research is to develop a metadata model to aggregate dispersed digital archive resources (institutional) and non-institutional archive resources into a solitary unit of access. The One-to-One Principle of Metadata is used as the base concept to distinguish between objects and their metadata descriptions. Through this aggregation, CHI enrichment is expected as the CHI is linked to various web resources.

This kind of CHI aggregation is not common in S/SE Asia and this will be significantly important to the region. In addition, the identification of Physical and Digital Spaces of CHI and their metadata separately is also another significant attempt observed through this research.

## 3 Related Models and Concepts

EDM is one major related model in this research which specifically designed for CHI aggregation [4]. Secondly, OAI-ORE<sup>1</sup> is another model-based approach designed to aggregate web resources.

Since this research is handling metadata descriptions, Metadata Encoding and Transmission Standard (METS)<sup>3</sup> is an important standard because it tries to organize metadata using a set of elements.

In addition, CIDOC-Conceptual Reference Model (CRM)<sup>4</sup> and its extension known as FRBRoo (FRBR-object oriented)<sup>5</sup> can be identified as some other models and ontologies specifically designed to organize cultural heritage resources through various entities and relationships.

Finally, DDCMI's One-to-One Principle of Metadata [9] and DCAP [2] are two main concepts that were used in this study to distinguish different metadata descriptions on the web.

---

<sup>1</sup> <https://www.openarchives.org/ore/>

<sup>2</sup> <https://www.europeana.eu/portal/en>

<sup>3</sup> <https://www.loc.gov/standards/mets/>

<sup>4</sup> <http://www.cidoc-crm.org/>

<sup>5</sup> [https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo\\_v\\_2.4.pdf](https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf)

## 4 Cultural Heritage in Digital Environment (CHDE) Model: An Approach to Organize CHI on the Web

### 4.1 Model Overview of Cultural Heritage in Digital Environment (CHDE)

The proposed CHDE model has two main divisions: the *Physical Space* and the *Digital Space* (Fig. 1). Whether tangible or intangible, any cultural heritage can physically exist or occur in the real world. On the other hand, memory institutions develop their digital collections by digitizing those physical instances. CHDE defines the entities in the process of digital curation by memory institutions. It uses the One-to-One Principle of metadata to identify the entities and to aggregate metadata collected from multiple sources, i.e., institutional digital archives and non-institutional web resources about cultural heritage [6].

According to Figure 1, the left portion represents a *Tangible Cultural Heritage (TCH)* and the right portion is for an *Intangible Cultural Heritage (ICH)*. In the TCH section, heritage objects can exist as physical objects and they can be directly recorded as photographs and audio recordings which may be stored in various media such as compact discs, tapes and local databases. These entities are modeled as *Offline Resources* in the *Physical Space*. The *Offline Resources* can be converted or just utilized as online *Digital Resources* and aggregated into a *Set of Digital Resources* which is named as *Curated Digital Instance* in the figure.

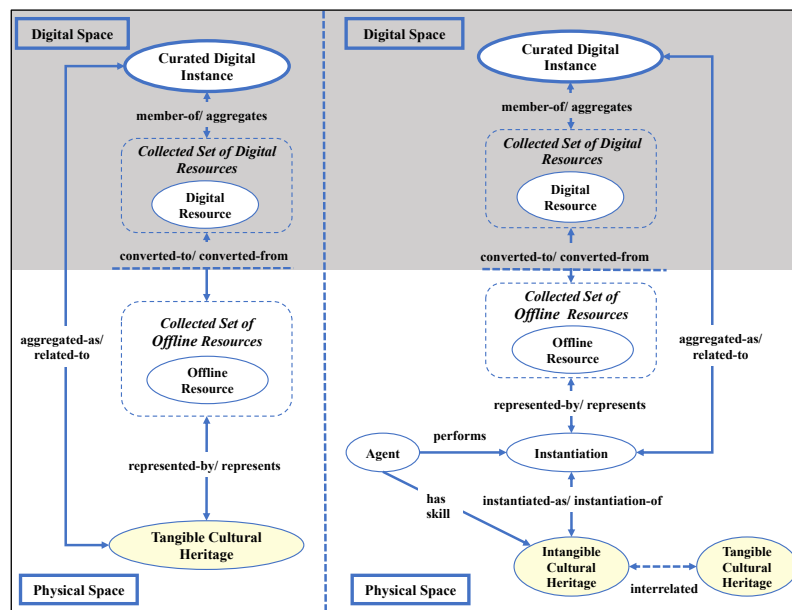


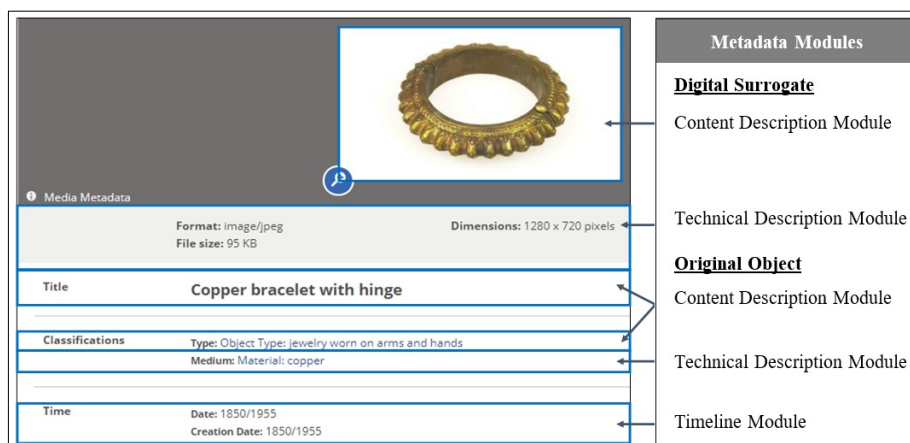
Fig. 1. CHDE Model: Describing Tangible and Intangible CHI Aggregation

In the ICH section, it is not possible to directly capture ICH but their instantiations, such as dance performance and traditional paper making skill, which physically exist

within a given space and time. This specific occurrence of ICH is represented as an *Instantiation*, and it can be captured as TCH objects via *Offline Resources*. The *Offline Resource* can be converted to online *Digital Resources* and finally collected into a *Curated Digital Instance* in the *Digital Space*.

## 4.2 Distinguishing Metadata Descriptions by Description Modules

Above CHDE model does not describe a real metadata aggregation situation. It is only an abstract level representation of possible TCH and ICH information aggregation in different levels. To apply CHDE for aggregating metadata extracted from existing CHI resources, the author developed the *Description Modules* (connected to the CHDE Model concept), which is a structural component enabling the identification of different metadata descriptions in different records (Fig. 2), created using different schemas.



**Fig. 2.** Description Modules Mapped to Metadata Record Instances

These *Modules* exist as four main categories: *Original Object*, *Digital Surrogate*, *Administrative* and *External Resource*. Each category was further described using several modules e.g., *Content Description Module*, *Timeline Module*, *Agent Module*, *Rights Module* etc. To create these modules author investigated more than ten institutional plus non-institutional resources on the web. Figure 2 shows an example taken from Europeana<sup>6</sup> to show how the mapping was done between digital archival records and the proposed metadata modules. These *Modules* can be used to differentiate these distinct resource descriptions individually. Further, these Metadata Modules were aligned with the DCAP. The Description Set Profiles (DCP) in the DCAP is utilized in this context to clarify the proposed *Description Module* idea.

<sup>6</sup> <https://www.europeana.eu/portal/en/record/2048221/81452.html>

## 5 Discussion and Future Direction

The paper proposed a metadata model to aggregate heterogeneous CHI on the web. Identification of different levels associated with the CHI and their relationship with the TCH, ICH instances are described by the CHDE model. In addition, avenues of aggregation were found through mapping diverse metadata records and revealed that *property-by-property* level (or *attribute-by-attribute*) mapping is not effective in the context of miscellaneous information defined using various schemas. Therefore, the author proposed some structural units of metadata, i.e., *Description Modules*, for mapping, instead of conventional *property-by-property* level mapping.

There is criticism to the One-to-One Principle, though researcher learned that One-to-One relationship is crucial to handle metadata collected from multiple sources and aggregate them [5]. It is a known factor that DCAP is a crucial model to develop interoperable metadata, but there is a need to extend it as *Description Module* model for CHI metadata aggregation.

The future direction of this research is to apply the model to cultural heritage of S/SE Asia and to build a system based on these investigations. Author understand that rights and provenance metadata descriptions are vital for long-term maintenance of aggregated metadata, but they are also left for future work.

**Acknowledgement.** This research is partially supported by JSPS KAKENHI Grant (Grant-in-Aid for Scientific Research) number: (A) #16H01754. The author would like to express her great appreciation to professor Shigeo Sugimoto who is the main supervisor of this research, for his invaluable guidance and support during planning and development of this research.

## References

1. Baca, M. (Ed.). Introduction to metadata. Getty Publications (2008)
2. Coyle, K., Baker, T.: Guidelines for Dublin core application profiles. Dublin core metadata initiative. (2009). <http://dublincore.org/documents/profile-guidelines/>
3. Hyvönen, E.: Publishing and using cultural heritage linked data on the semantic web. Synthesis Lectures on the Semantic Web: Theory and Technology, 2(1), pp. 1–159 (2012)
4. Isaac, A. (ed.): Europeana data model primer (2013). [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)
5. Miller, S. J.: The one-to-one principle: challenges in current practice. International Conference on Dublin Core and Metadata Applications, pp. 150-164 (2010)
6. Wijesundara, C., Monika W., Sugimoto S.: A metadata model to organize cultural heritage resources in heterogeneous information environments. Choemprayong S., Crestani F., Cunningham S. (eds). ICADL 2017. LNCS, vol 10647, pp. 81-94. Springer, Heidelberg (2017). doi: [https://doi.org/10.1007/978-3-319-70232-2\\_7](https://doi.org/10.1007/978-3-319-70232-2_7)
7. Wijesundara, C., Sugimoto, S., Narayan, B., Tuamsuk, K.: Bringing cultural heritage information from developing regions to the global information space as linked open data: an exploratory metadata aggregation model for Sri Lankan heritage and its extension. A-LIEP-2016, pp. 117–132 (2016)

8. Woodley, M. S.: Metadata matters: connecting people and information. Baca, M. (Ed.): Introduction to metadata. Getty Publications (2016)
9. Woodley, M. S., Clement, G., Winn, P.: DCMI glossary (2005). <http://www.dublincore.org/documents/usageguide/glossary/>