A Study on Organizing Heterogeneous Cultural Heritage Information in Networked Information Environments

A Study on Organizing Heterogeneous Cultural Heritage Information in Networked Information Environments

Chiranthi Jayampathini Wijesundara

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

August 2019

A Study on Organizing Heterogeneous Cultural Heritage Information in Networked Information Environments

Abstract

Cultural Heritage is a showcase of a particular human society which demonstrates its values, traditions and characteristics. Preserving cultural heritage is essential to bestow it to future generations. This study was initiated as a way of finding avenues to help support safeguarding cultural heritage via informational interventions. Basically, "information" is the key to understanding a specific cultural heritage entity, as well as giving it meaning and context. Therefore, the idea of organizing and linking cultural heritage information for better access and for more context is the foundation of this study.

As digital information about cultural heritage is important in this research, the study introduces the term Cultural Heritage Information (CHI) which denotes digital information related to cultural heritage. Memory institutions are generally responsible for handling CHI; they organize digital archives as a collection of CHI and disseminate CHI on the Internet. CHI in digital archives is often based on individual cultural heritage objects, referred to as "item-centric" information in this research. Generally, a user's information needs are diverse and complex, and they sometimes require additional information related to a certain cultural heritage object that conventional digital archives - composed of item-centric CHI - are unable to deliver. On the other hand, the Web provides a large amount of cultural information resources delivered by third-party, non-memory institutions, such as Wikipedia. Those resources may not be well standardized to describe cultural heritage, but are popular among the general public. Thus, linking institutional and non-institutional CHI into a single platform will give opportunities to understand cultural heritage as a complete entity while fulfilling complex user needs.

The main goal of this research is to develop a model for metadata to organize heterogeneous cultural heritage resources in networked information environments by aggregating institutional and non-institutional CHI. This goal includes developing two metadata models: (i) a model named Cultural Heritage in Digital Environment (CHDE) which defines entities included in a development process of digital archives and (ii) a model named Description Modules model which defines metadata mapping aimed for aggregating metadata descriptions in diverse schemas.

Accordingly, the author conducted a preliminary survey of the CHI situation in the networked information environments related to South and Southeast Asia. The rich cultural heritage background associated with socio-religious factors such as Buddhism and Hinduism led to the selection this region as the main study area of this research. This survey provided some insight used to identify the prevailing information conditions and issues in the South and Southeast Asian region. The author learned that inadequacy of web-based information, rigid institutional policies, digitization, and information sharing limitations were some common issues associated with the CHI in the region. Furthermore, she learned intangible cultural heritage information organization was given less priority, and the region had no proper controlled vocabularies or digital information related to the same. All of these factors prevented the creation of a rich digital archive of tangible and intangible cultural heritage in the South and Southeast Asia.

At this point, the study is left with some fundamental questions, such as what kind of CHI should be linked, and what is the suitable technology for this linking?

The study focused on linking institutional and non-institutional CHI related to both tangible and intangible cultural heritage objects. However, in the digital environment, tangible and intangible deviation has no meaning, as everything is being considered as digital information. Since South and Southeast Asia do not possess many digital archives and web-services, this research relied on alternative digital archives openly accessible via memory institutions in Europe and North America. These digital archives consist of potentially related CHI which can be used to enrich poorly resourced services in South and Southeast Asia. As a result, the study utilized existing CHI freely available on the Web without having to develop new ones.

Subsequently, "information aggregation" is the main technology behind linking information. The author investigated various well-known aggregation efforts, such as Europeana, prior to designing the proposed models. The Europeana Data Model (EDM) defines a basic model to aggregate diverse digital cultural heritage resources provided by memory institutions in Europe and disseminate the aggregated information via the its collection. As Europeana uses EDM to aggregate their information and it is specifically designed for CHI aggregation, this could be recognized as one of the main related studies of this research. However, Europeana's well-structured data ingestion process and top-down data model approach cannot be fully utilized to aggregate CHI in South and Southeast Asia. In addition, cultural heritage ontologies such as CIDOC-CRM are used in cultural heritage knowledge management in different scales. However, none of these models or ontologies can be fully

utilized to aggregate both tangible and intangible CHI in the South and Southeast Asian region which are fragmented on the Web and created based on diverse schemas.

Therefore, the thesis presents a novel conceptual model for metadata aggregation. The model has two main component models as mentioned above. Primarily, Cultural Heritage in Digital Environment (CHDE) model defines entities that should be included in the development process of digital archives. The CHDE model defines a generalized CHI aggregation scheme for both tangible and intangible cultural heritage.

The CHDE model gives a special reference to organizing and capturing intangible cultural heritage entities. Identification of an "Object" of intangible cultural heritage is an important requirement in this setting. Intangible heritage is a well-organized collection of knowledge and skills (e.g., traditional dances, rituals) and have no physical existence, unlike tangible cultural heritage objects. The knowledge and skills achieve a physical existence through performances and actions in/at a specific time and location. Based on this perspective, the CHDE model defines an entity called *Instantiation* which acts as a physically existing "Object" in the real world. *Instantiations* can be captured and recorded by memory institutions using electronic tools. Often memory institutions collect and exhibit the records of *Instantiations* as surrogates of intangible cultural heritage and we often see memory institutions exhibit intangible cultural heritage by various recordings of *Instantiations.* However, digital archives created by these memory institutions do not explicitly identify the *Instantiations* as a surrogate of intangible cultural heritage. When it comes to tangible cultural heritage objects e.g., paintings, sculptures, this situation is not relevant as they can be directly captured by various mediums. CHDE explicitly identify the collected records as *Instantiation*s of intangible cultural heritage so that the records can be aggregated into a larger unit, i.e., a collection of CHIs, which may represent intangible cultural heritage as a collection of the records with contextual information.

As a whole, the captured records can be converted into digital formats regardless of their tangibility and then aggregated into a single entity named *Curated Digital Instance*, which can be later used to form a digital archive of cultural heritage. The CHDE model proposed in this thesis presents a comprehensive set of definitions of these entities in the physical and digital spaces and their relationships.

Theoretically, the CHDE model provides a way to aggregate diverse resources in networked information environments. CHDE's *Curated Digital Instance* consists of a set of digital objects, descriptions and links related to a specific cultural heritage object, which encompasses a great deal of metadata belonging to different schemas. The second model of

this study named the Description Modules Model is developed based on the above concept and to define a metadata mapping framework aimed to aggregate metadata described in diverse schemas. The idea of *"one metadata description should describe one object only"* known as the One-to-One Principle of Metadata is used as the fundamental concept when creating this model. The author realized that the institutional metadata mostly contains descriptions about multiple objects in a single records which is often called "Hybrid". Hybrid metadata is useful in a self-contained environment, but it may cause problems in metadata aggregation across different schemas because we need to identify a target object of each component in metadata and find correspondence between the components across different schemas prior to aggregation. This study developed the Description Modules model to overcome this Hybrid CHI problem using the One-to-One Principle.

The Description Modules Model can be utilized to overcome crosswalk problems such as property-to-property level metadata mapping used in conventional metadata aggregation. Primarily, multiple schemas increase the number of combination of property pairs for metadata mapping which leads to an unmanageable and ambiguous crosswalk. In this situation, Description Modules can be utilized to distinguish metadata chunks according to their objectives and further apply this model to diverse CHI records on the Web enabling better aggregation. This kind of application reduces the complexity of the crosswalk and it provides a chance to identify different types of metadata descriptions within a single CHI record which is useful for information aggregation. However, this discussion needs to be explored more in the future.

The feasibility study for these proposed models was manually done. The author has clarified the feasibility of the models from this study, but she has learned that aggregation of non-institutional CHI (e.g., Wikipedia) should be explored more.

To conclude, this novel approach will benefit memory institutions which have insufficient resources to create digital resources such as those in South and Southeast Asia. The intangible cultural heritage organization through instantiation is a useful approach that is discussed in this research. In addition, identifying CHI components via Description Modules is a novel aspect presented in this thesis.

Table of Content

List of Figures

List of Tables

Appendices

Chapter 1: Introduction

The cultural heritage of a human society includes unique characteristics and values related to the society's knowledge, traditions, economy, craftsmanship, and more. Cultural heritage can be tangible; statue, painting or monument and intangible; traditional dance, handicraft or ritual. This research identifies these tangible and intangible cultural heritage entities as *cultural heritage objects* regardless of their differences. Cultural heritage objects can be found all over the world and they are preserved and exhibited in galleries, libraries, archives and museums (GLAM), often called memory institutions. Large quantities of Cultural Heritage Information (CHI) based on cultural heritage objects are continuously produced and added to the Web by the memory institutions, other public and private institutions, and by many individuals.

The CHI exists as isolated services or websites. Sometimes users might need a range of diverse information on a cultural heritage object which is not limited to a conventional museum record or a single digital recording. Therefore, the main goal of this research is to present an approach to help aggregate and organize both tangible and intangible cultural heritage related information which are held in isolated data silos on the Web. This aggregation can enable users to gain more context-rich information than item-centric information retrieved from a conventional digital archive on the Web.

Therefore, this research recognizes CHI as a key term defined as, "description or representation of a particular cultural heritage object which can be tangible or intangible". Here, representations characterize any digital surrogate (image, video or audio recording) related to a cultural heritage object. Previously CHI has been considered difficult to deal with from the viewpoint of interoperability on the Web because of its heterogeneity. Nevertheless, memory institutions (referred to as *institutions* in this thesis) accept this challenge and intervene in this process. They collect cultural heritage objects as their resources and digitize them, organizing the digital cultural heritage as a part of their collections, and provide this information to their patrons via the Web and/or their in-house services. As scholars identify, there are three types of data of memory institutions: (i) structured data, e.g., bibliographies, indexing and abstracting databases, (ii) semi-structured data, e.g., unstructured sections within metadata descriptions (for instance, notes in bibliographic records) and (iii) unstructured data, characterized as "everything else", e.g., documents and other information-bearing objects (textual or non-textual, digitized or non-digitized) (Zeng, 2019). Memory institutions use metadata standards to organize this data/information about their holdings in

accordance with their demands, for instance, heritage related metadata schemas and ontologies such as CDWA (Categories for the Description of Works of Art)[1], LIDO (Lightweight Information Describing Objects)[2], and CIDOC-CRM (CIDOC-Conceptual Reference Model)[3], information aggregation models like EDM (Europeana Data Model)[4], and authority control standards such as the Getty thesauri[5]. This range arose from the heterogeneity of memory institutions and their requirements. For instance, objects of a gallery and objects of an ordinary museum are different, and they need different standards and schemas to document their objects. An intangible cultural heritage museum may be completely different from the above and need to deal with multiple recordings, virtual reality (VR) programs and objects as well. Therefore, handling and documenting cultural heritage objects is not homogenous compared to materials such as books in a library. As a result, institutions follow different standards to organize their resources and finally end up creating isolated information gateways with multiple schemas. According to scholars, such diversity has drawbacks as follows. "… differences in descriptive schema across museums, libraries and archives, although necessary for individual applications, can seriously hinder cross-domain discovery and interoperability of cultural information resources in the global context of the Internet" (Gill, 2004).

When the information is not linked, information users end up searching institutional portals one-by-one and it takes more time and effort and ultimately, they end up with some limited information. In addition, maintaining individual portals or digital archives is costly and if the information is linked it might give more benefits to the institutions as they can share their information easily. Hence, enhancing metadata has become an important concern in the institutional data enhancement efforts, in order to overcome challenges relating to data quality and discoverability in the digital age, while providing more context and multilingual information for cultural heritage objects (Isaac et al., 2015).

With the advance of Web technologies, GLAM is currently trying to find novel approaches to link these isolated collections built by individual institutions and present them as connected large-information portals, aiming for easy accessibility and context-rich information. The Europeana data portal[6] which has been developed to collect and disseminate

---

[1] http://www.getty.edu/research/publications/electronic_publications/cdwa/
[2] http://network.icom.museum/cidoc/working-groups/lido/
[3] http://www.cidoc-crm.org/
[4] https://pro.europeana.eu/resources/standardization-tools/edm-documentation
[5] http://www.getty.edu/research/tools/vocabularies/index.html
[6] https://www.europeana.eu/portal/en

digital cultural heritage on the Web is a typical example of such an effort. Regardless of the standards used in the institutions, Europeana gathers digital cultural heritage information provided by institutions across the European Union. Europeana does not get involved in the institution's digitization activities, nor do they provide any instructions. Although they collect existing data in diverse schemas from different data providers (basically the GLAM), they finally ingest this information into their system based on the Europeana Semantic Elements (ESE). In 2010, they introduced a richer aggregation model widely known as the Europeana Data Model (EDM) for their digital cultural heritage aggregation process (Europeana Data Model Primer, 2013). Since this research is also focusing on CHI aggregation avenues, Europeana can be considered as one of the most relevant studies in this setting.

The initiation of this research was not solely based on information aggregation. The original idea was *"Safeguarding of cultural heritage via informational intervention"*. Since cultural heritage can be preserved physically and informationally, the author investigated the informational aspects of cultural heritage specific to Sri Lanka. How metadata standards can be used at the ground-level of heritage documentation was investigated as the very first stage of this study. Documentation of Spatial (location) and Temporal (time) information related to tangible-immovable heritage is critical to identification of a specific monument, its surrounding environment and for future planning processes, etc. Specially, when heritage is destroyed due to natural or man-made activities, this information is vital for reconstruction or restoration of the cultural heritage object back to its previous state. Thus, the author proposed a customized metadata standard to document Sri Lankan heritage based on the MIDAS Heritage which is a data standard for historic monuments in the UK (Wijesundara, Sugimoto, & Narayan, 2015).

Later, the author narrowed the research towards the Sri Lankan museum information aggregation and experienced many hitches particular to the museum information domain. One of the major limitations of Sri Lankan museums is the absence of remotely accessible CHI. Currently, Sri Lankan museums have no ability to provide online access to their patrons. National museums and regional museums of Sri Lanka maintain a standard manual recording system and a computer application to record information within the organization. But this information is not available to the general public. During this preliminary survey, the author found this problem and finally had to depend on published printed catalogs. Though, the national museums do have published catalogs, which are not restricted, but they do not cover all the objects in the museums in Sri Lanka, and not even all the objects in the national museum in Colombo (Wijesundara, Sugimoto, Narayan, & Tuamsuk, 2016).

Due to the scarcity of online information, alternative data sources had to be explored and consequently, the author found that there are well-standardized, openly accessible CHI related to Sri Lanka available in Europe and North American institutional digital archives (*digital portals* are referred to as *digital archives* in this thesis). Since Sri Lanka and many other South and Southeast Asian countries were under colonial occupation, an enormous amount of Sri Lankan artifacts can be found in European countries. These artifacts were not returned back to their original countries but were detained by the institutions and private collectors. Fortunately, most of these artifacts can be openly accessed currently via well-established institutional digital archives such as the British Museum, Rijksmuseum, etc. Therefore, we can use this rich information to enrich the poorly developed CHI in Sri Lanka by aggregating them in different memory institutions through a semantic metadata model defined using the Resource Description Framework (RDF), a fundamental standard for a Linked Open Data (LOD) environment on the Web. This aggregation was conducted by a metadata crosswalk approach between museum vocabularies, which was achieved by investigating over 2600 object records across four museums inside and outside Sri Lanka, and by mapping them to the Getty Art and Architecture Thesaurus (AAT) (Wijesundara, Sugimoto, Narayan, & Tuamsuk, 2016). By this research, the author learned that aggregation of multiple resources is beneficial to create context-rich information and it gives better accessibility to CHI as well.

The same concept of "*aggregating dispersed information for better access and data enrichment*" was later applied in a broader context. This time the research was expanded towards the South and Southeast Asia aiming at cultural similarities such as Hinduism and Buddhism. For instance, the "Ramayana" [7] is a common tradition shared by the whole region and many cultural heritage objects, monuments and intangible heritage related events are created based on the "Ramayana" story. If we consider a traditional dance performance, such as "Khon" dance performance in Thailand, "Ramlila" Performance in India and Indonesian "Sendratari Ramayana" performance, they all exist as individual performances but share a common base in "Ramayana". On the other hand, the objects associated with performances such as props, masks, musical instruments may be very different. If we can aggregate these things into a single platform, however, it will be beneficial for enriching the CHI content and might also be advantageous from the user's point of view. For instance, if a museum collects *masks* related to a "Khon" dance performance in Thailand, this particular artifact/s can be

---

[7]https://www.britannica.com/topic/Ramayana-Indian-epic

associated with various other entities such as a photograph related to the *mask* or video recording of a performance by someone wearing the *mask*. When a user needs additional context related to the *mask,* he/she should retrieve these details separately from individual Web-services. If all the associated information is linked together, it enables easy access and more-richer information to the user. Hence, aggregating dispersed cultural heritage objects (which may be tangible or intangible) based on their relationships has become one of the main goals of this research.

Unfortunately, the CHI situation of developing regions such as South and Southeast Asia is not as in the developed countries. It is obvious that South and Southeast Asian memory institutions do intervene in the CHI creation, management, and dissemination, but they operate as individual data silos without interconnection. For instance, the initiatives of the Thailand government, such as the Cultural Knowledge Center by the Ministry of Culture and Ministry of Science and Technology, and the Princess Maha Chakri Sirindhorn Anthropology Centre, help promote the knowledge structure in cultural heritage by providing various services to the public (Wijesundara, Sugimoto, Narayan, & Tuamsuk, 2016). In addition, the lack of widely accepted standards to share information among the institutions leads to many barriers when linking information on the Web. For example, the region does not have controlled vocabularies which can be utilized in the CHI aggregation. Moreover, the information existing on the Web provided by the South and Southeast Asian institutions is incomplete and is not enough to describe cultural heritage objects in the region. As well, most of the institutions in the region are reluctant to provide their information to the general public due to various social factors such as security issues in the museums.

Nevertheless, other than from the institutions, there are many Web resources created by third parties (non-institutions), such as Wikipedia, which is a widely used encyclopedia, and also individual websites/ services, and blogs created by experts and various individuals. Those Web resources are useful for many end-users wanting to understand (contextual information about) the cultural heritage objects. This non-institutional information might not be higher quality, but still it can provide some context to cultural heritage and these resources are abundant and openly available on the Internet. Since this research focusses on existing, usable CHI on the Web related to the South and Southeast Asia, the author identifies this non-institutional CHI as a key information resource which can be utilized in this research. Also, this research uses well-standardized, openly available digital archival information from Europe and North American institutions as they relate to the South and Southeast Asian

heritage. It is crucial to link this institutional CHI and the non-institutional CHI to add value to the cultural heritage objects provided by the GLAMs or memory institutions.

Considering all these factors and issues related to the CHI situation in the study area, the author has understood the CHI should be aggregated in such a way to provide better accessibility while enabling more context-rich information. Therefore, this study basically proposes two metadata models to achieve this result. The two models are (i) the Cultural Heritage in Digital Environment (CHDE) model which defines entities that should be included in the development of digital archives and (ii) the Description Modules model which defines metadata mapping aimed for aggregating metadata descriptions available in diverse schemas.

As a digital archive of cultural heritage is a key feature to lower the above-mentioned barriers, the author defined a model called Cultural Heritage in Digital Environment (CHDE) which organizes and aggregates metadata for both tangible and intangible cultural heritage aiming to form a digital archive (Wijesundara & Sugimoto, 2018; Wijesundara, Monika, & Sugimoto, 2017). For various reasons, EDM-like aggregation models cannot be fully incorporated into the South and Southeast Asian CHI aggregation process. Consequently, the author designed this CHDE model as a solution to the region's CHI aggregation and organization process. CHDE clearly distinguishes the physical and digital environments of a CHI. Hence, the research specially focuses on the *Digital Space* (or digital environment) as CHI aggregation is taking place in the networked environment. The final outcome of the CHDE is called *Curated Digital Instance* which is a composite cultural heritage resource corresponding to a particular cultural heritage object in the *Physical Space*.

Since intangible cultural heritage is more diverse compared with the tangible cultural heritage, capturing and aggregating this information is challenging. Current CHI aggregation models favor tangible heritage aggregation and there is no suitable model for aggregating intangible cultural heritage. Consequently, the CHDE model gives a special focus to model the intangible cultural heritage through an *Instantiation* based information organization approach which is unique to this study. This can be identified as a focal point of this model proposal.

CHDE is aiming to clearly identify the objects which may be tangible, intangible or digital. The model uses a fundamental concept known as the One-to-One Principle of Metadata[8] coined by the Dublin Core Metadata Initiative (DCMI) for this identification.

---

[8]http://www.dublincore.org/documents/usageguide/glossary/

While CHDE defines entities essential for an abstract level CHI organization and aggregation aiming to create digital archives, in reality, we have to deal with various CHI metadata records on the Web. When dealing with these metadata, the author has identified some common problems specifically associated with the CHI metadata descriptions. Basically, institutional digital archives and non-institutional means provide a mixture of information (*Hybrid* records) about original heritage objects and their digital surrogates. A significant problem with this is that the objectives of the GLAM metadata description and those of Web resources are quite different. One of the primary contributions of this study is to clearly identify the objectives of metadata descriptions to link the GLAM's metadata and Web resources based on the One-to-One Principal of Metadata.

Moreover, conventional metadata aggregation is done through mapping across different schemes of metadata descriptions, i.e., property-by-property level mapping. However, this conventional method is very complex and time-consuming. This research proposes to use structural units of metadata for mapping instead of individual properties. Dublin Core Application Profiles (DCAP) (Coyle & Baker, 2009) is used as a related work when defining the structural units of metadata termed as *Description Modules* in this research.

The second model proposed by this study which is named as the Description Modules model helps support the metadata aggregation effectively. The Description Modules model uses the One-to-One Principal of Metadata as a base principle, to identify objects and the objectives of metadata descriptions can be carried out using the Description Modules model. However, identifying an object in the physical space, e.g., Original Object, is not possible with the *Description Modules* as they are representing the data structure of an object only. Therefore, the research created *Facets* which act as solid entities, to identify the *Object* in the real world. Thus, the identification of a structural data view of an object could be made using the *Description Modules* while the outer perspective of an object can be viewed through the *Facet*s, as an extension to the *Description Modules*. The feasibility of the CHDE model is tested based on the real CHI instances and Description Modules model.

Aggregation of heterogeneous CHI from institutional and non-institutional Web services aiming for a digital archive of CHI is a new approach to the South and Southeast Asian CHI domain. The proposed model is designed to be region neutral. In addition, the organization of intangible cultural heritage through *Instantiation* and utilization of the One-to-One Principle of Metadata to distinguish the metadata of CHI records is a major outcome of this research. The Description Modules model and *Facet* idea gives some new insights of

viewing objects via different perspectives which is essential for database creation and information retrieval.

The main contents of this thesis are structured as follows. Chapter 2 includes definitions and concepts related to the study, research problems, requirements for model-based metadata aggregation and the novelty of the research. Chapter 3 discusses the literature review based on the cultural heritage information modeling and organization and cultural heritage information aggregation. Chapter 4 is dedicated to introducing the main model known as the Cultural Heritage in Digital Environment (CHDE) and its components and functionalities. Chapter 5 introduces the Description Modules model concept which describes the aggregation technology in this model while Chapter 6 is focused on the case studies and crosswalk based on the CHDE model and Description Modules model and *Facet* idea which are crucial for the feasibility evaluation of the models. Chapter 7 is dedicated to the results and discussion, while the last chapter (Chapter 8) summarizes the whole thesis.

Chapter 2: Cultural Heritage and Its Information

2.1.   Related Definitions and Concepts

This section presents definitions of major concepts and terms used in this study, including those defined specifically for this study.

2.1.1. Cultural Heritage and Information Organization: Definitions Associated

Cultural Heritage. As defined by the UNESCO, "Cultural heritage is the legacy of physical artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations"[9]. Cultural heritage has many sub-facets and numerous variations. Since it is involved with culture and society, it is always complex and sometimes ambiguous. This study uses the term *cultural heritage objects* to denote the entire cultural heritage domain including tangible and intangible heritage.

Tangible Cultural Heritage. Tangible cultural heritage means "objects significant to the archaeology, architecture, science or technology of a specific culture"[9]. Tangible cultural heritage can be further subdivided into movable objects such as paintings, coins, sculptures in museums and immovable objects as archaeological sites and monuments. The term *tangible cultural heritage object* is used to refer to a tangible object in the real world (Appendix 1).

Intangible Cultural Heritage. UNESCO defines intangible cultural heritage as the "traditions or living expressions inherited from our ancestors and passed on to our descendants"[10]. Intangible cultural heritage cannot be touched. It occurs in a given space (i.e., in the physical world) during a given time, enabling a physical existence for a while. Traditional dance performance, culinary art and handicraft come under this category. Intangible cultural heritage can be split into five main categories as follows.

  i.   Oral traditions
  ii.   Performing arts
  iii. Social practices, rituals, festive events
  iv.   Knowledge and skills to produce traditional crafts

---

[9] http://www.unesco.org/new/en/cairo/culture/tangible-cultural-heritage/
[10] https://ich.unesco.org/en/what-is-intangible-heritage-00003

v.   Knowledge and practices concerning nature and the universe[10] (see Appendix 1)

This study identifies Cultural Heritage, Tangible Cultural Heritage, and Intangible Cultural Heritage according to the UNESCO definitions and categorizations. Further, this research made a mapping between these tangible and intangible cultural heritage categories with the terms and classes from the Getty Art and Architecture Thesaurus, American Folklore Society Ethnographic Thesaurus, CIDOC-CRM and FRBRoo (Functional Requirements for Bibliographic Records - object oriented) ontologies for better understanding and formalization of the terms (Wijesundara, Monika, & Sugimoto, 2017). The full mapping table is included in Appendix 1 of this thesis.

Cultural Heritage Information (CHI). This research identifies CHI as "a description or representation of a certain heritage object which can be tangible or intangible". Catalog records at memory institutions, Wikipedia articles about cultural heritage, and bibliographic data provided by digital archives are typical CHI. In this study, digital surrogates being digital copies of original cultural heritage objects are included in CHI (e.g., a digital image of Mona Lisa and a virtual reality (VR) program of Angkor Wat). CHI includes metadata such as administrative data and external/additional information about the cultural heritage object. CHI may or may not be digital.

A digital surrogate is created from the original/primary cultural heritage object. It is a secondary instance and a sort of metadata, e.g., a digital image of Mona Lisa is a secondary instance/surrogate of the original Mona Lisa painting. On the other hand, digital surrogates which are born digital or converted into digital can be recognized as primary cultural heritage objects in a digital archive, e.g., inside a digital archive (as it consists of no real-world cultural heritage objects) so a VR program of Angkor Wat might become a primary object. This recognition does not present a significant problem for conventional digital cultural heritage archives which are created by digitizing original tangible objects. However, in the case of digital archives of intangible cultural heritage, there exists no original object for a digital resource as intangible cultural heritage is a conceptual entity. Therefore, records created for performances of intangible cultural heritage are digital surrogates of *Instantiations* of an intangible cultural heritage entity, e.g., a particular Thai "Khon" dance performance (which is intangible) may have many *Instantiations* which can be captured by various media and further realized as digital surrogates related to that *Instantiation*. This idea will be further investigated in Section 4.1.1 of this thesis.

Metadata. Metadata is generally defined as "data about data". Basically, an information object may consist of features such as content, context, and structure. All these features may carry with them some metadata that originated during creation management and use of the information object. According to NISO, metadata is "structured information associated with an object for purposes of discovery, description, use, management, and preservation" (NISO, 2007, p. 58). Meanwhile, NISO introduces six metadata principles to identify "Good Metadata". Good metadata: i) conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection, ii) supports interoperability, iii) uses authority control and content standards to describe objects and collocate related objects, iv) includes a clear statement of the conditions and terms of use for the digital object, v) supports the long-term curation and preservation of objects in collections, vi) records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification (NISO, 2007).

As Anne Gilliland says, "cultural heritage information professionals such as museum registrars, library catalogers, and archival processors often apply the term metadata to the value-added information they create to arrange, describe, track, and otherwise enhance access to information objects and the physical items and collections related to those objects. Such metadata is frequently governed by community-developed and community-fostered standards and best practices in order to ensure quality, consistency, and interoperability" (Gilliland, 2008, p. 2). However, embedding more metadata into a digital object should be done by the metadata creators as it is essential to understand and share an object effectively.

In this thesis, any CHI is considered as metadata. Digital surrogates (which are basically digital objects) such as a photograph of Mona Lisa and a VR image of Angkor Wat may be considered as a kind of metadata based on this definition. This research, however, uses metadata in a slightly narrower sense - a textual description of a cultural heritage object, that is, CHI expressed in a textual form such as plain text, XML texts, Excel sheets, relational databases, etc. This study categorizes metadata into two types, institutional and non-institutional metadata, explained below.

Metadata Description. A metadata description uses metadata elements to describe something; for instance, a record describing a heritage object via its properties and values. Requirements of a metadata description are diverse and depend on the standards, institutional requirements, context, content, etc. (Zeng & Qin, 2016). In addition, metadata descriptions

are based on some basic standards. However, in this research all the institutional and non-institutional CHI records are considered as metadata descriptions and the author is just reusing the existing metadata descriptions rather than recreating them again.

Institutional Metadata. According to this research, "metadata descriptions stored in the form of catalog records which are created, maintained and hosted by memory institutions can be considered Institutional Metadata". A metadata record of this type is generally created for a single item collected by a memory institution so that it is item-centric metadata. Digital archives created by memory institutions use this item-centric metadata for search and access.

Non-Institutional Metadata: Wikipedia articles and tourism websites can be categorized as non-institutional metadata. This type of metadata can exist at both item-centric and thematic (subject) levels. For instance, a Wikipedia article can describe a single heritage object or a specific theme. In most cases, non-institutional metadata are not based on well-standardized metadata schemas. However, some non-institutional metadata follow their own data standards (e.g., UNESCO Intangible Heritage Lists[11]).

Digital Archives: According to Sugimoto, a Digital Archive refers to "a collection of digital resources selected, collected, organized and maintained for long-term use" (Sugimoto, 2014, p. 62). A digital archive acts as a portal to disseminate information provided by a memory institution. Since digital archives of cultural heritage are created mostly by memory institutions, they mostly disseminate item-centric information. The content of digital archives may be born-digital or digitized resources. In this thesis, metadata of digital archive resources fall into the category of institutional metadata. A set of CHIs collected and organized for use is called a *digital archive of cultural heritage*.

As NISO defines, a "Good Digital" collection consists of the following characteristics. A good digital collection is: i) created according to an explicit collection development policy, ii) described so that a user can discover characteristics of the collection, including scope, format, restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity, and interpretation, iii) curated, which is to say, its resources are actively managed during their entire lifecycle, iv) broadly available and avoids unnecessary impediments to use while accessible to persons with disabilities, and usable effectively, v) respects intellectual property rights, vi) has mechanisms to supply usage data and other data that allows standardized measures of usefulness to be recorded, vii) supports

---

[11]https://ich.unesco.org/en/purpose-of-the-lists-00807

interoperability, viii) integrates into the users own workflow, and ix) is sustainable over time (NISO, 2007).

2.1.2. Related Concepts and Technologies

One-to-One Principle of Metadata. The One-to-One Principle of Metadata is a concept introduced by DCMI in the 1990s. DCMI defines this as a "principle whereby related but conceptually different entities, for example, a painting and a digital image of the painting, are described by separate metadata records" (Woodley, 2009). It implies one metadata description should describe or represent only one resource/object (Figure 1). The identification of the cultural heritage objects as an objective of the metadata description is crucial if the relationship between the heritage objects and their metadata is not One-to-One. Therefore, this research uses the One-to-One concept as a foundation to make this identification. However, there are some arguments and misconceptions (Urban, 2014) about the practice of the One-to-One principle. Miller specifically mentioned the practical usage, advantages, and challenges of using the One-to-One principle in cultural heritage institutions while digitizing and creating CHI related contents (Miller, 2010). Hence this research identifies it as a crucial concept for digital archives.
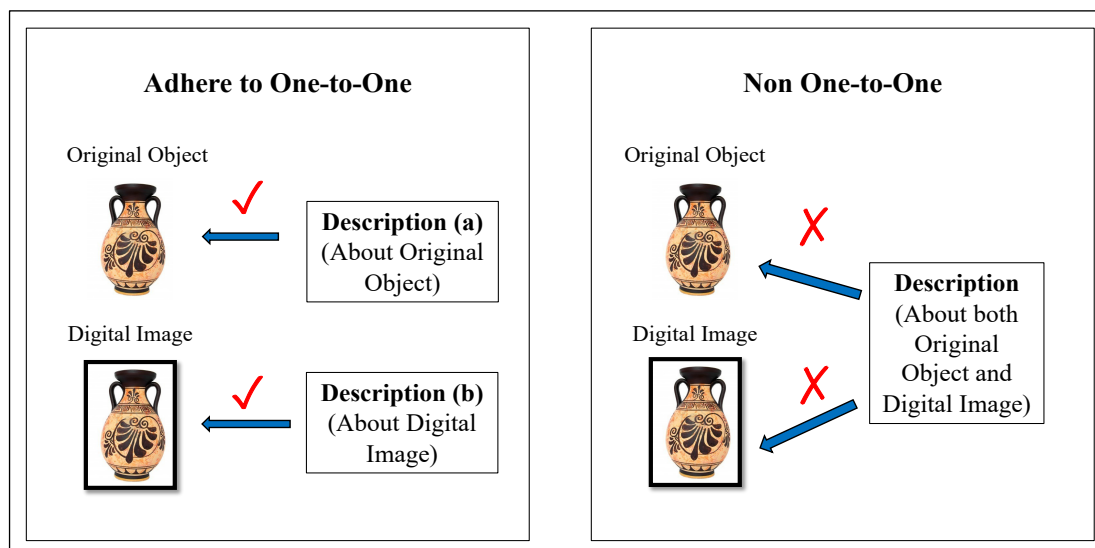
Figure 1. One-to-One Principle of Metadata

Dublin Core Application Profiles (DCAP). The Dublin Core Application Profile (DCAP) model outlines a general framework of metadata schemas in both semantic and structural

aspects. "DCAP defines metadata records which meet specific application needs while providing semantic interoperability with other applications on the basis of globally defined vocabularies and models" (Coyle & Baker, 2009). This research mainly used the structural aspect of DCAP, called a Description Set Profile (DSP), which has a hierarchical structure. DSP basically defines the metadata structure through its templates. Each of these templates has some specific constraints intended for the metadata. This research does not discuss the metadata creation but, the *Description Set* concept can be closely coupled with the CHI metadata records. For instance, a single CHI metadata record is a combination of metadata descriptions and it can be considered as a *Description Set* in the DSP model.

Metadata Modularity. Discussions of Metadata Modularity emerged within the Dublin Core community. In mid-1990s, the first discussions about modularization of metadata took place under the Warwick Framework which is a proposal for a container architecture for diverse metadata sets (Lagoze, 1996). Then, in another Dublin Core paper in 1998, there was a clustering of the 15 Dublin Core elements based on three classes; (i) elements related mainly to the content of resources, (ii) elements related mainly to the resources when viewed as intellectual property, and (iii) elements related mainly to the instantiation of the resource. This work is also a very good example of early modularization of metadata descriptions based on categories. As Carl Lagoze states "Modularity is the basis of metadata architectures such as the Resource Description Framework (RDF), which permit different communities of expertise to associate and maintain multiple metadata packages for Web resources" (Lagoze, 2001). Simply, a module of metadata is chunks of metadata which can be combined to create richer and more complex metadata descriptions. The basic structure is intended to capture most of the fundamental descriptive categories necessary to promote effective search and retrieval (Weibel & Miller, 2000).

Facets. *Facets* and *faceted search* are some important terms in this thesis. The term *Facet* implies "one side of something many-sided"[12]. This idea is being heavily used in organizing bibliographical information in libraries, e.g., Dewey Decimal Classification (DDC) and the origin of the idea goes back 300 years (Schulte-Albert, 1974). The *faceted classification* is used in organizing knowledge into a systematic order. *Faceted classification* utilizes semantic categories, either general or subject-specific, and later they are integrated to form a full classification list. The same idea was later adapted as a search paradigm known as *faceted search*. *Faceted search* is an approach which involves increasing the conventional search

---

[12]https://en.oxforddictionaries.com/definition/facet

techniques with a *faceted navigation* system, allowing users to narrow down search results by applying multiple filters based on *faceted classification* of the items (Tunkelang, 2009; Hjørland, 2013).

The same idea is being used in this research to identify the "cultural heritage objects via facets" and form a class-based mapping between the *objects* and *facets* forming a relationship.

Linked Open Data (LOD). Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. Technically Linked Data is data published on the Web and they are machine-readable, their external and internal links are well described, etc. Technologies that support Linked Data are URIs (Uniform Resource Identifiers), HTTP (Hypertext Transfer Protocol), and RDF (Resource Description Framework) (Bizer et al., 2009). Berners-Lee (2006), who is a pioneer in web-based approaches, describes a set of rules for publishing data on the Web in a way that all published data becomes a part of a single global data space. These rules can be recognized as Linked Data Principles. "Linked Open Data (LOD) is Linked Data which is released under an open license, which does not impede its reuse for free" (Berners-Lee, 2006).

Anyhow, unlike Linked Open Data, Linked Data does not have to be open. Since the focus is on online CHI, the term LOD becomes an important factor when enriching CHI information related to this study.

Resource Description Framework (RDF). RDF (Resource Description Framework) is a "standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed" (W3C, 2014). However, RDF can be identified as a formal and flexible technology capable of addressing a variety of problems. It was developed as a World Wide Web Consortium (W3C) specification. According to the RDF 1.1 Primer by W3C, reasons for using RDF are as follows.

- Adding machine-readable information to Web pages to enable them to be displayed in an enhanced format on search engines or to be processed automatically by third-party applications.
- Enriching a dataset by linking it to third-party datasets.
- Interlinking API feeds, making sure that clients can easily discover how to access more information.

- Using the datasets currently published as Linked Data.
- Building distributed social networks by interlinking RDF descriptions of people across multiple Web sites.
- Providing a standard compliant way for exchanging data between databases.
- Interlinking various datasets within an organization (W3C, 2014)

The RDF data model is based on three core object types (RDF triples) known as Subject (Resource), Predicate (Property/Relation) and Object (Resource/Literal). Through these triples can express any relationship and these triples can be connected. Figure 2 shows an RDF triple which shows a relationship between a Resource and a Literal.
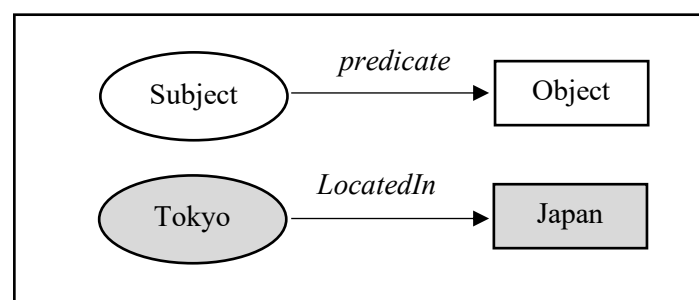
Figure 2. RDF Triples Example

Frequently used technologies/resources based on RDF are FOAF, DC, SKOS, schema.org and programming languages such as Turtle syntax of RDF languages, JSON-LD, RDFa and RDF/XML. Many professionals in various backgrounds use RDF technology to describe their resources and the CHI sector also utilizes the same technology when dealing with Web-based semantic research. The most prominent example is the Europeana Data Model (EDM) of the Europeana project. Europeana uses the RDF graph and RDF Syntax (e.g. Turtle and RDF/XML) to describe their model (Europeana Data Model Primer, 2013).

2.2. Issues Related to the Cultural Heritage Information

2.2.1. General Issues

CHI is the primary factor which makes the cultural heritage meaningful and usable. CHI can be in various forms and can be recorded in various ways. Many scholars identify CHI as a unique type of information. According to Lanzi (1998), CHI has ten characteristics. Similarly, Hyvönen (2012) defines five features of cultural heritage data.

i. Multi-format: contents are presented in various formats

ii.   Multi-topical: contents concern various topics

iii.   Multi-lingual: content is available in different languages

iv.   Multi-cultural: content is related and interpreted in terms of different cultures

v.   Multi-targeted: contents are targeted to different people (Hyvönen, E., 2012).

Due to CHI diversity, institutions use different standards (e.g., CARARE Metadata Schema[13], CDWA (Categories for the Description of Works of Art)[14], LIDO (Lightweight Information Describing Objects)[15], SPECTRUM[16], etc. to create and organize their information which leads to data interoperability issues in the future (Figure 3). Correspondingly, there are many local standards developed by each country depending on their own institutional requirements. Alternatively, Data standards used by other domains, for example, Dublin Core, MODS (Metadata Object Description Schema) and VRA (Visual Resources Association) Core Categories, etc. are also utilized by the CHI domain where necessary. To give an instance, the Cultural Heritage Metadata Task Group of Dublin Core Metadata Initiative (DCMI) tried to identify the challenges of metadata for cultural heritage by developing a simple cross-community metadata model for cultural heritage objects.
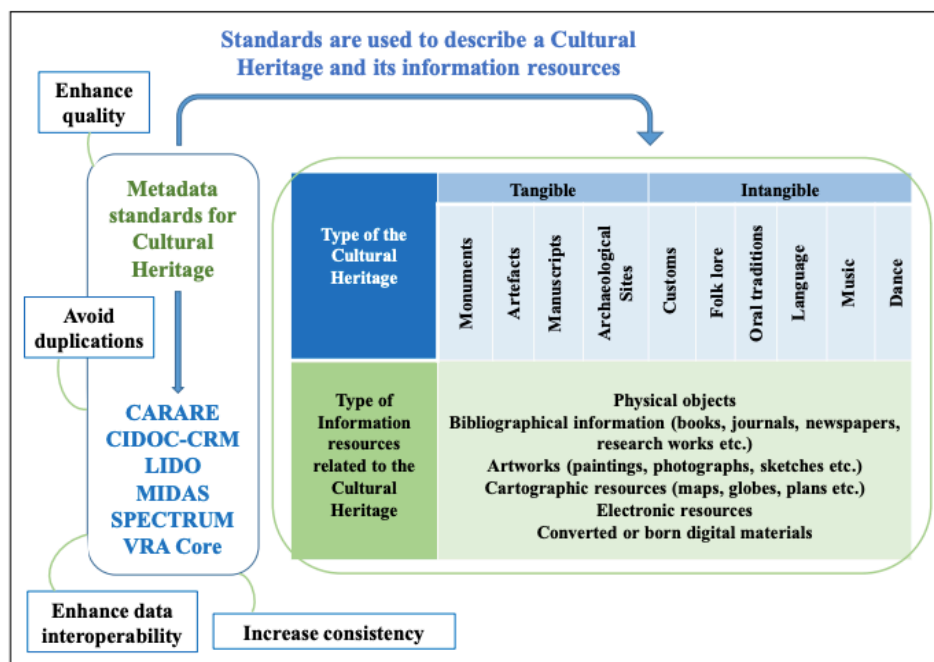
Figure 3. Different Types of Cultural Heritage Objects, Their Information Resources and its Connection Between the Metadata Standards (Wijesundara, Sugimoto, & Narayan, 2015)

---

[13]http://pro.carare.eu/doku.php?id=support:metadata-schema

[14]http://www.getty.edu/research/publications/electronic_publications/cdwa/

[15]http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/

[16]http://collectionstrust.org.uk/spectrum/

Besides they intended to give a recommendation for the development of DCMI Application Profiles based on the task above[17].

In addition, there are data models specifically designed for CHI which can be used to organize data and define their relationships with real-world entities. Some of these well-known CHI models will be discussed in Sections 3.2 and 3.3 of this thesis.

Nevertheless, none of these standards could completely cover the entire cultural heritage domain to describe its properties. Previously mentioned standards are developed to capture tangible objects only. Thus, there is a gap between tangible and intangible heritage data standards which has yet to be filled. When considering the intangible cultural heritage, it is difficult to express intangible assets as individual items. Similarly, intangible cultural heritage can be realized only if it is recorded. Memory institutions cannot curate a concept such as a skill or a performance related to intangible cultural heritage, but they can use various media to capture the intangible heritage and record them as individual records.

Whether there is a deviation of tangible and intangible heritage, these assets are interrelated sometimes. Unfortunately, current CHI on the Web provided by various means does not deliver such contextual information to patrons. As a result, we need a scheme to cover the heterogeneity by linking cultural heritage objects of different kinds.

According to the author, another critical point that can be seen in conventional digital archives is the scarcity of LOD. This problem is partly associated with the heterogeneous nature of the CHI as well. There are some efforts which provide their CHI in LOD friendly schemes, e.g. Europeana and British Museum Online. However, many of the institutional digital archives stay as isolated information silos and they tend to describe only their own information .

Long-term Use of CHI. The longevity of digital resources and digital archives is a well-known and important issue. A proper scheme which keeps track of changes in the digital collections is mandatory for digital archives. This scheme heavily relies on the metadata used by the digital archives. Clearly defined relationships between metadata and its objectives are crucial to make the maintenance process simpler.

2.2.2. Issues Related to the Cultural Heritage Information Aggregation

When considering the LAM environment, they mostly record CHI related to a single object.

---

[17]http://dublincore.org/archive/mediawiki_wiki/Cultural_Heritage_Metadata_Task_Group/

For instance, a museum record describes a single object, with all the properties and values which corresponds only to a single item.

According to Zeng and Qin (2016, p. 93), "when discussing levels of granularity, the "term item-level" (in contrast to "collection-level") is often used to refer to the individual objects that, together, constitute a whole collection". This item-centric feature came from the long tradition of management of museum holdings. In general, this item-oriented perspective is a benefit for knowledge organization, and it provides for convenient and easy user interfaces too. However, user needs are not always concentrated on a single item, especially if they need more complex and linked information associated with a heritage object.

On the other hand, "hybrid records" (Woodley, 2016, p. 44) of CHI provided by digital archives are confusing as the objectives of the objects and their metadata are not clear. As a result, users sometimes cannot distinguish between the original objects and their digital surrogate information in a digital archive. Figure 4 shows some metadata related to the original heritage object (vessel) and its digital surrogate (digital image) put together as a single metadata record (This example is taken from Europeana[18] and the same example will be used in several other places in this thesis).

Property/ Attribute	Value
Title	vessel
Description	A glazed clay vessel, *hattiya* or *muttiya* (Sin.), used as a kitchen utensil to boil water all over Sri Lanka…
Image	
Format	image/jpeg
File Size	59 KB
Type	vessel
Size	Hight 12 cm, Diameter 17.5 cm
Date	Late 19th c. A.D.

Figure 4. Instance of a Hybrid CHI Record

Hence, the purpose of the metadata becomes unrecognizable and complicated. Yet, identification of objects, objectives and their metadata are very important when collecting metadata from different sources and aggregating them into a single database. This issue is connected with the One-to-One Principle of Metadata as well. The One-to-One Principle of

---

[18]https://www.europeana.eu/portal/en/record/91619/SMVK_EM_objekt_1024261.html

Metadata is utilized as a foundation in this research and the author tries to differentiate and identify the CHI metadata the same way. As CHI records are hybrid, the One-to-One Principle helps to distinguish and separately identify the objectives of the metadata.

In metadata aggregation, merging of metadata schemas is common, and known as a *crosswalk*. According to NISO (NISO, 2004, p.11), a crosswalk is a "mapping of the elements, semantics, and syntax from one metadata scheme to those of another." This practice begins with independent schemas and then efforts are made to map or to create a crosswalk between equivalent or comparable metadata elements enabling data interoperability. These metadata elements are known as *properties* in this research followed by the RDF definitions. However, the same term is referred to as an *attribute*, *field* or *label* in other domains. The basic *crosswalk* is a direct mapping or an equivalency among properties from diverse schemas. This mapping can be done as an *absolute crosswalk*; exact mapping between properties or a *relative crosswalk*; map all the properties to at least one property or a targeted vocabulary, regardless the equivalency of the properties (Zeng & Qin, 2016).

The *property-by-property* level mapping denotes this *absolute mapping,* and this is the one-to-one mapping method. However, this needs equivalent (or closely matching) properties. When there are no matching property means, there is no crosswalk. "Metadata mapping is often done based only on properties. Property-level mapping has risks of losing the context of properties given in the schema in which the properties are included, such as mandatory levels and value types" (Sugimoto et al., 2018, p.97). Therefore, these non-mappable elements are left out and as a result, valuable data might be missing from the mapping. Table 1 shows such a scenario. There are three source schemas (British Museum[19], Metropolitan Museum of Art[20] and Europeana[21]). These three institutional digital archives have similar properties such as *Title, Description*.

E.g., British_Museum: Museum Number ≈ CIDOC-CRM: E42 Identifier

Europeana: Creation Date ≠ CIDOC-CRM: E50 Date

(Approximately equal- ≈, Not equal- ≠)

These properties are mapped to the CIDOC-CRM entities to create a crosswalk. Still, it is obvious that some properties (marked in RED color) cannot be fully mapped to the targeted schema.

---

[19]https://www.britishmuseum.org/research/collection_online/search.aspx
[20]https://www.metmuseum.org/art/collection/
[21]https://www.europeana.eu/portal/en

Table 1. Crosswalk Difficulty Between Diverse Schemas

Source Schema			Target Schema
British Museum	**Metropolitan Museum**	**Europeana**	**CIDOC-CRM**
Title	Title	Title	E35 Title
Museum Number	Accession Number	---	E42 Identifier
Date	Date	Creation Date	E50 Date
Dimensions	Dimensions	Size	E54 Dimension

Dealing with CHI means, dealing with heterogeneous metadata schemas. Applying *property-by-property* level mapping to some diverse cultural heritage schemas makes mapping very complex and ambiguous. On top of this, when the number of schemas grows, the property mapping combinations also grow exponentially, resulting in an uncontrollable crosswalk. "Each metadata schema has its advantages and challenges, and no two schemes are created 100% equal. Whereas having two fields that are consistent across the schemes and can be fully mapped from one schema to another (one-to-one mapping), however rare, works well in metadata crosswalks, mapping one field in one schema to several fields in another and vice versa can result in ambiguity and, consequently, reduce the accuracy of search results. Finally, having to omit certain fields from the source schema because the target schema is less granular than the source schema and does not accommodate for inclusion of respective fields, ultimately results in data loss" (Gogina, 2016, p. 17).

2.3. Metadata Aggregation in Cultural Heritage Domain

This study aims to define conceptual metadata models for building digital archives of heterogeneous cultural heritage objects by aggregating CHI collected from various sources. Therefore, metadata aggregation is a key aspect of this research. Metadata schemas of the digital archive are often based on well-known standards and best practices in order to ensure the quality, consistency, and interoperability of data which is an important factor when it comes to metadata aggregation (Gilliland, 2008).

Simply, metadata aggregation is linking or connecting different metadata through their relationships. The previously discussed technologies such as crosswalk, RDF and LOD can be incorporated in the metadata aggregation process. Swan & Awre (2006) in their research called

Linking UK Repositories, outline the benefits of metadata aggregation as follows.

- Aggregations offer a breadth of access across many repositories, relieving end-users from accessing each one individually.

- Aggregations provide a single point of access to multiple sources of research and other materials to aid discovery.

- Aggregations offer an alternative route for enhancing metadata held within a repository.

- Aggregation can provide preservation and metadata enhancement capabilities to support the long-term storage of and access to the content, etc.

Metadata aggregation is associated with many fields such as library information, computer information, geographic information, etc. This is a reliable approach to linking information with diverse standards while sharing and enriching it. Other than data sharing, metadata aggregation allows one to enhance and create more complete resource profiles by aggregating both complete and fragmentary metadata from many resource providers (Hillmann, Dushay, & Phipps, 2004).

The requirement for metadata aggregation can be either i) Data access services: providing unified machine interfaces for searching and linking metadata held in different types of repositories through the use of standardized access protocols, or ii) Data mapping services: repurposing metadata of various schemas from disparate sources into formats which are coherent and of consistent quality for specific use contexts (Low, 2006). In this thesis, the focus is on the data mapping services and so uses a metadata model approach as the baseline technology.

There are well-known models for metadata aggregation such as Open Archives Initiative-Object Reuse and Exchange (OAI-ORE)[22] and the Europeana Data Model (EDM) (Europeana Data Model Primer, 2013). Both models are defined primarily for digital objects which are organized with conventional institutional metadata. These models define data models of aggregated metadata, but not the metadata creation and aggregation process. The metadata creation process depends on factors such as if it is tangible or intangible cultural heritage, while it being born-digital or digitized may affect the metadata aggregation process.

There are several fundamental components which have to be taken into account for metadata aggregation: identification of objects described by metadata (i.e., objectives of metadata description); identification of purposes of metadata description such as administrative, technical, and descriptive; underlying data models and metadata schemas of

---

[22] https://www.openarchives.org/ore/

metadata instances to be aggregated which define syntactic, structural and semantic features; and metadata interoperability schemes to make metadata aggregatable such as vocabulary mapping. A simple example of identification of objectives is identifier assignment to cultural heritage objects. On one hand, memory institutions have identification schemes for physical items curated as cultural heritage. Instead, digital archives need to identify every digital surrogate created from a particular cultural heritage object. Compared with tangible objects, identification of intangible cultural heritage may not be so simple. Identification of objectives is related to the identification of purposes. For example, descriptive metadata about a cultural heritage object is used to find and access the object, descriptions about stakeholders of the object may be used to know the roles of the stakeholders, and so forth. Thus, identification of objectives and purposes is crucial. However, in many cases of conventional metadata schemas, a single metadata record describes more than one objective, and it is not obvious which part of the metadata describes what objective. From this point of view, the One-to-One Principle of Metadata is a useful underlying concept for identifying different objects and their descriptions separately.

The Dublin Core Application Profile (DCAP) (Coyle & Baker, 2009) provides a framework for metadata interoperability, which is also useful for metadata aggregation. According to DCAP, an application profile defines a metadata schema where a metadata instance consists of one or more descriptions, and each description is a set of statements. Metadata terms used in the statements should be defined separately from an application profile. Metadata of digital cultural heritage object is usually complex, consisting of descriptions of different objects such as an original cultural heritage object, its digital surrogates, stakeholders, access conditions, and so forth. Hence DCAP provides a framework to identify the objectives of descriptions. Clear identification of the objects described by metadata is fundamental for aggregation.

The thesis proposes conceptual metadata models as a foundation for the organization and aggregation of heterogeneous data sources. According to scholars, information aggregation enables a global view of diverse information contents, semantic searching, linking and sharing content, data enrichment, data reuse and longevity of information, etc. (Hyvönen, 2012). The proposed models are designed to create such aggregation, enabling many functionalities required for digital archives of the tangible and intangible cultural heritage. Thus, the model proposed in this thesis is neither a model for application metadata nor one to express an aggregated metadata but is defined to systematically connect both ends.

2.4. Requirements for Metadata Models

Organizing and connecting dispersed CHI into a single platform using a metadata model is the main goal of this research. The metadata model for a cultural heritage digital archive must contain features that will provide users with efficient services to access and retrieve data about the information objects either by browsing the collections or by searching those using keywords. The metadata model needs to be focused both on the collection level and item level. (Lourdi, & Nikolaidou, 2009). In addition, identifying and aggregating intangible cultural heritage resources is also a crucial aspect of this study.

This study was started with two fundamental research questions and some requirements based on the research questions which arose from discussions of digital archives of cultural heritage for South and Southeast Asia. The two research questions and its requirements follow.

Research Question 1: "How can we model metadata for digital objects to be created by aggregation of fragments extracted from existing digital archives and other Web resources?"

Requirements:

i) Identification of facets required for aggregation of institutional and non-institutional CHI: It is important to provide information about a cultural heritage object together with contextual information about the object for users who want to learn about it and its cultural contexts. This linkage between the two types of information can be realized by linking item-centric institutional CHI and general-description-oriented non-institutional CHI. Basic requirements for linking this very different CHI need to be clarified.

ii) Identification of objectives of metadata aggregation, using the One-to-One Principle of Metadata as a foundation: It is essential to identify metadata description objectives individually, that is, identifying the original cultural heritage object and its surrogates as separate entities. However, current digital archives and other Web services tend to provide mixed descriptions (Hybrid Records) of more than one object. Clear identification of the relationship between a metadata description and its objective is important for metadata aggregation. The One-to-One Principle of Metadata concept is a reasonable foundation to satisfy this requirement. This thesis clarifies how the One-to-One Principle can be applied to cultural heritage resources on the Web through the discussions of the two proposed metadata models.

iii) Identification of "Entities" in the process of organizing CHI: It is essential to identify entities which work in the process of organizing CHI into a digital archive. The process models (Figure 7) help us identify those entities and understand how we can create, collect and aggregate CHI in accordance with those entities.

Research Question 2: "How can we describe intangible cultural heritage for digital archives?" Conventional digital archives provide digital records of intangible cultural heritage such as dance performance and music performance. However, a single performance is not a particular intangible cultural heritage entity, but we have to use performance which is physically shown to record intangible cultural heritage.

Requirement:

i)   Identification of an "Object" of intangible cultural heritage: What objects of intangible cultural heritage archived in a digital archive have to be identifiable? This is a fundamental requirement for digital archives and databases which have to record temporal objects such as events, performance, installations, and so on. This requirement is crucial even in the case of tangible cultural heritage because they may be lost by natural and/or man-made disasters.

The thesis is trying to answer these main research questions and requirements based on its proposed model/s.

2.5.   Objectives and Novelty of the Research

This study proposes a metadata model (CHDE model) to collect diverse CHI on the Web. The model explicitly distinguishes the physical and digital spaces/environments of cultural heritage objects and its information which cannot be clearly seen in the existing CHI aggregation or organization models. However, this study gives a special focus on the Web-based CHI in the networked information environments.

In addition, the model is designed to aggregate both institutional item-centric information and non-institutional information. Many existing CHI models intended to aggregate just one type of information, e.g., EDM - Institutional information. The proposed model aggregates both types of information enabling more context for the cultural heritage objects.

The identification of intangible cultural heritage via a special entity called *Instantiation* is another novel approach proposed by the CHDE model. Since the knowledge organization

of intangible cultural heritage information is not well discussed in the cultural heritage domain, the proposed model is trying to provide a new solution to fill the gap.

The author researched widely on the metadata aspects of the CHI as it is the key for the CHI aggregation. Identifying CHI based on the One-to-One Principle of Metadata is fundamental when building a digital archive of cultural heritage. The second half of the thesis considers the role of the One-to-One Principle in the Description Modules model which is intended specifically for CHI metadata identification and aggregation. Connecting and disseminating information on the Web is not a new idea but a metadata model that adheres to the One-to-One Principle provides a clear foundation for connecting and aggregating heterogeneous CHI on the Web. It will help memory institutions in South and Southeast Asia build digital archives of their regional cultural heritage.

A complete compression between the proposed CHDE model and the existing models will be separately discussed in Section 3.3 of this thesis. The novelty of the proposed models and what is lacking in the current models will be discussed in more detail with supporting examples in Section 3.3 of Chapter 3.

Chapter 3: Literature Review

As this study is focusing on organizing and aggregating fragmented CHI on the Web, this chapter focuses on discussing each aspect separately. Further, Section 3.3 tries to differentiate the proposed models from the existing models utilized in CHI aggregation and organization.

3.1.   Cultural Heritage Information Modeling and Organization

Various studies have been conducted on data models for CHI organization and aggregation.

Hu et al. (2014) used CRM as a framework for describing the Pang Wang Festival in China and provided analysis for mapping their data to CIDOC-CRM. Similarly, Tan et al. (2009) constructed an ontology model based on CIDOC-CRM to represent the Dragon Boat Festival, China. Finally, they proposed a Browser/Server architecture to implement a prototype, which involves several key functionalities such as semantic knowledge retrieving. Unfortunately, neither of them discusses the significance of the festivals or defines any contextual elements that describe the intangible asset.

Chen et al. (2013) used FRBRoo as an ontological approach to aggregate diverse CHI metadata and transform it from a human-understandable format to a machine-understandable format for semantic query. They have collected data (accompanied by Dublin Core terms) from two collections and mapped them into the FRBRoo classes and properties to make heterogeneous metadata integration possible.

Smiraglia (2005) tried to model artifacts in museums using the *Work* concept in the FRBR (Functional Requirements for Bibliographic Records) model. He used Etruscan artifacts from the University of Pennsylvania Museum of Archaeology and Anthropology to demonstrate the connection of the *Work* concept in non-documentary artifacts. The model he presented, called the Content Genealogy Model, tries to conceptualize the representations of museum artifacts and their metadata. Both of these studies are focused on work level CHI aggregation, but they used different approaches.

Carboni and Luca (2016) analyzed the dichotomy between tangible and intangible heritage and proposed a way to document the same. They used CIDOC-CRM as the base, and modeled information using a use case to show that a cultural object has multiple facets and dimensions that incorporate both tangible and intangible elements.

Some scholars have proposed their own models for organizing and describing heritage resources. Amin et al. (2012) proposed one such model called a knowledge repository model

for intangible cultural heritage as a framework and guideline to archive Malay Intangible Culture Heritage in Malaysia. The proposal was to digitize intangible to tangible heritage without losing their originality and archive them using the proposed model.

Creating metadata schemas helps CHI organization and it formalizes and standardizes the aggregation. For instance, a study conducted by a group of researchers shows creating a new metadata schema to describe intangible cultural heritage. They have developed a schema to capture metadata related to folklore dance which is necessary to describe a particular dance. As intangible cultural heritage metadata schemas are rare, they wanted to investigate problems associated with their proposal and implementation based on the same proposal as the next step of their research (Giannoulakis et al., 2018).

Collection development in CHI is also an important aspect in CHI organization. As Low and Doerr stated, most institutions like museums privately maintain collections of documents or other objects that relate to a specific theme or activity, which are often referred to as "folders" (Low & Doerr 2010; Doerr et al. 1997). They further study the internal and external knowledge collection and transfer processes of several museums. Based on their study, digital representations of museum collections for research purposes and the general public should be distinctive from the conventional institutional documentation practice.

Geisler et al. (2002) proposed the development of "virtual collections" within digital libraries, which were conceptualized as sub-collections of digital library collections based on a common attribute or relation to a common subject. But these approaches are not restrictive enough. For instance, they could not necessarily distinguish a group of items retrieved through an online search from the collections that are developed by libraries, archives, museums through a systematic selection of items, or the research collections created by scholars, etc.

The studies above followed different approaches to organize CHI, some of which are item-centric information organization, and some are focused on specific kinds of cultural heritage objects. The proposed CHDE model tries to clearly identify the physical and digital objects via the One-to-One Principle of Metadata. In addition, the model proposes a novel information organization approach specifically for intangible cultural heritage.

3.2.   Cultural Heritage Information Aggregation

CHI aggregation is a well-known approach for heritage domain specialists to help resource discovery and to enable data interoperability.

A group of researchers (Freire et al., 2017) surveyed and conducted a preliminary study of Web technologies used in the contemporary cultural heritage information aggregation domain. The authors identified OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) that has a long history in the aggregation of cultural resources. Though, with the technological advancements and computational capacity, etc. the motivation and urge of utilizing OAI-PMH have become uncertain. Instead, the authors introduced new approaches such as IIIF (International Image Interoperability Framework), Webmention, Linked Data Notifications, Sitemaps, ResourceSync, Open Publication Distribution System (OPDS); and the Linked Data Platform as potential technologies in the domain. This study gives an overview of the present CHI aggregation technologies used across the globe.

Scholars such as Orgel et al. (2015) proposed a data model that combines federated search results from different cultural heritage data sources. They then proposed an approach for metadata mapping, with a focus on easy configurability of mappings, which—once properly configured—can then be executed on the fly by an automatic service (Orgel, Höffernig, Bailer, & Russegger, 2015). The system they developed is known as EEXCESS (Enhancing Europe's Exchange in Cultural Educational and Scientific reSources). The final aim of the system is to harmonize and enrich cultural heritage resources enabling enhanced search functionality.

Wickett et al. (2014), discuss cultural heritage collections in digital aggregation and exchange environments. In their proposal, they developed a set of general requirements for the representation of collections in digital aggregation systems and used EDM as a base model to support their idea (Wickett, Isaac, Doerr, & Fenlon, 2014). This research is supporting the idea of collection level resource aggregation and used a metadata model approach.

Another team of researchers proposed a cultural heritage monumental and architectural property aggregation approach which is known as CARARE (Connecting Archaeology and Architecture in Europeana) (Papatheodorou et al., 2011). CARARE facilitates a mapping approach and aggregates heterogeneous resources related to the archeological assets and monuments in the Europeana collection. As archeological and monuments assets consist of complex information resources - for instance, 3D models, section drawings and also, they include critical metadata such as geographic coordinates, place names, there is a need to aggregate this information by a unique way. CARARE fulfill this difficulty facilitating new projects such as 3D ICONS which digitize a series of architectural and archaeological assets and provide 3D models and related digital content to Europeana.

DPLA (Digital Public Library of America)[23] is another well-known example of metadata aggregation. DPLA aggregates existing metadata from libraries, archives, and museums to enable users to search and find collections and individual items. The resources of the DPLA vary from print to digital media and they provide a single point of access to millions of resources around the United States. In addition, DPLA-API provides access to metadata of the resources and all these data are freely available to the users (Guthro, 2013).

Signore (2008) discussed CHI on the semantic web. He argues that metadata level aggregation is not enough to fulfill the current trends and it should be replaced by a core ontological approach. He further lists issues related to CHI and related applications and then tries to introduce ontological and semantic web approaches for information integration.

A study by Peroni et al., (2013) stated issues in digital repositories which aggregated cultural heritage resources and inadequacy of data models in the domain. They used Europeana collection as the base and discussed the metadata standards used in Europeana, the necessity of richer aggregation data model such as; the 'media type' concept, the multi-layer description, and the connection between roles and values, etc. In the end, they analyzed the sophistication of the current implementation of Europeana with regard to their Linked Data offering.

Freire et al. (2017) discussed using IIIF (International Image Interoperability Framework) which is a standardized method of describing and delivering images over the Web, and Sitemaps which is a way of organizing a website, identifying the URLs to aggregate cultural heritage metadata on the Web. They use EDM as the base aggregation model and conducted some case studies based on the above technologies to conduct this aggregation.

The "Sampo" project uses another interesting approach which is designed for aligning metadata models, and sharing domain ontologies for populating the metadata models. The final aim of this project is to realize data in Linked Open Data (LOD) formats. There are many types of "Sampo" models such as CultureSampo[24], BibliographicSampo[25], WarSampo[26], etc. All these models aggregate cultural heritage resources forming a large platform which disseminates information related to Finish cultural heritage in Finland (Hyvönen, 2016).

---

[23]https://dp.la/

[24]http://www.kulttuurisampo.fi/?lang=en

[25]http://biografiasampo.fi/

[26]https://www.sotasampo.fi/en/

In addition, a technical report by CIRISS comprehensively discusses a collaborative effort of modeling digital collections on cultural heritage for aggregation. They discuss the usefulness of collections such as improvement of user search experiences and the provision of contextual information about cultural heritage resources, etc. The goal is to enhance the representation facilities of EDM, and to make EDM suitable for representing collection-level data from DCC (Digital Collections and Content) and other digital content providers (Wickett, 2013).

The author would like to introduce one last initiative known as Open Archives Initiative Object Reuse & Exchange (OAI-ORE)[27] as a standard for the description and exchange of aggregations of Web resources. OAI-ORE proposes the ORE Model which uses Resource Map (ReM) to make the relationship between resources semantically. ORE aggregation supports connecting dispersed resources with multiple media types such as text, images, data, and video (OAI-ORE, 2008). EDM along with OAI-ORE defined structures of metadata collected and aggregated from multiple metadata sources. As a result, EDM uses all the main classes and properties in ORE model. Scholars stated that "the core data structure of EDM, for instance, is based on the OAI-ORE, a reference model for the description and exchange of aggregations of Web resources. ORE aggregations are used to represent a data provider's contribution to Europeana, which consists of the "provided item" together with its digital "view(s)" (modeled as *web resources*)" (Haslhofer, & Isaac, 2011, p. 98). Besides, NSDL (National Science Digital Library)[28], National Diet Library: Great East Japan Earthquake Archive (ひなぎく)[29] is a famous digital archive which uses the ORE model for their data aggregation.

3.3. Current Models Compared with the Proposed Model: What is Unsolved and What is New?

This section discusses some features related to the existing CHI models—EDM, CIDOC-CRM, and FRBRoo- in contrast to the proposed CHDE model.

EDM, CIDOC-CRM, and FRBRoo are intended to aggregate and organize CHI. However, before discussing issues related to each model it is better to introduce them very briefly.

---

[27]https://www.openarchives.org/ore/
[28]https://nsdl.oercommons.org/nsdl-overview
[29]http://kn.ndl.go.jp/#/

Europeana Data Model (EDM). Europeana Data Model (EDM) acts as a typical CHI aggregator in the cultural heritage domain. EDM aggregates metadata from 3,000 cultural heritage institutions all over the European Union and enriches them further for better accessibility. The aggregated content is disseminated via the Europeana Collections which host nearly 6 million artifacts, books, films, music information, etc. as digital content. According to the EDM primer, "EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO for museums, EAD for archives or METS for digital libraries" (Europeana Data Model, 2013, p. 5). The model is created using RDF and uses classes (e.g., *edm:ProvidedCHO* for provided cultural heritage objects) and properties (e.g., *edm:hasView* for one or more resources that are digital representations of the provided object), plus RDF syntax in Turtle to describe their model semantically.

CIDOC- Conceptual Reference Model (CRM). The second prominent model is CIDOC-CRM by the International Council of Museums (ICOM). CIDOC-CRM "provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation" (Le Boeuf et al., 2015). This model takes an ontological approach, and it consists of a comprehensive set of classes and properties to describe artifacts, monuments and any form of cultural heritage entity.

FRBRoo (Functional Requirements for Bibliographical Records - object-oriented). FRBRoo is known as an extension to the CIDOC-CRM, and it is another important model in the CHI domain. FRBRoo basically uses the Work, Expression, Manifestation, and Item (WEMI) entities of FRBR and adds some more classes and properties aligned with CIDOC-CRM. The main aim of FRBRoo is to integrate bibliographical information related to cultural heritage and facilitate library and museum information harmonization (Chryssoula et al., 2015).

Untangling the mixed nature of the CHI records and distinguishing CHI according to their objectives is a main focus of this research. Basically, South and Southeast Asian memory institutions do not have a common base to share and aggregate CHI as Europeana does. As a result, CHDE uses a bottom-up approach for metadata aggregation which relies on the existing Web-based CHI provided by memory institutions and non-institutions. A fundamental issue in metadata aggregation is to identify every single object to be described by a metadata record or presented as a CHI.

As introduced previously, digital archives of memory institutions organize CHI records as individual items. Data models such as EDM aggregate digital surrogates related to a heritage entity. This aggregation is an item-to-item information aggregation (Figure 5). EDM is considered as a top-down approach where data providers submit their CHI as packages to the Europeana in a form conforming to EDM. According to EDM, Mona Lisa by Leonardo da Vinci is a single item (known as *edm:ProvidedCHO*) and its digital images may be collected from more than one institution. These are collected using *edm:hasView* and *edm:aggregatedCHO* properties (Europeana Data Model Primer, 2013). This method adheres to the One-to-One Principle of Metadata in the CHDE model. However, the cultural heritage objects provided by these institutions are not 100% reliable. Although they represent individual heritage items, the information provided sometimes consists of a mixture of original heritage information and digital surrogate information as a single record. This mixed nature of CHI records and difficulty of understanding the objectives of metadata can be further described as follows. According to Orgel et al. (2015), EDM failed to provide provenance information related to the annotations they created for the digital cultural heritage in multiple views. Another paper states that the "distinction between provided objects (painting, book, movie, etc.) and their digital representations" is a core principle in the EDM. However, "although the Europeana core classes stress the difference between the provided object (*edm:ProvidedCHO*), i.e., the "real object", and its digital representation (*edm:WebResource*), i.e., its Web resource, sometimes this difference is not evident in the aggregated metadata exposed to the final user, generating confusion. Sometimes the description seems to be addressed to the electronic version, some other to the original work, without a clear distinction" (Peroni et al., 2012).

The models proposed in this thesis identify each metadata component in a CHI record based on the One-to-One Principle of Metadata. The Description Modules model acts as the foundation to do this identification and it acts as a separate model which lies behind the CHDE main model which is different from the EDM scenario. Therefore, it gives a solution to the non-One-to-One CHI problem while aggregating CHI from diverse Web resources.

The inclusion of web-based resources provided by non-institutions is an important aspect in the proposed CHDE aggregation scenario which is out of the scope of the Europeana. Figure 5 describes a comparison of the Europeana aggregation with the CHDE. Europeana tries to collect similar Web resources created and provided by institutions. Nevertheless, the CHDE model tries to aggregate different Web resources (both institutional and non-institutional) with the help of the *Description Modules*.

According to the preliminary studies, neither EDM nor CIDOC has specific entities designed for expressing intangible cultural heritage. Since an intangible cultural heritage entity is not an item physically collectible by memory institutions, the item-centric resource aggregation is not suitable. The CHDE model provides a solution via *Instantiation* as a bridge to aggregate those resources related to intangible cultural heritage. This *Instantiation* acts as a specific aggregator which forms a collection/set of resources related to a specific intangible cultural heritage entity.
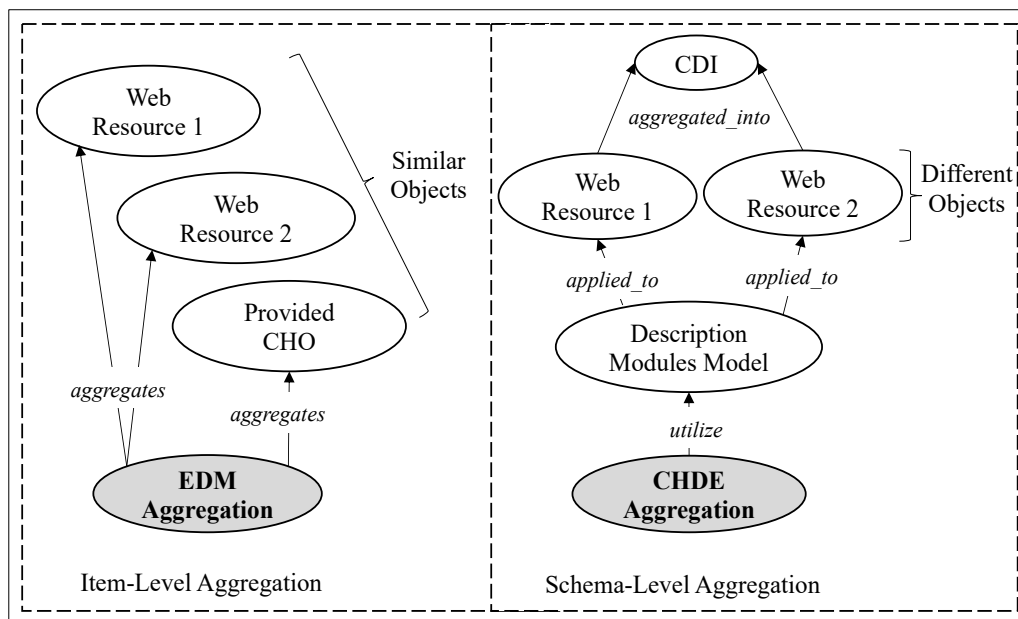
Figure 5. EDM and CHDE Aggregation Levels

The author has mapped CHDE classes to those in CIDOC-CRM and FRBRoo to check the compatibility of the models (Appendix 1). Later the study did a comparison between the proposed models with the EDM, CIDOC and FRBRoo ontologies as follows (Table 2).

Additionally, the Metadata Encoding and Transmission Standard (METS) can be presented as another model/ schema related to this research. METS is an extensible, XML based schema designed for encoding descriptive, administrative, and structural metadata regarding objects within a digital library. For instance**,** objects in digital libraries include fields that differentiate between certain stewardship roles involved in the maintenance and dissemination of digital objects. METS identifies this and introduces a container-oriented structure and has sections in accordance with the categories, such as *METS Header, Descriptive Metadata Section, File Section, Behavior Section,* etc. (METS, 2010). In addition, METS is particularly designed to capture information about many entities responsible for a METS document such as preparing metadata for encoding, for the document

or collection being described, for preservation functions, and for dissemination functions and so on. However, these fields attend more directly to recording information about metadata records than cultural resources themselves (Wickett et al., 2013).

The Modular Metadata model discussion in Chapter 5 is trying to reveal a similar approach specific for cultural heritage resources. This research prioritizes identifying metadata descriptions of digital and physical cultural heritage objects based on the modules like container-oriented structure which is similar to the METS approach. However, the *Description Modules* are more dynamic, and its requirements are distinct from the METS Schema.

Table 2. Comparison Between the Proposed Model, EDM and CIDOC-CRM/ FRBRoo

Proposed Model (CHDE)	EDM	CIDOC-CRM/ FRBRoo
Enables both item-level and collection-level aggregation	Item level aggregation only	More towards item level aggregation
Relationships are made through specific instances, e.g., *Instantiation*	Relationships are centralized on the *Provided CHO*	Relationships are created based on individual instances/ situations
No data providers	Have data providers	
Depends on existing Web resources	Digital data are submitted as packages	Priority is given to the offline CHI
No object ID	Have identifications for every *Provided CHO*	
More towards contextual information aggregation	Contextual information aggregation is not clear	Contextual information aggregation is not visible
Has a bottom-up approach	Well organized and has a top-down approach	
For Asian resources (can be extended)	For European regional resources	For any region
Deviations are clear (tangible, intangible, digital, physical)	Only for digital resources	More towards physical tangible resources. Other deviations are not clear

3.4.   Discussion

This chapter surveyed the literature related to this research. Each section investigated the main areas of information organization, modeling, and aggregation in the CHI domain.

The literature shows different approaches used in CHI modeling and organization. The scholars have used current models and ontologies plus brand-new models and approaches to achieve this task. For instance, the CIDOC-CRM and FRBR concept were utilized by many scholars for cultural heritage knowledge organization (Hu et al., 2014; Tan et al., 2009; Carboni & Luca, 2016) (Section 3.1).

Somehow it is obvious that CIDOC-CRM is used to organize both tangible and intangible cultural heritage while FRBR based FRBRoo model is being used to organize bibliographical CHI materials (Chen et al., 2013; Smiraglia, 2005) (Section 3.1). At some instances, the author utilized both CIDOC and FRBRoo terms to form a mapping between the proposed CHDE entities with the current models (Table 7 and Appendix 1). Through this mapping, the author tried to give a formalization to the proposed model classes, and it gives some understanding of the missing and unexplained classes in the CIDOC-CRM and FRBRoo entities as well (see Chapter 7 for more details).

Also, collection level knowledge organization in the CHI domain is another key factor in this research. The idea of *"collection"* and "*collection-level aggregation*" is a promising approach for the CHI domain as well. Specially aggregation of intangible cultural heritage should be done this way rather than item-level aggregation. Many CHI aggregation approaches concentrate on items and some scholars identify the potential of collection-level aggregation compared to the item-level approach (Low & Doerr 2010; Doerr et al. 1997; Geisler et al., 2002) (Section 3.1).

In Section 3.2, the author discussed a few studies based on the metadata aggregation platforms. Most of these platforms used the EDM as the base-model or they were designed to support the Europeana collection (Wickett et al., 2014; Wickett, Isaac, Doerr, & Fenlon, 2014; Papatheodorou et al., 2011; Freire et al., 2017).

The author found some studies based on independent CHI aggregation platforms and models that were developed by a few scholars, e.g., EEXCESS (Amin et al., 2012; Orgel et al., 2015; Orgel, Höffernig; Bailer, & Russegger, 2015; Hyvönen, 2016) (Section 3.1).

There were a few notable aggregation approaches such as DPLA and OAI-ORE which can be related to metadata aggregation. However, this study focused mainly on the EDM, CIDOC-CRM, and FRBRoo as the main related models. This chapter tried to make a

comparison between these models with the proposed models of this research and identified what is missing in the existing models (Table 2). Mainly, identification of intangible cultural heritage, distinguishing metadata based on their objectives were some problematic sectors in this discussion and the proposed CHDE model is trying to give some solutions to bridge the gap between these unsolved CHI situations.

The CHDE entities were mapped to existing schemas. This mapping confirmed that the CIDOC-CRM and FRBRoo do not provide an underlying model to explicitly express the relationships between digital surrogates and their source objects which may be tangible or intangible (Appendix 1). Besides, despite the vast ontology provided by CIDOC-CRM, it was difficult to find appropriate classes which clearly represent digital surrogates of a cultural heritage object. Similarly, these existing models have no definite classes to represent intangible cultural heritage entities (Wijesundara, Monika, & Sugimoto, 2017).

The author has introduced the METS standard (Section 3.3) as another related standard when it comes to identifying the CHI metadata via module like entities. Although the research does not use the METS schema as a whole, the concept of METS can be applied to this research as well, e.g., Description Modules model idea (see Chapter 5).

Chapter 4: Cultural Heritage in Digital Environment (CHDE): An Abstract Model for Information Organization and Aggregation

As a solution to the problems discussed above, the author designed the CHDE model as a conceptual model for CHI resource organization and aggregation which can be utilized to develop digital archives of cultural heritage (Wijesundara, Monika, & Sugimoto, 2017; Wijesundara & Sugimoto, 2018).

4.1. CHDE Main Components

4.1.1. Overview

Figure 6 represents the proposed CHDE model. The model defines entities and their relationships between CHI resources, and explicitly differentiates the physical and digital environments of the CHI. Memory institutions mainly collect resources that are realized in the physical environment and further digitize them in the digital environment.

This situation is applicable to both tangible and intangible cultural heritage. CHDE defines metadata for each of these instances based on the One-to-One Principle of Metadata.

Figure 6 presents the CHDE model for tangible and intangible cultural heritage. The entities which constitute the CHDE model are defined in Table 3. Besides, CHDE identifies two main resource environments, *Physical Space* and *Digital Spaces*.

i. *Physical Space*: All the resources and entities that exist and occur in the physical environment (without any connection to the networked environment) belong to this category, including all *Offline Resources, Tangible Cultural Heritage Objects*, and *Intangible Cultural Heritage* and their *Instantiations*. Tangible cultural heritage exists as physical objects which humans sense, and they can be recorded in physical mediums such as photographs and videos. An intangible cultural heritage entity has no physical existence as it is an abstract entity, and we can see intangible cultural heritage only through human activities in the physical space, such as dance performance and craftsmanship performance, which are called *Instantiations* of intangible cultural heritage. Therefore, the CHDE model identifies *Tangible Cultural Heritage Objects, Instantiations,* and their recordings as *Offline Resources* in the *Physical Space.* In addition, *Agent* for one who is involved in creating a tangible

cultural heritage object or performing an intangible cultural heritage event also exists in the *Physical Space.*

ii.   *Digital Space:* All the Digital Resources (converted from Offline Resources or born-digital) plus Curated Digital Instances (created from aggregated Digital Resources corresponding to a cultural heritage object) on the networked environment come under this category. Digital Space has no tangible-intangible differentiation.
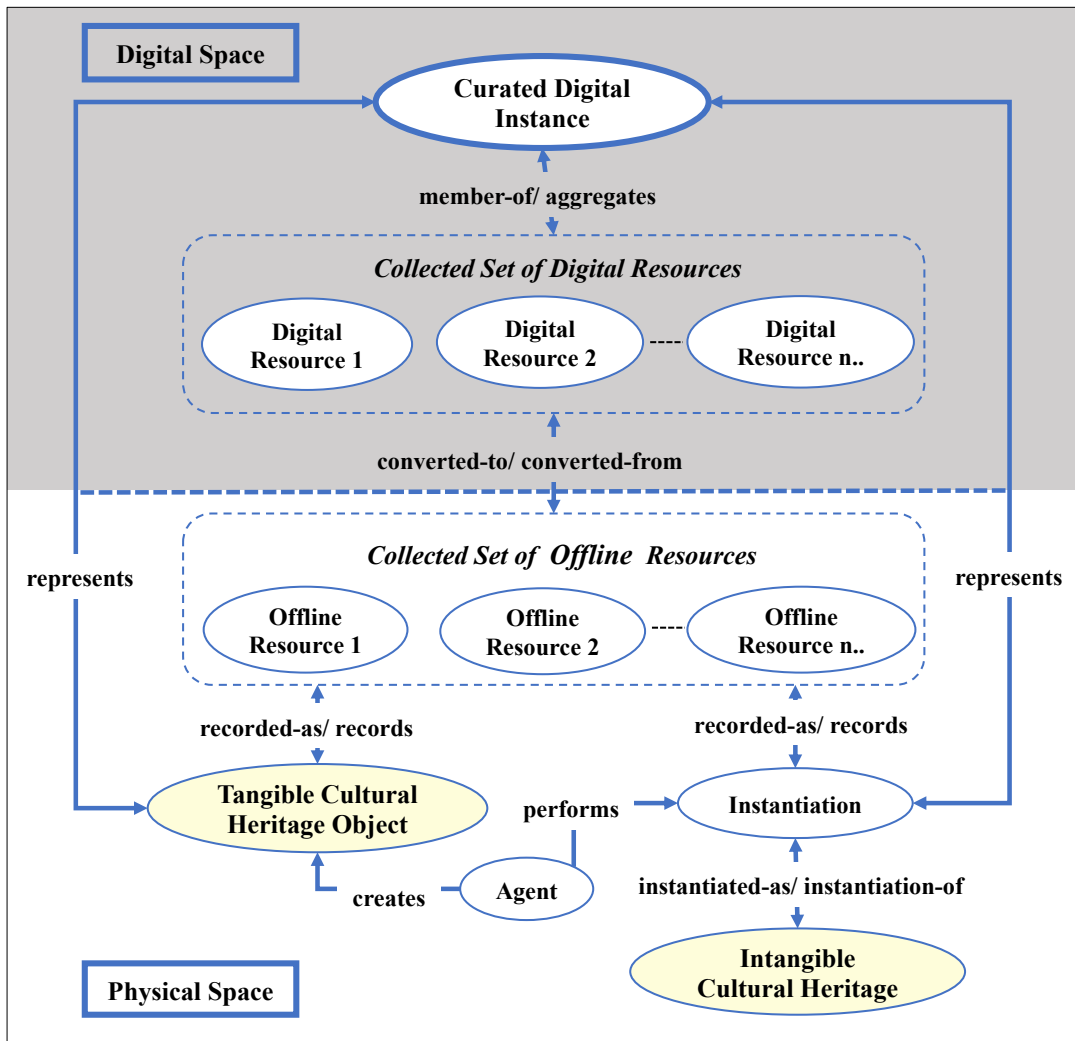
Figure 6. CHDE Model Describing Tangible and Intangible CHI Aggregation

The CHDE model represented in Figure 6 aggregates both tangible and intangible CHI. The entities in the upper section of Figure 6 (*Curated Digital Instances, Collected Set of Digital Resources* and *Collected Set of Offline Resources*) are common to tangible or intangible CHI aggregation. However, the bottom section of the diagram is slightly different as *Intangible Cultural Heritage* (which exists as conceptual in the *Physical Space*) cannot be captured directly as a *Tangible Cultural Heritage Object*. Therefore, the study uses a special feature called *Instantiation* as a physical entity to represent *Intangible Cultural Heritage* and

it can be aligned in parallel to the *Tangible Cultural Heritage Object* and can be later connected to the *Curated Digital Instance*.

Table 3. CHDE Entities and their Descriptions

	Entity Label	Description
1.	Offline Resource	A tangible cultural heritage object or Instantiation of intangible cultural heritage can be recorded by the memory institutions in various media such as image (printed photograph), sound (audio tape) or textual formats (printed book). In addition, there can be readily available digital resources such as an image file on a CD or a USB flash drive. These resources are named as *Offline Resources* in this model as they are still not connected to the networked environment. A single *Tangible Cultural Heritage Object* or an *Instantiation* can consist of one or several *Offline Resource/s,* e.g., a printed photo, a video on a CD, etc.
2.	Collected Set of Offline Resources	One or more of these *Offline Resource/s* (corresponding to a *Tangible Cultural Heritage Object* or *Instantiation* of an intangible cultural heritage) can be identified as a *Collected Set of Offline Resources*
3.	Digital Resource	*Offline Resources* can be converted into *Digital Resources* or utilized as is (if they are already in digital formats) in the *Digital Space*. For instance, a printed photo or an audio record on a tape can be converted to JPEG and MP3 format in the *Digital Space,* while a JPEG image on a CD can be used as it is without any conversion. In addition, there can be born-digital materials such as games and animations, which are readily available as *Digital Resource/s* in the *Digital Space*.
4.	Collected Set of Digital Resources	Subsequently, one or more of these *Digital Resource/s* can be identified as the *Collected Set of Digital Resources*.
5.	Curated Digital Instance	The topmost circle denoted *Curated Digital Instance* (CDI) acts as the aggregated set of digital resources corresponding to the *Tangible Cultural Heritage Object* or *Instantiation* of intangible cultural heritage entity at the bottom. This entity may include one or more *Digital Resource/s* and their metadata descriptions

		corresponding to a particular cultural heritage object.
6.	Tangible Cultural Heritage Object	This represents a tangible/ physical heritage object in the real world. These *Tangible Cultural Heritage Objects* may be housed in a museum or can be positioned as an immovable monument.
7.	Intangible Cultural Heritage	*Intangible Cultural Heritage* is basically a conceptual entity and can be realized as performance, ritual, skill, etc. This intangible cultural heritage has to be performed during a particular time and at a location, and once it has occurred, only the performance can be captured by any medium.
8.	Instantiation	Unlike tangible cultural heritage, intangible cultural heritage does not exist as physical items and so cannot be represented as items. However, an intangible cultural heritage entity can have many *Instantiations*. For example, a particular traditional dance performance can be performed in many places. Therefore, each performance is an *Instantiation* of the corresponding intangible cultural heritage entity. A specific intangible cultural heritage is represented by a combination of *Instantiations* and their associated resources. These intangible cultural heritage *Instantiations* can be based on temporal, location, category, agent, activity or concept associated with the intangible cultural heritage entity (Wijesundara, Monika, & Sugimoto, 2017). In addition, once an *Instantiation* has been created, it can be used to connect *Tangible Cultural Heritage Objects* which are interrelated with the intangible cultural heritage.
9.	Agent	The Agent entity (e.g., person or group of people) is associated with both tangible and *Instantiation* of the intangible cultural heritage, e.g., a painting done by a famous artist, traditional performance performed by a group of dancers, etc.

4.1.2. Instantiation Based Intangible Cultural Heritage

An intangible cultural heritage entity has no physical existence as it is an abstract entity, and we can see intangible cultural heritage only through human activities in the *Physical Space*; such as dance performance and craftsmanship which are called *Instantiations* of intangible

cultural heritage (Figure 6 and 7) (Table 3). Therefore, an *Instantiation* can exist as a physical entity which humans can directly see, hear, smell, taste and/or sometimes feel and they can be recorded in physical media such as videotapes, audio tapes, etc. Figure 7 shows the difference in capturing the tangible and intangible cultural heritage in a real-world situation. According to Figure 7, a tangible object (*Mask*) can be directly recorded via various media. Intangible cultural heritage is a conceptual entity. Therefore, it should have once occurred as an *Instantiation* which can be captured and recorded for future use.

Apart from that, intangible cultural heritage and their *Instantiations* are associated with tangible objects as well. For example, an outcome of traditional craftsmanship and cuisine will be a tangible entity such as an ornament and food. The CHDE model explicitly distinguishes the intangible heritage assets and their instantiations in physical forms in order to identify classes and properties in the metadata. According to Figure 6, *Instantiations* are placed in parallel to the *Tangible Cultural Heritage Object.* Therefore, in the CHDE model, an *Instantiation* acts as a physical instance (or "*object*" like entity) similar to a *Tangible Cultural Heritage Object* in the *Physical Space*.
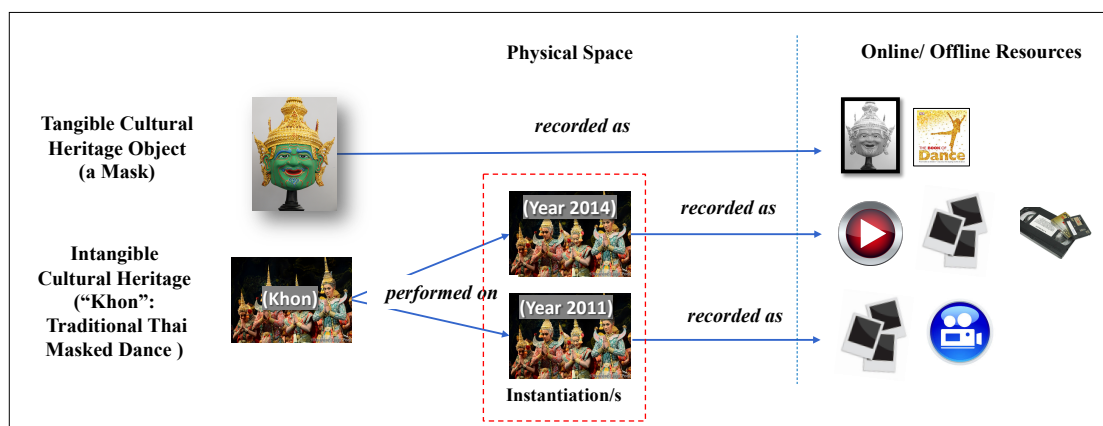
Figure 7. Capturing of Tangible and Intangible Cultural Heritage

Table 4 represents the *Instantiation* Classes related to intangible cultural heritage. Instance classes are realized according to 5W1H questions: when, where, what, who and how. Based on those questions, the author created *Temporal, Location, Category, Agent* and *Activity* Classes which represent attributes of *Instantiations*. Apart from these, an additional instance class was added to represent the conceptual entities, and named as *Concept* Class. The created classes were further mapped to the CIDOC-CRM, FRBRoo classes and terms from the Getty AAT to make a formalization.

Table 4. Intangible Cultural Heritage Instance Classes Mapped to CIDOC-CRM and FRBRoo

Instance Classes	AFS/ AAT Related Terms	CIDOC Classes	FRBRoo Classes
Temporal	time (AAT/AFS)	E2 Temporal Entity	
	time-related attributes (AAT)	E4 Period	F8 Event (=)
		E5 Event	
Location	location (AAT)	E53 place	F9 Place (=)
Category	_	E55 Type	F3 Manifestation Product Type (*) (≠)
Agent	agents (AAT)	E21 Person	F10 Person (=)
		E39 Actor	
Activity	activities (AAT)	E7 Activity	F31 Performance (*)
Concept	concepts (AFS)	E28 Conceptual Object	F6 Concept (=)

Key:	**AAT**: Art & Architecture Thesaurus (Getty)	
	AFS: American Folklore Society Ethnographic Thesaurus	
(*) Subclass	**(=)** Equal Class	**(≠)** Not Equal

4.1.3. Curated Digital Instance (CDI)

The Curated Digital Instance (CDI) in the CHDE model acts as an aggregated instance in the Digital Space and is created from cultural heritage objects in the Physical Space (Figure 6). As discussed above, a Digital Archive in the CHDE process is a collection of these Curated Digital Instances (CDIs) which is created according to the One-to-One Principle of Metadata concept (Figure 8).

This research defines a CDI as "a collection of digital resources and their descriptions representing a single cultural heritage object." A CDI is not a single CHI instance but a collection of CHI instances.

As a whole, Figure 8 depicts a CDI instance representing an intangible cultural heritage entity labeled Performance "A". The CDI aggregates various digital objects/resources (e.g., a video, a photograph, and an audio record), each of which should be given a metadata record. Similarly, the instantiation of intangible cultural heritage, which is denoted as Performance "A", has its own metadata (i.e., description about the performance). External Resources are vital when identifying and enriching the CDI via non-institutional resources. Moreover, Original Descriptions are given to the intangible cultural heritage in the physical environment

can be aggregated into the CDI. Each of these resource components has its own set of metadata, and a CDI itself has a separate metadata record. Identification of individual metadata descriptions related to individual resources is based on the One-to-One Principle of Metadata . This differentiates individual objects separately without any confusion with other resources/objects.

As CDI is the final outcome of the CHDE model this is the main aggregated instance that corresponds to a certain tangible cultural heritage or *Instantiation*. By linking, CDIs can form a large digital platform and can be used as the foundation for creating digital archives of cultural heritage.
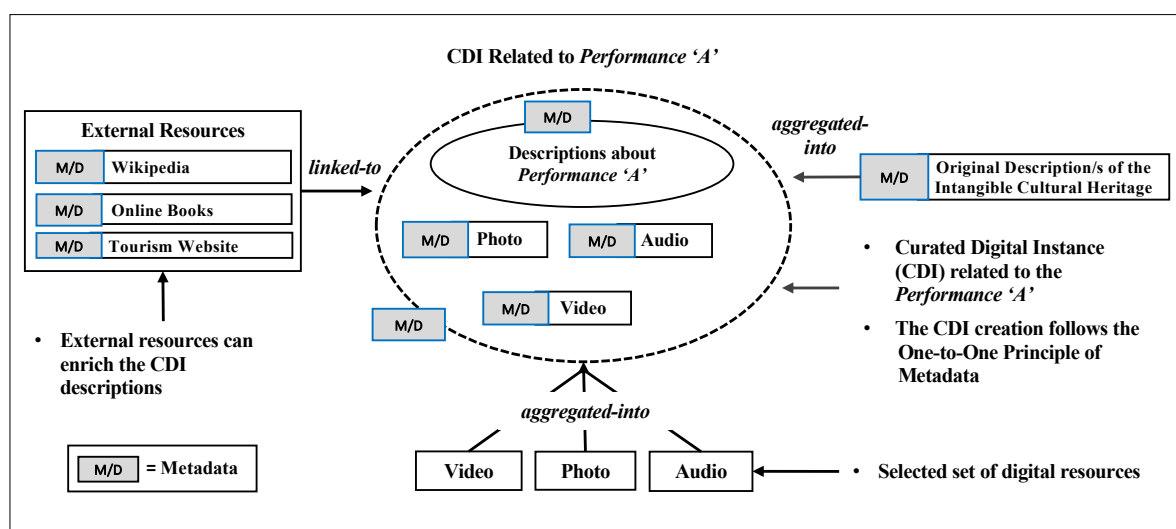
Figure 8. Curated Digital Instance (CDI)

4.2.   CHDE Curation Process

Digital Archives curate relevant digital resources into one platform and deliver them to the users in a suitable way. Europeana, DigitalNZ[30] and the British Museum[31]host such comprehensive digital archives and there are various other small-scale digital cultural heritage archives on the Web, e.g., Princess Maha Chakri Sirindhorn Anthropology Centre, Thailand[32]. These digital archives have their own digital curation process to organize CHI into a digital archive. This section discusses a generalized digital curation process model defined in accordance with the entities included in the CHDE model. The CHDE Curation

---

[30]https://digitalnz.org/

[31]https://www.britishmuseum.org/research/collection_online/search.aspx

[32]https://www.sac.or.th/main/en/database/

Process diagram shown in Figure 9 helps recognize entities in the CHI creation, selection and organization activities in a digital archival system.

Digital curation has emerged as a new inter-disciplinary practice that seeks to set guidelines for the disciplined management of information. For example, the Digital Curation Center (DCC) is an internationally recognized institution which is involved in R&D activities related to the digital curation domain. According to the DCC, "digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle" [33]. They have introduced a model called the DCC curation lifecycle model which facilitates a graphical high-level overview of phases necessary for successful curation and preservation of a digital object from its start to the end (Higgins, 2008). The same model was further extended by the Digital Curation Unit at the Athena Research Centre in Greece (Constantopoulos et al., 2009). The model shown in Figure 9 is based on these general digital curation process models.

This study focuses on a few key phases in the process shown in Figure 9. The oval labeled as Cultural Heritage Object (CHO) denotes the intangible and tangible heritage entities in the physical world.

CHOs can be recorded into a carrier which may or may not be digital, where tangible cultural heritage can be recorded directly, and intangible cultural heritage can be recorded as their Instantiations.

Figure 9 consists of three main *Activities* (gray rectangles) and *Resources* (ovals). In the *Ingestion Activity*, the *Non-Digital Resources* of various physical carriers are digitized, and digital resources such as *Born-Digital* are imported from various places (e.g., digital archives). After the *Ingestion Activity*, the resources are selected according to the institutional and user requirements. The digital resources are selected depending on factors such as the usefulness of the content, data capacity, reliability, cost, etc. Then, the digital resources are given contextual information as a part of their metadata in the *Description Activity*.

Curated Digital Instance (CDI) is an essential output of the *Description Activity*. A single *Curated Digital Instance* has to contain a non-empty set of digital resources. The *Curated Digital Instance* acts as a container for those context-rich digital resources related to a cultural heritage object in the physical world. This idea will be further described later in Chapter 6.

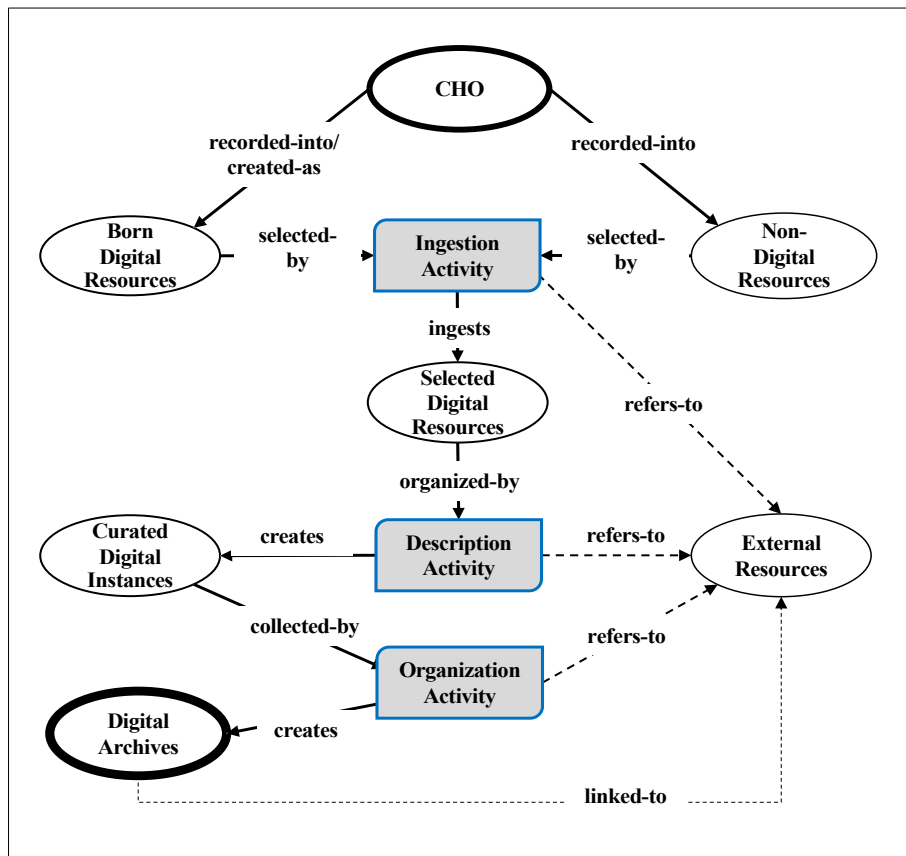---

[33]http://www.dcc.ac.uk/digital-curation/what-digital-curation

Figure 9. CHDE Curation Process

The third activity is known as the *Organization Activity*. The resources in the Curated Digital Instance is arranged and classified during this phase and for this various standards and techniques can be utilized. During all three Activity phases (*Ingestion, Description,* and *Organization*) *External Resources* are applicable. These *External Resources* can be online or offline and, in each phase, they can be used to identify and enrich the context of a digital resource corresponding to a heritage entity. The *Description Activity* forms the main information aggregation which is the key objective of this whole process. The output of this aggregation is *the Curated Digital Instance* (CDI) which is the main entity corresponding to the CHO on the top.

Conclusively, the *Digital Archive* is created after aggregating collection of those CDIs and it will be the final output of the whole digital curation process. After the entire process, the digital resources (or CHIs) are created, contextualized and categorized and now they are ready for dissemination via a *Digital Archive*.

4.3. Discussion

This chapter described the proposed CHDE model and its entities comprehensively. The CHDE model is developed to define the entities (Figure 6 and Table 3) which are involved in creating a digital archive of CHI. The model is designed to aggregate both tangible and intangible cultural heritage. The CHDE model is an abstract model which consists of some broader entities which permit aggregation of any kind of CHI. The identification of the digital and physical information spaces of CHI is a prominent feature in the model (Figure 6). It shows a clear difference between two information spaces which is not visible in existing CHI aggregation models. *Offline Resources* explicitly identify the resources in the CHDE model for archiving of tangible but large objects or ephemeral objects that cannot be housed in a museum.

The identification of intangible cultural heritage through *Instantiation* permits the memory institutions to organize resources based on a physical entity (*Instantiation*) related to an intangible cultural heritage entity (Figure 7 and Table 4). The *Instantiation* approach facilitates identification, organization and aggregation of CHI into the Curated Digital Instance, and it adds more contextual information to the intangible cultural heritage entity. Identifying and creating *Instantiations* are solely dependent on the institutional requirements and according to the CHDE model, a single intangible cultural heritage can have one or more *Instantiations,* while a single *Instantiation* should have at least one or more CHI record/s.

Section 4.1.2. presented the Curated Digital Instance (CDI) which is the main outcome of the CHDE model. Since it is the aggregated component of all tangible cultural heritage objects and *Instantiations* of intangible cultural heritage it plays the main role in the CHDE model. A CDI consists of one or more digital resource/s in many formats and their metadata corresponds to a cultural heritage object in the *Physical Space*. Instances like CDIs are common among existing aggregation platforms in the CHI domain. Still, the CDI in the CHDE model is unique as it aggregates diverse resources, e.g., selected digital resources of a cultural heritage object, related external resources, original descriptions of heritage object, metadata of individual resources, and so on. Therefore, CDI is not another simple resource aggregation, but rather for complex and higher-level aggregation compared to the existing methods. However, resources such as descriptions about cultural heritage objects, e.g., books and articles, are not explicitly included in the CHDE as it is focusing on primary objects only.

The CHDE process discussed in Section 4.2 is intended to explain the overall process behind the CHDE model. The process discussed in this thesis is not intended to describe the

entire digital curation process in current practice, but it exhibits a few essential *Activities* and *Resources* related to the CHDE model (Figure 9). Basically, we can identify three main activities such as *Ingestion, Description,* and *Organization* which are key tasks in the CHDE model. CHDE curation process presents two aggregation levels i) Curated Digital Instances, created as an aggregation outcome of the CHDE model, and (ii) Digital Archives (of cultural heritage), created by aggregating a collection of Curated Digital Instances. However, although this second aggregation level is out of the scope of the current study, it supports the general idea of developing a digital archive of cultural heritage as an extensive scenario of this study.

Chapter 5: Identifying CHI Metadata Through Description Modules

Previously the author proposed the CHDE model which is generic and abstract for organizing and aggregating cultural heritage resources curated into digital archives. This chapter is dedicated to overview of the second metadata model connected to the CHDE model proposed by this study. The model is known as Description Modules model and it gives some conceptual idea for metadata mapping aiming aggregation of metadata descriptions from multiple schemas. Therefore, the key ideas discussed in this model are;

i.   How to aggregate both institutional and non-institutional metadata and

ii.   How to utilize semantic and structural features of metadata for mapping across different application metadata.

5.1.   Introduction

Even though CHDE proposes a theoretical method to the development process of digital archives via aggregating diverse CHI resources, it does not explicitly mention how it can be carried out. Also, CHDE does not mention much about the metadata aspects of the CHI which is the key to the aggregation. This chapter is dedicated to discussing the data model that lies behind the CHDE model which supports the CHI aggregation.

Figure 10 supports this idea by showing the connection between the CHDE model and the Description Modules model. Basically, the CHDE's Curated Digital Instance (CDI) acts as the main aggregated instance corresponding to a cultural heritage object which may be tangible or intangible. According to Figure 10, the upper portion of the figure shows this abstract level CHI aggregation process by aggregating various information resources and creation of the CDI related to a cultural heritage object known as the "Statue of Tara"[34] from the British Museum. According to Figure 10, each resource (e.g., photo, audio, video and original description) has its own metadata, and in this chapter, the author specifically focused on distinguishing and aggregating those metadata accurately. In the CHDE model discussion, the study did not focus on these metadata of the CHI. The second half of the figure depicts a single metadata record instance corresponding to a resource (an audio record extracted from BBC, UK). This metadata record consists of a mixture of information related to the audio file (digital surrogate) and the cultural heritage object ("Statue of Tara" which is the original

---

[34]https://www.bbc.co.uk/programmes/b00snm1x

object). This chapter is basically concentrating on how these mixed metadata records can be differentiated correctly without violating the One-to-One Principle of Metadata concept and to create a CHI aggregation across multiple schemas.
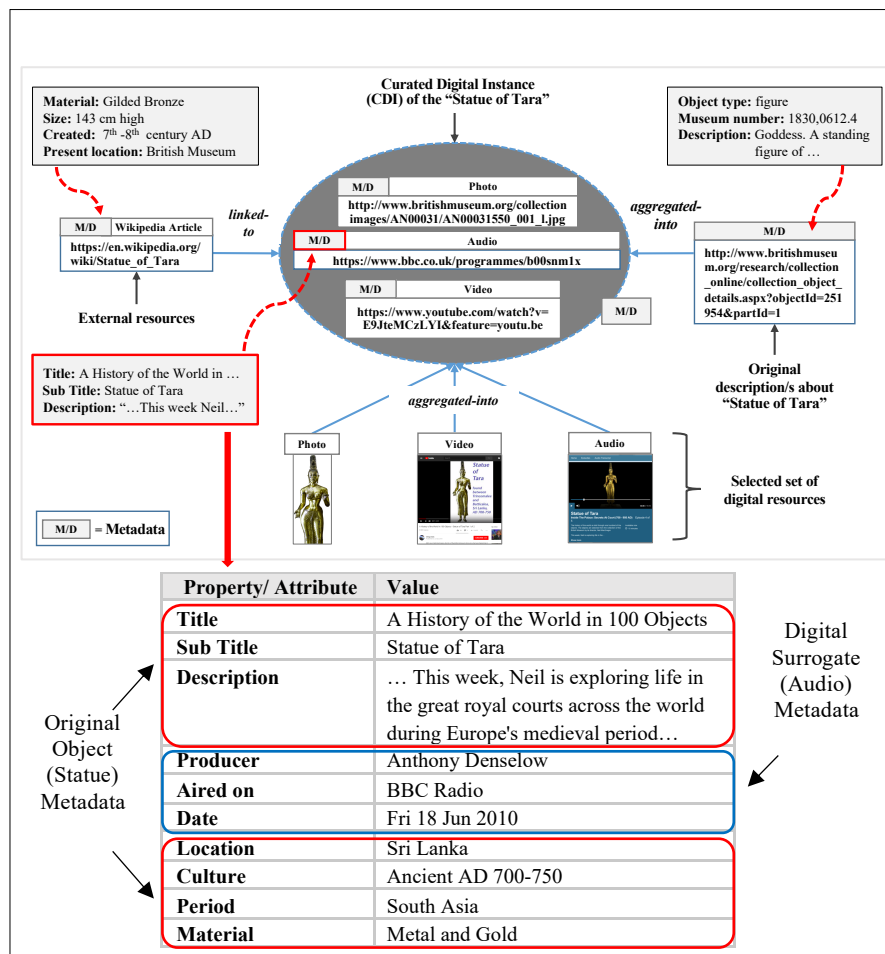

Figure 10. Connecting Description Modules Model with the CHDE Model

5.2.   Description Set Profiles (DSP) Involvement

The One-to-One Principle of Metadata and DCAP (Dublin Core Application Profiles) are used as the basis of the Description Modules model presented in this chapter. As introduced in Section 2.4, DCAP's DSP (Description Set Profiles) can be closely matched to a conventional CHI record on the Internet (Figure 11). For instance, a particular CHI record[35] consists of multiple descriptions (related to various objects) which can be considered as a *Set of Description Sets* and a single block of description/s related to a particular object that may

---

[35]https://www.europeana.eu/portal/en/record/91619/SMVK_EM_objekt_1024261.html

be a *Description Set* while a set of property-value pairs may imply a single *Description*. This idea is presented in Figure 11 as follows.

The concept of *Description Sets* and *Description* is crucial when identifying individual objects and their corresponding metadata described within a single record which is based on the One-to-One Principle of Metadata . The model presented in this chapter investigated this basic concept and the author followed a similar but unique method to identify metadata components of a CHI record in a different way.
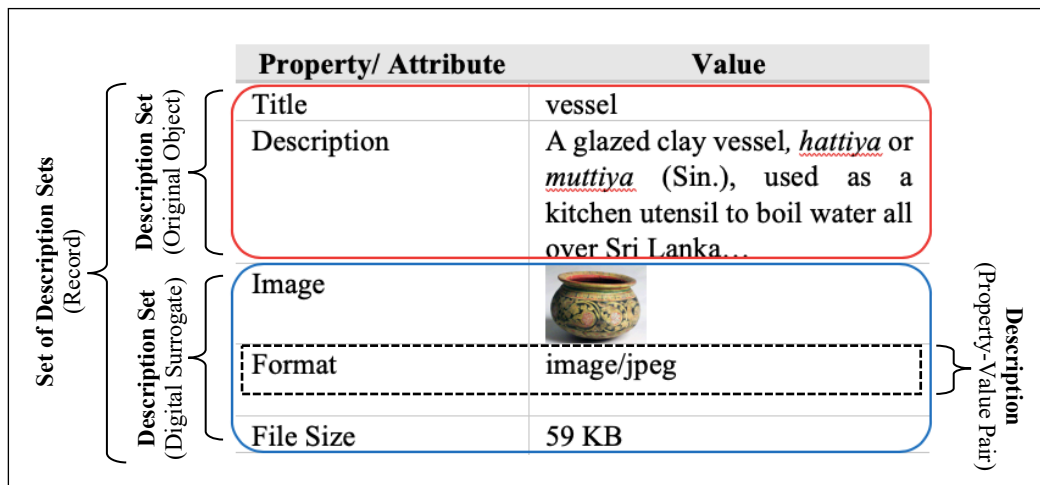

Figure 11. DSP Structure of a Cultural Heritage Record

5.3. Description Modules Model as a Metadata Structure to Identify Metadata

This research is dealing with institutional and non-institutional CHI belongs to diverse metadata schemas. Therefore, the primary target of this study is to aggregate heterogeneous CHI and we need to think of a specific strategy to realize this aggregation effectively. Semantic mapping across different metadata schemas is the basic technique used to aggregate metadata collected from different sources. Conventionally, properties of two or more application metadata schemas are mapped. Though it is a promising approach, we have learned that there are a few issues and challenges associated with property-to-property level metadata mapping (Section 2.2.2).

Secondly, the mixed nature of institutional CHI records makes aggregation difficult and it is a challenge for the digital archives and users of the digital archives when the objectives of the CHI are unclear (Section 2.2.2).

Therefore, the study created the Description Modules model to organize diverse metadata related to CHI for decreasing the complexity of metadata mapping and as an answer for the hybrid nature of the CHI records. Here modularization of metadata is used, and the idea of modular metadata was introduced in Section 2.1.2 of this thesis.

Prior to developing the Description Modules model, the author collected metadata instances from ten different CHI services (Table 5). The Web services were selected based on the heritage categories and institutional and non-institutional basis. As there is a need to collect and aggregate diverse information, the author used non-institutional Web services such as Wikipedia in this analysis. In addition, different types of cultural heritage resources, tangible and intangible, were collected to represent all cultural heritage categories.

After careful extraction of metadata from the Web resources identified above, the author determined four main categories for *Description Sets* (referred to as *Object Categories* in this thesis) (Table 6) which are represented by the metadata in these records. This research uses the term *Object Categories* referring to those four main category types which can be listed as follows.

Table 5. List of Investigated Digital Archives and Web Services

	Web Resource	URL	Resource Type
1	British Museum, UK	http://www.britishmuseum.org/research/collection_online/search.aspx	TCH (artifacts)
2	Metropolitan Museum of Art, USA	https://www.metmuseum.org/art/landing-page	TCH (artifacts)
3	Asian Art Museum, USA	http://www.asianart.org/collections/collection	TCH (artifacts)
4	Rijksmuseum, Netherlands	https://www.rijksmuseum.nl/en/search	TCH (artifacts)
5	Europeana Collection	https://www.europeana.eu/portal/en	TCH (artifacts) & ICH
6	Technical University of Cluj-Napoca, Romania	https://dacit.utcluj.ro/scandb/?page=scandb#/models/en	TCH (artifacts as 3D objects)
7	Wikipedia	https://www.wikipedia.org/	TCH & ICH
8	UNESCO World Heritage Centre	https://whc.unesco.org/en/list/	TCH (Monuments)
9	Asia/Pacific Database on Intangible Cultural Heritage (ICH)	http://www.accu.or.jp/ich/en/index.html	ICH
10	UNESCO- ICH	https://ich.unesco.org/en/lists	ICH
	TCH: Tangible Cultural Heritage, ICH: Intangible Cultural Heritage		

i. *Original Object.* Consist of physical (original) artifacts/ monuments which belongs to tangible cultural heritage. In the intangible cultural heritage domain, all the instantiations and their recording objects (e.g., photographs, an audio recording of an event) can also be considered as *Original Objects* based on the definition in CHDE.

ii. *Digital Surrogate.* All the digital copies corresponding to an original intangible or tangible heritage can be put into this category. Captured images, 3D scanned objects, digital video of a dance performance are a few such digital surrogates. *Digital surrogates* are semantically associated with *Original Objects*.

iii. *Administrative.* All *Original Objects and Digital Surrogates* have their own administrative information. They are not directly expressing the original or digital object, but they are important when handling the corresponding objects individually.

iv. *External Resource.* This is another significant resource related to a CHI, from which you may obtain additional information about the CHI, e.g., blogs, websites and online books and journal articles.

Table 6. Identified Object Categories and Description Modules

Object Categories	Description Modules
Original Object	Content Description Module
	Agent Module
	Location Module
	Timeline Module
	Technical Description Module
Digital Surrogate	Content Description Module
	Agent Module
	Location Module
	Timeline Module
	Technical Description Module
Administrative	Content Description Module
	Agent Module
	Location Module
	Timeline Module
	Provenance Module
	Rights Module
External Resource	External Link Module
	Bibliographic Module

These four *Object Categories* were created based on the requirements of the model. This study wanted to distinguish the metadata related to the original object in the physical world and the digital surrogate metadata of the same as two separate metadata descriptions. Based on that, the *Original Object* and *Digital Surrogate* categories were created. In contrast, we can see administrative metadata related to the objects and sometimes external information associated with the objects included as URLs. Accordingly, another two *Object Categories* were developed known as *Administrative* and *External Resources.* Each *Object Category* consists of several *Description Modules* which represent the content of the target *Object Category*. Some *Description Modules* may be repeated among several *Object Categories*, e.g., Original Object and Digital Surrogate both have Agent Module. However, the meaning and the content of *Original Object – Agent Module* and *Digital Surrogate- Agent Module* are distinct. Class definitions related to these *Object Categories* and *Description Modules* will be presented in Chapter 6 (Table 7).

Figure 12 depicts an identification of metadata descriptions based on the One-to-One concept.
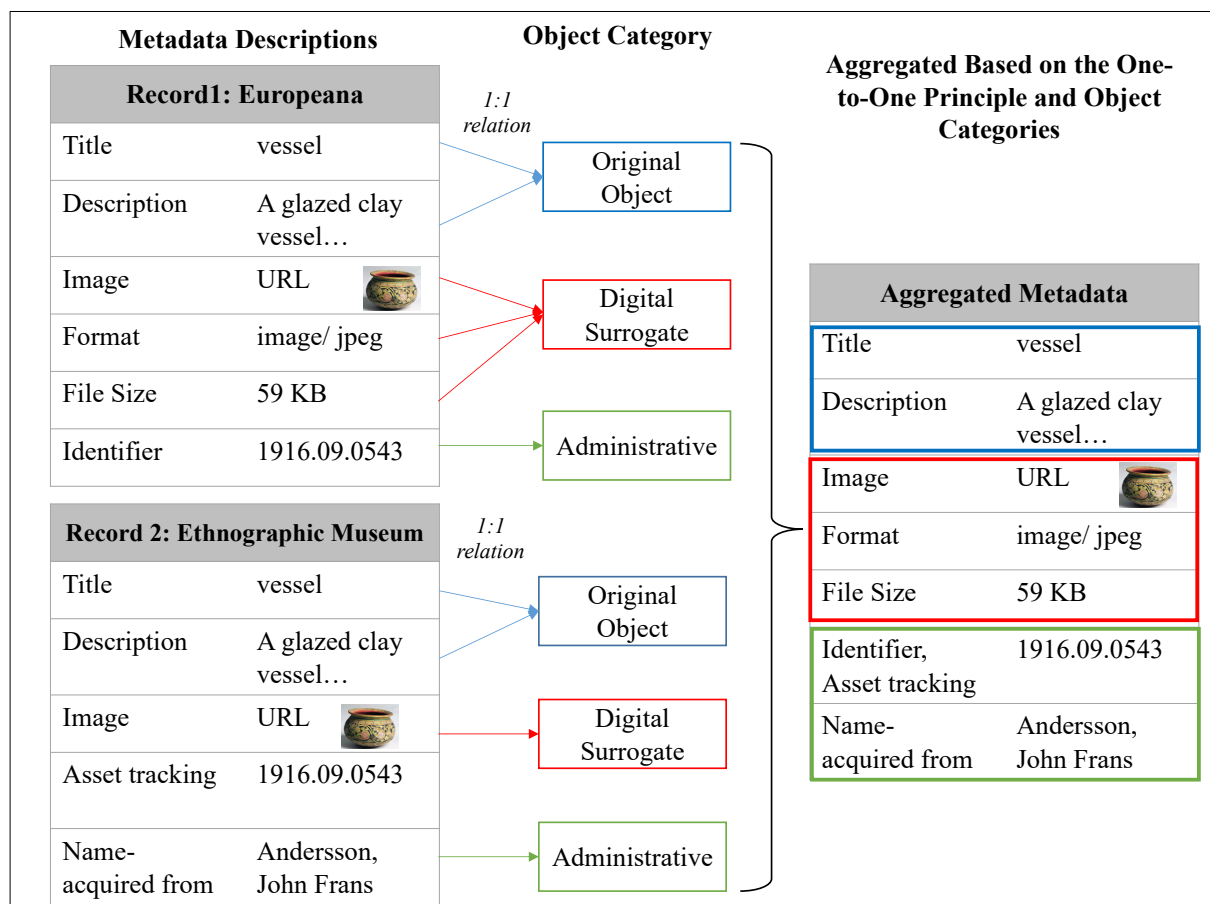

Figure 12. Mapping Object Categories to Metadata Descriptions to Perform Aggregation Based on the One-to-One Principle

Since the study has determined four *Object Categories* each description can be mapped into those objects creating a one-to-one relation between the metadata description and the *Object Category*. Here, two record instances extracted from Europeana[36] and the Ethnography Museum of Sweden[37] and each of these metadata descriptions, can be mapped to its corresponding *Object Category*. The outcome of this aggregation may be an *Aggregated Metadata Description* as above (Figure 12). Blue, Red and Green colors are used to portray the *Original Object, Digital Surrogate* and *Administrative* object categories of the metadata descriptions.

Theoretically, we can identify objects and corresponding metadata descriptions as above (Figure 12). Yet, the study needed finer structure to distinguish metadata in a more detailed manner as it makes the aggregation more concrete. Hence, a set of sub-modules were created under four main *Object Categories* which are known as *Description Modules* (Table 6). (A full list of the instances used to create *Description Modules* are available in Appendix 2). Each object category can have multiple sub-modules and a single module acts as a structural component which consists of one or more metadata descriptions. This research identifies a *Description Module* as "a data entity/instance which consists of one or more property-value pair/s". Therefore, a single *Description Module* acts as a *Description Set* and it works as a container to capture metadata to identify the objectives of the metadata. Figure 13 depicts how these *Description Modules* and *Object Categories* are used to distinguish metadata instances in a cultural heritage record[38] in a more detailed manner.

Figure 13. Assigning Description Modules to Metadata Instances

[36] https://www.europeana.eu/portal/en/record/91619/SMVK_EM_objekt_1024261.html

[37] http://collections.smvk.se/carlotta-em/web/object/1024261

[38] https://www.europeana.eu/portal/en/record/91619/SMVK_EM_objekt_1024261.html

Likewise, any metadata instances represented by a record can be mapped into this *Description Modules* scheme.

Basically, a set of *Description Modules* (*Description Set*) depicts a single object. Therefore, *Description Modules* are a sort of identification of an object which may be digital or physical. Hence, the ultimate goal of this Description Modules model is to aggregate diverse cultural heritage descriptions based on a module-like structure. As this research deals with multiple schemas, the author applied the Description Modules model to map between schemas to create the aggregation (Figure 14 and 15).
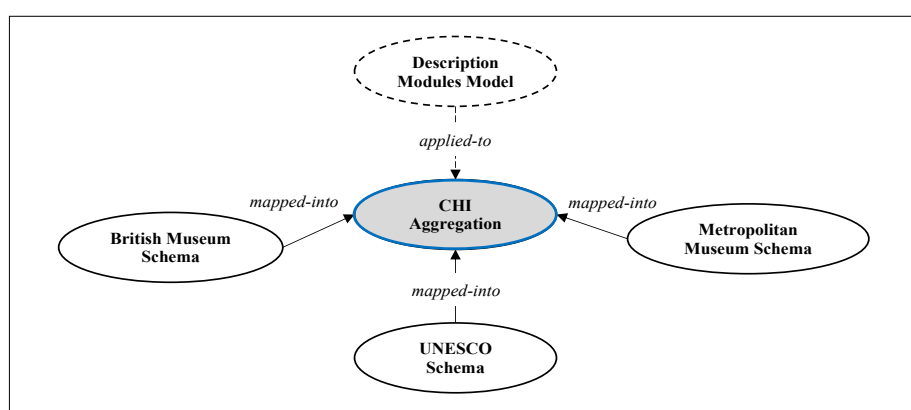
Figure 14. Applying Description Modules Model to Map Diverse Schemas

The *Description Module* acts as a schema to capture the metadata of a heritage record, e.g., *Content Description Module* may include metadata related to Title, Description and Identification of an object (Figure 15). Each of these descriptions may have separate structures and substructures.

However, identifying the data structure is not the focus of this study. Figure 15 represents two records from the British Museum[39] and Wikipedia[40] which describe a tangible object. The metadata which describes contextual information, such as *Object Type, Description* and *Identification* of two records can be mapped into the *Content Description Module* of the Description Modules model. The mapping example in Figure 15 is done manually. Theoretically we can assign *Description Modules* to metadata descriptions without any trouble. But, this kind of mapping is a bit tricky and it does not consider the finer details of a given metadata unit.

---

[39]https://www.britishmuseum.org/research/collection_online/collection_object_details.aspx?objectId=251954&partId=1

[40]https://en.wikipedia.org/wiki/Statue_of_Tara

Also, mapping finer descriptions into a single module might create complications during aggregation as it does not fully represent the context of the metadata. Therefore, the study proposes an extension to the Description Modules model to carry-out the aggregation more effectively.
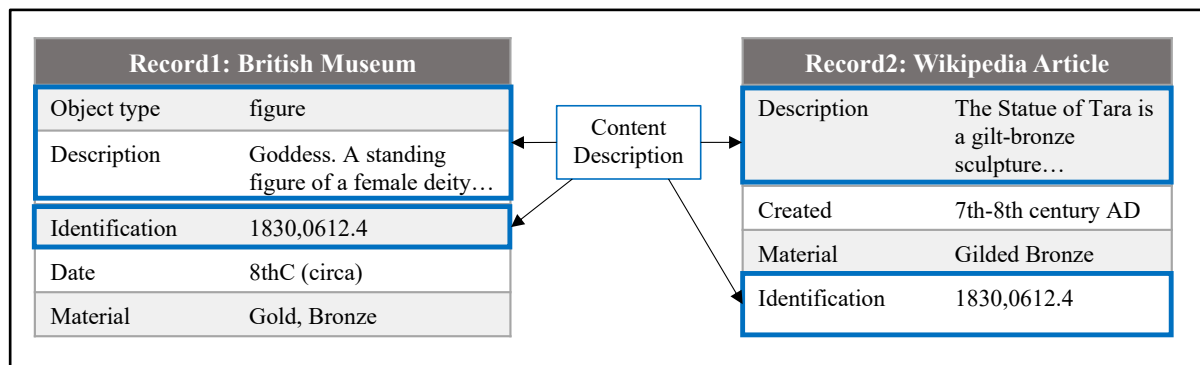

Figure 15. Mapping via Description Modules

5.4.   Facets: A Window to an Object

As introduced above, metadata aggregation needs finer attributes to perform the aggregation effectively. However, *Description Modules* are more generalized in this context. Since we need to avoid property-to-property mapping as well, the author proposed an alternative way to perform the mapping based on the classes. Basically, the mapping needs to define classes and classes of *Description Modules* cannot be used to connect real objects (such as artifacts or digital surrogates) as the meanings of those classes are very different. For instance, *Content Description* module may have classes called *Date Class* or *Place Class*. However, these *Date* or *Place* classes are intended to describe some attributes related to the *Content Description module,* but not the real *Object* which the *Content Description Module* is connected to via its metadata. It is true that the *Description Modules* can be used to identify the metadata structure of a CHI record, but it has no connection with the *Object* itself, as *Object* exists outside the metadata or schema levels.

Therefore, this research further identifies some finer entities related to the *Description Modules* called *Facets* (Figure 16). While the *Description Module* explicitly describes a grouping of metadata descriptions, a *Facet* explicitly identifies a particular description based on the One-to-One Principle.

Facet in this research is a particular grouping to identify the attributes of an *Object* and it enables creating a link between the *Object* and *Facet* via classes. The connection between the *Facets, Objects,* and *Description Modules* are shown in Figure 16.

This research created the *Facet* after examining the metadata instances extracted from the Web services mentioned in Table 5 (The full list of *Facets* is available in Appendix 3).
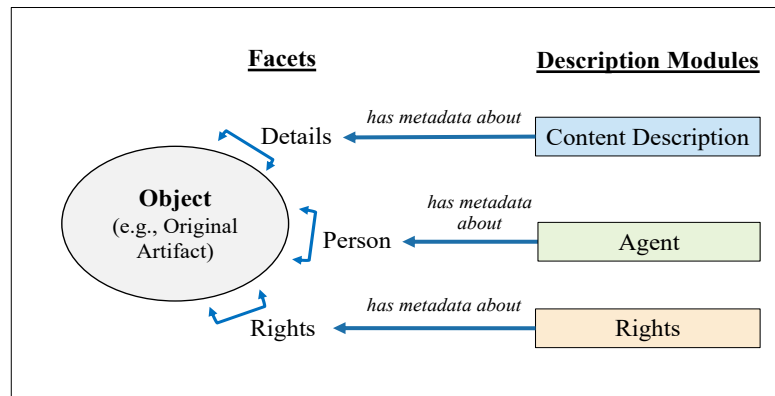
Figure 16. Viewing an Object via Facets

5.5.   Discussion

This chapter is dedicated to the second metadata model proposed by this study called the Description Modules model. This model is basically connected with the CHDE model and the main aim is to facilitate metadata aggregation via mapping heterogenous CHI across multiple schemas.

After considering a few concerns associated with the CHI records and conventional aggregation issues (Section 2.2.2. and Section 5.3), the author proposed the Description Modules model to organize diverse information contents to decrease the complexity of metadata mapping and the hybrid record issue. Here modularization of metadata (Section 2.1.2) and *Facet*s (Section 2.1.2) are introduced as basic concepts related to this model and the idea was backup by the DCAP's DSP model as well (Section 2.1.2 and Section 5.2).

The Description Modules model (Section 5.3) consists of two main streams and the first one is called *Description Modules*. *Description Modules* acts as a grouping for metadata and it can be used as a structure to capture metadata according to different objectives. The objectives of the metadata are mainly categorized according to the *Object Categories* and *Description Modules* proposed in the model (Figure 12, 13 and 15).

Tangible and intangible CHI from different Digital archives and Web services related to institutions and non-institutions are examined to create these modules. Therefore, it covers many heritage types from diverse metadata schemes and can be used to help aggregate the same (See Figure 14, Table 6 and Appendix 2).

Secondly, the model introduced the *Facet* concept which can be utilized to identify and map *Objects* in the real world (Figure 16). Since *Description Modules* are typically representing the data structure of an object it cannot be directly linked to the *Object*. Basically, the *Description Modules* are highly dependent on the metadata of an *Object*, as an implementation-based structure. Subsequently, the *Facet* gives an implementation neutral perspective as it is interacting with the view of the *Object*. Therefore, *Facets* presented through the Description Modules model can be used as a reference model to identify the view of an *Object* without considering its metadata structure (Section 5.4).

Some aggregation examples based on the *Description Modules* and *Facets* will be discussed in the next chapter of this thesis.

Chapter 6: Feasibility of the CHDE Model and Description Modules Model

6.1.   CHDE Model Use Cases

6.1.1. Tangible and Intangible Cultural Heritage Aggregation

This section shows a few use cases of the CHDE model applied to real-world examples. Examples involving tangible cultural heritage and intangible cultural heritage are illustrated in Figure 17 and 18, respectively. Figure 17 and 18 are based on the CHDE model (Figure 6) but they are illustrated as two separate examples for better understanding.
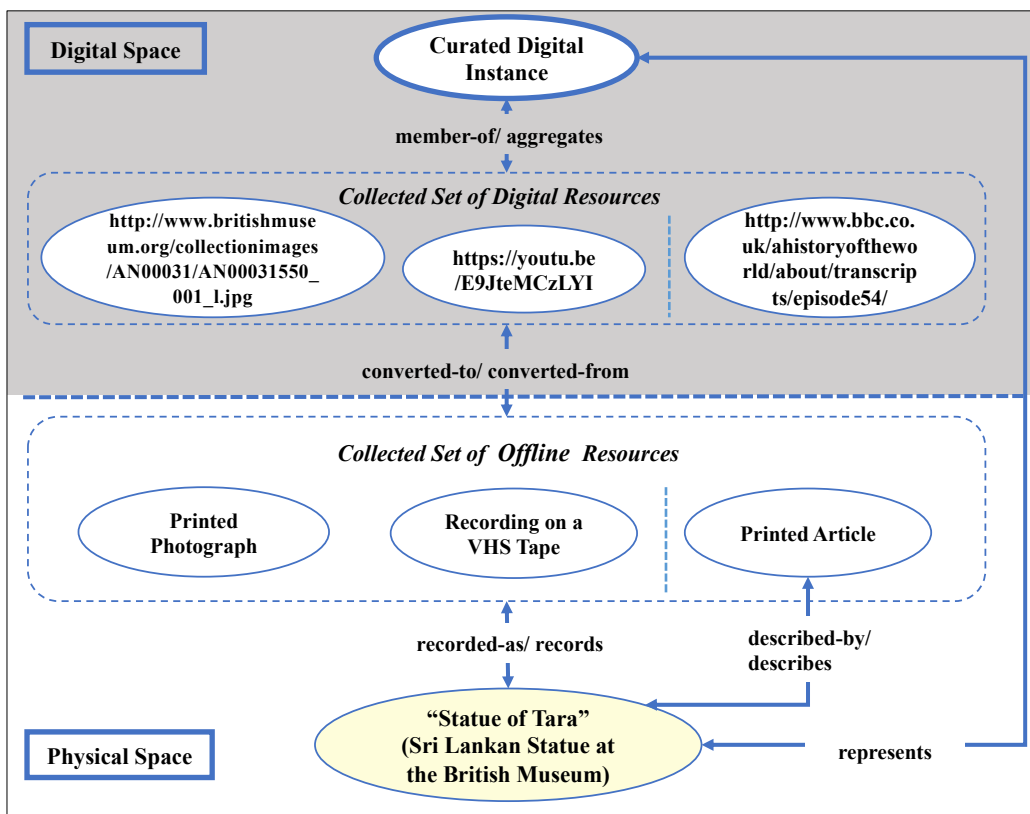
Figure 17. CHDE Model: Tangible Cultural Heritage Aggregation

Figure 17 shows a tangible cultural heritage object known as the *"Statue of Tara"*, which is a bronze artifact of 8th Century AD originated in Sri Lanka, and currently housed at the British Museum, UK[41]. This statue can be recorded directly into physical media as a

---

[41]https://www.britishmuseum.org/research/collection_online/collection_object_details.aspx?objectId=251954&partId=1

printed photograph, video recording or printed article which consists of some background information of the object. Further, these *Offline Resources* can be converted into *Digital Resources* in *Digital Space*. For instance, a VHS recording can be converted into a YouTube video which can be in MOV, MP4 or WMV format. In addition, there can be resources which are already in digital formats which exist as *Offline Resources* and can be directly collected in the *Digital Space*. Finally, all these collected resources are aggregated as a *Curated Digital Instance* (CDI) which represents the *Statue of Tara* as a single comprehensive unit. Hence, resources related to the tangible cultural heritage artifact can be identified as separated CHIs. As a result, CDI supports the possibility of identifying metadata descriptions individually. Since it collects and aggregates institutional and item-centric metadata from the British Museum and non-institutional metadata such as YouTube videos and BBC articles, it makes the CDI more context-rich as well. The *Agent* entity in the original CHDE model (Figure 6) has been removed from this example as the creator of the artifact (*Statue of Tara)* is anonymous.

Figure 18 shows an intangible cultural heritage entity known as the *"Kandy Esala Perahera"* festival in Sri Lanka.
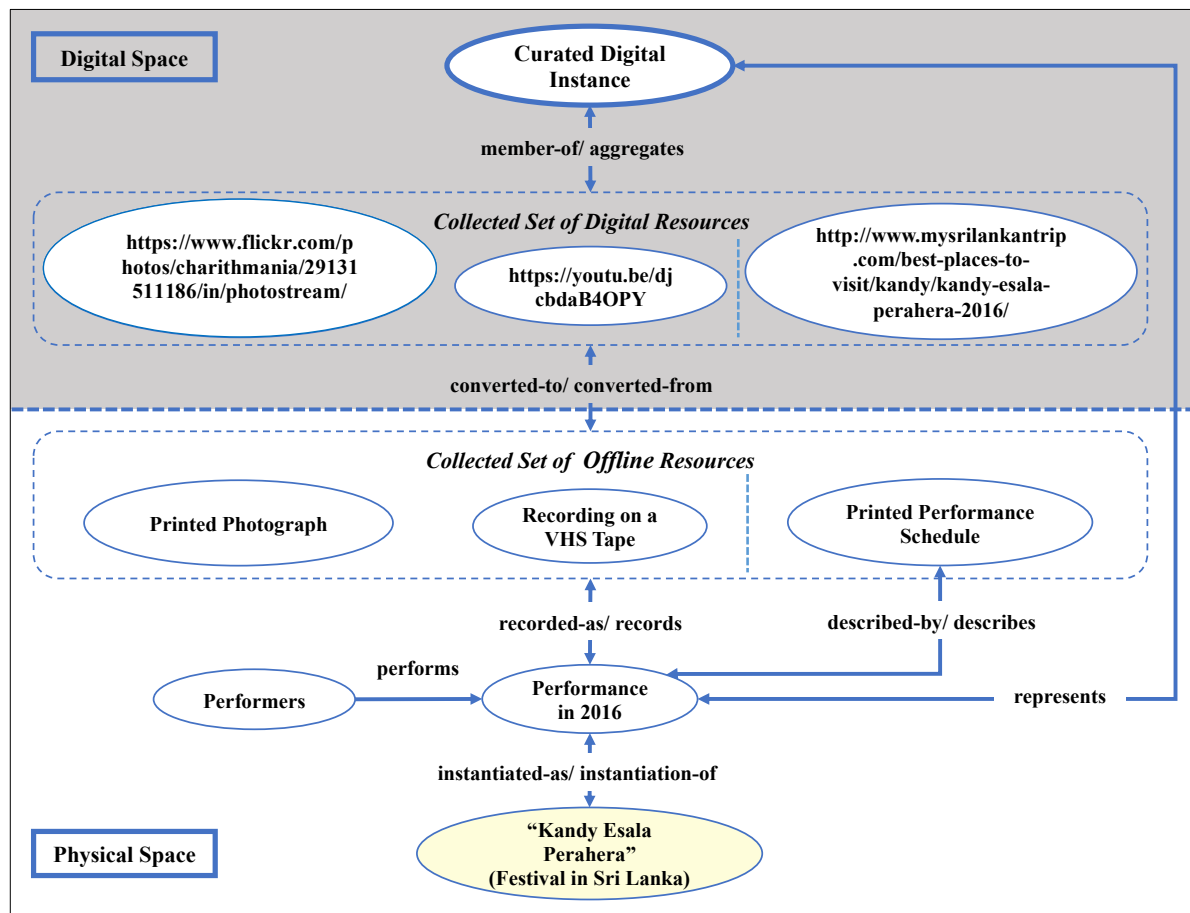

Figure 18. CHDE Model: Intangible Cultural Heritage Aggregation

This can be identified as a major intangible heritage event related to Sri Lanka which occurs annually. It is associated with many *Agents* such as dancers, performers, singers, etc.

The *Kandy Esala Perahera* festival is a tradition and a conceptual entity. The local community carries out the festival as an annual event, which is modeled as an *Instantiation* in the CHDE model. For example, the *Performance in 2016* is a single intangible cultural heritage *Instantiation* (Figure 18). This specific instantiation is the physical existence of the *Kandy Esala Perahera* which can be recorded into physical and/or digital resources. Thus, *Kandy Esala Perahera* can be instantiated at a particular time and place, and by/with particular agents. The *Performance in 2016 Instantiation* of *Kandy Esala Perahera* can be captured by different physical media such as a printed photo or a VHS tape. Also, there can be a printed performance schedule which describes the *Performance in 2016.* None of these resources are connected to the networked information environment and all exist as *Offline Resources* in the *Physical Space*. Later, these *Offline Resources* are converted into *Digital Resources*, such as a JPEG image and MPEG file on a website. Similarly, the Printed Performance Schedule can be converted into an HTML webpage. Finally, all these collected *Digital Resources* are aggregated under the CDI which represents the *Performance in 2016* (Figure 18).

An intangible cultural heritage entity is frequently accompanied by various tangible cultural heritage objects such as props and instruments, which are excluded from Figure 18 to keep the figure simple. Generally, these tangible cultural heritage objects are utilized during some intangible cultural heritage activities, e.g., flags, masks and costumes used during *Kandy Esala Perahera*. Similarly, some tangible cultural heritage objects are produced during an intangible cultural heritage activity, e.g., traditional dancing costumes (known as *"Udarata Wes Andum Kattalaya"* in Sinhala language) worn by the dancers during the *Kandy Esala Perahera,* can be considered as tangible cultural heritage objects produced by the traditional craftsmen in the region).

6.1.2. Aggregation into the Curated Digital instance (CDI)

The CDI (Curated Digital Instance) entity in the model acts as a container of aggregated resources and corresponds to a cultural heritage object in the *Physical Space*, that is, a tangible cultural heritage object or an *Instantiation* of an intangible cultural heritage entity. Figure 19 shows an example of the aggregation in the *Digital Space* based on the CDI.

Starting from the bottom, the photo, video, and audio instances are digital resources related to the artifact *Statue of Tara* which is a tangible cultural heritage object in the *Physical Space*. These resources are collected and aggregated into the CDI. According to the One-to-One Principle of Metadata each of these digital resources has its own metadata denoted by M/D in Figure 19.

External resources such as Wikipedia article and a BBC article has some useful information related to the *Statue of Tara* which can be linked to enrich the CDI. The descriptions about the *Statue of Tara* on the right side is another critical information which should be aggregated to the CDI. The original description is from the British Museum and this can be linked to the CDI via a URL.

All these information resources have their own metadata and even the CDI has its own metadata descriptions which is at the "meta-metadata" level. The final result would be a CDI consisting of a combination of digital resources (corresponding to the original heritage object), descriptions of the original heritage object, and some linked (related) external resources. All these resources are individually identified by their metadata and aggregation is also based on the metadata description of these individual resources.

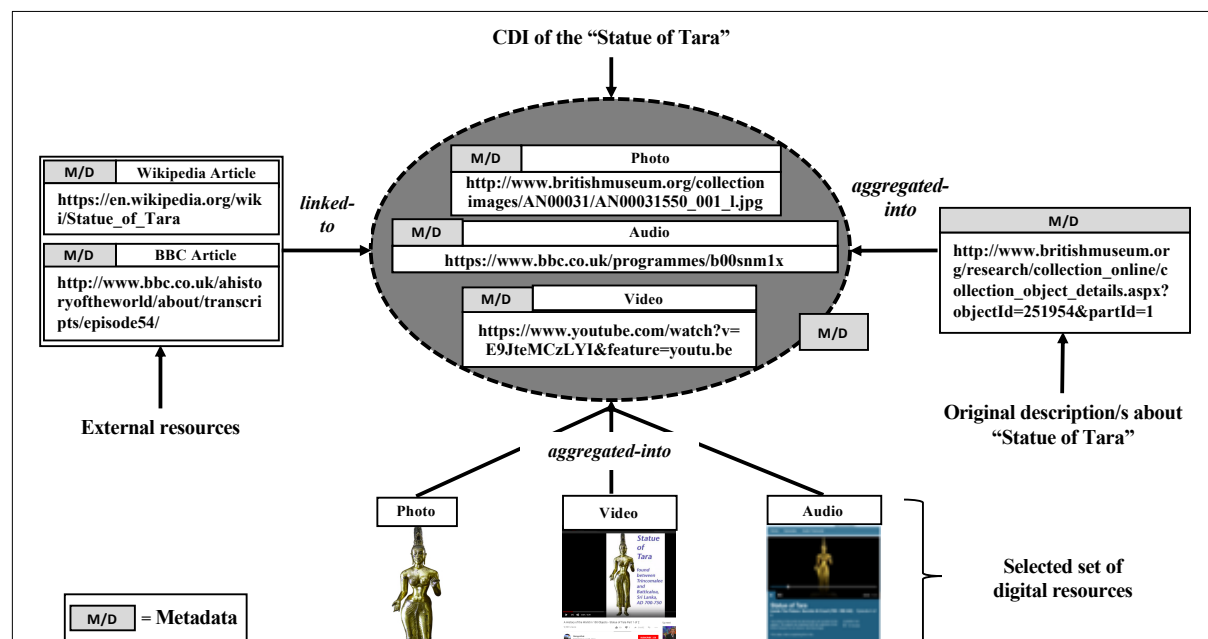**Figure 19. Example Depicting the Curated Digital Instance (CDI)**

Identification of the resources which should be aggregated is a decision by the institution and the study only proposed a conceptual model to help that aggregation. Besides, as discussed before (Section 4.2), a collection of CDIs is the foundation to form a digital archive of cultural heritage and it is the main outcome of the whole CHDE model.

6.1.3. Representation in RDF

In the Linked Open Data (LOD) environment, Resource Description Framework (RDF) plays a prominent role and converting the CHDE into a formal model is very helpful for this in the LOD environment. Figure 20 shows an RDF-based representation of a few entities of the CHDE model applied to *Kandy Esala Perahera*. Here, a few CHI instances are shown using simple RDF triples.
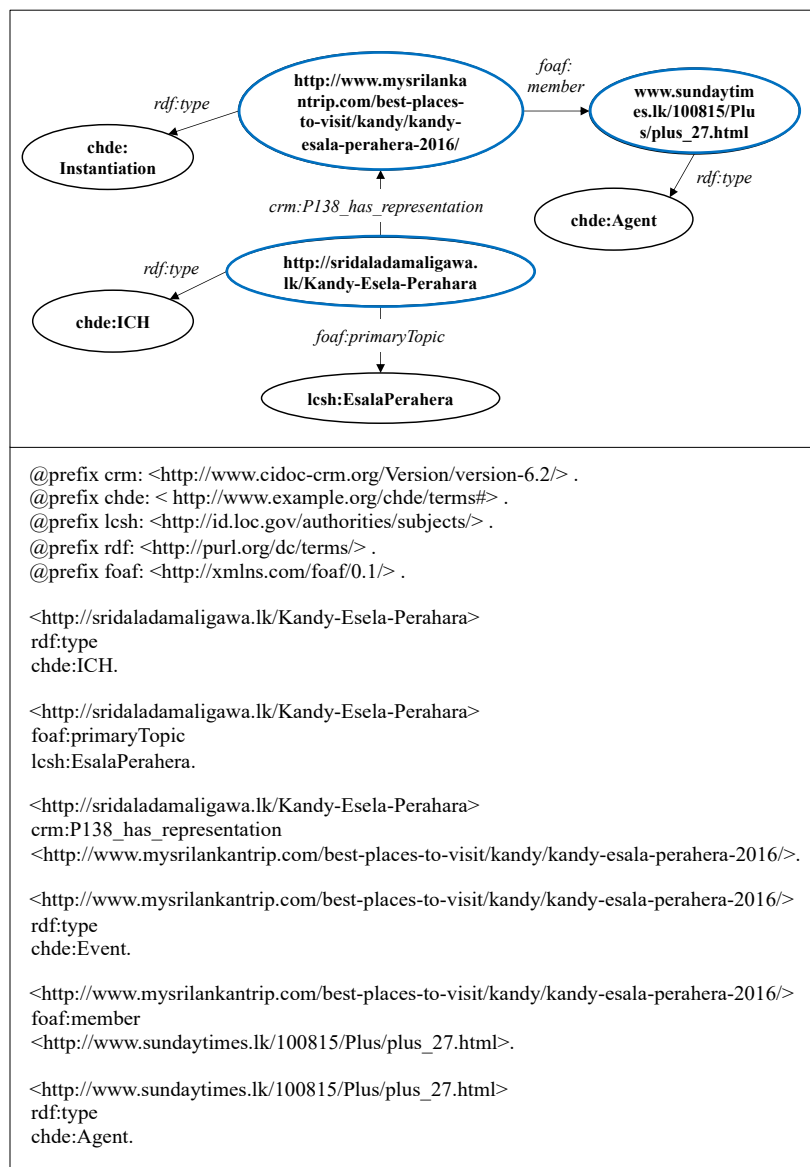


```
@prefix crm: <http://www.cidoc-crm.org/Version/version-6.2/> .
@prefix chde: < http://www.example.org/chde/terms#> .
@prefix lcsh: <http://id.loc.gov/authorities/subjects/> .
@prefix rdf: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://sridaladamaligawa.lk/Kandy-Esela-Perahara>
rdf:type
chde:ICH.

<http://sridaladamaligawa.lk/Kandy-Esela-Perahara>
foaf:primaryTopic
lcsh:EsalaPerahera.

<http://sridaladamaligawa.lk/Kandy-Esela-Perahara>
crm:P138_has_representation
<http://www.mysrilankantrip.com/best-places-to-visit/kandy/kandy-esala-perahera-2016/>.

<http://www.mysrilankantrip.com/best-places-to-visit/kandy/kandy-esala-perahera-2016/>
rdf:type
chde:Event.

<http://www.mysrilankantrip.com/best-places-to-visit/kandy/kandy-esala-perahera-2016/>
foaf:member
<http://www.sundaytimes.lk/100815/Plus/plus_27.html>.

<http://www.sundaytimes.lk/100815/Plus/plus_27.html>
rdf:type
chde:Agent.
```

Figure 20. Fragment of the CHDE Model in RDF

Resource URIs used in Figure 20 are as follows. (i) Intangible cultural heritage: as *Kandy Esala Perahera* festival in Sri Lanka (*http://sridaladamaligawa.lk/Kandy-Esela-Perahara*) (ii) *Instantiation* of the intangible cultural heritage as performance in 2016 (*http://www.mysrilankantrip.com/best-places-to-visit/kandy/kandy-esala-perahera-2016/*)

(iii) *Agent* as a famous dancer (*www.sundaytimes.lk/100815/Plus/plus_27.html*). Each resource is connected to its own classes and authors used the existing vocabularies to depict class property relationships such as AAT, LCSH, FOAF, RDF, etc.

6.2. Class-Based Mapping for Aggregation via the Description Modules Model

As described in Chapter 5, the Descriptions Module model is the foundation for the metadata mapping in the CHDE model. This kind of semantic mapping is essential to realize the CHI aggregation and it gives justification to the aggregation as well.

As the aggregation is intended to carry out as property-level mapping trough classes appropriate classes are required to perform this task. A class defines a "Type" of an object/instance and it provides classification to an object. According to the RDF Schema 1.1, "Resources may be divided into groups called classes. The members of a class are known as instances of the class. Classes are themselves resources……. A class may be a member of its own class extension and may be an instance of itself"[42]. Accordingly, a *Description Module* is a data entity which can be an instance of a class. The *Description Modules* and the *Facets* act as hints for the classes which are required in the mapping.

Subsequently, the author created a set of classes covering all the entities in the Description Modules model (Figure 21).
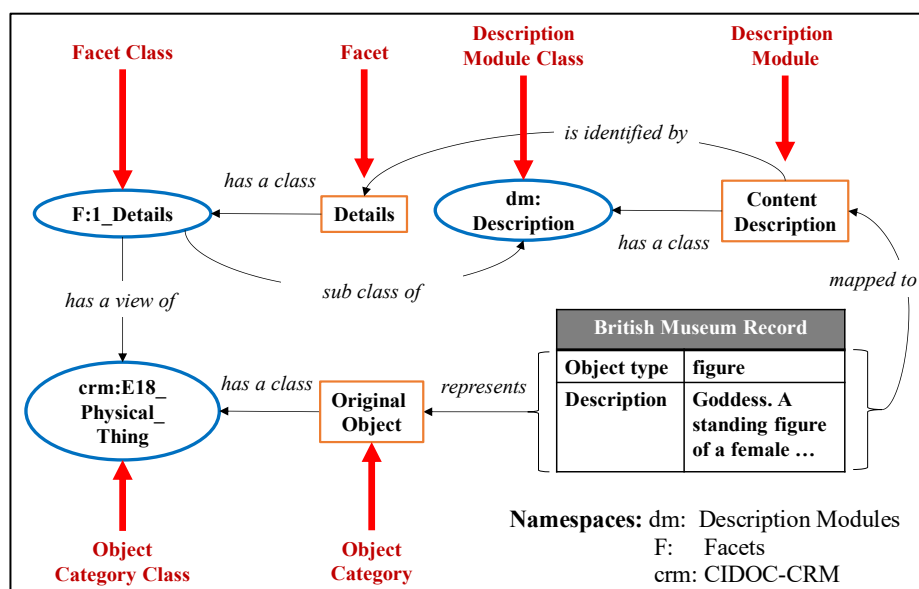

Figure 21. Relationships Between the Facets, Description Modules, Object Categories and Their Classes

---

[42]https://www.w3.org/TR/rdf-schema/#ch_class

Figure 21 shows all the related entities in the Description Modules model, e.g., *Object Categories, Description Modules*, *Facet*s and their corresponding classes, etc. All these classes in the Description Modules model are defined in Table 7 as separate sections (Table 7(a), 7(b) and 7(c)). The research used a few CIDOC-CRM classes in Table 7 (a) to define the classes for the *Object Categories* and other classes are defined by the author, based on the requirements of the Description Modules model entities.

Table 7. Classes and Definitions Used in the Description Modules Model

Table 7 (a). Classes Defined for Object Categories			
Entity Type	**Object Category**	**Object Category Class**	**Definition**
Object Categories	Original Object	crm:E28_Conceptual _Object	This class comprises non-material products of our minds and other human-produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication[43].
		crm:E18_Physical_ Thing	This class comprises all persistent physical items with a relatively stable form, man-made or natural[44].
	Digital Surrogate	crm:E73_Informatio n _Object	This class comprises identifiable immaterial items, such as a poem, jokes, data sets, images, texts, multimedia objects, procedural prescriptions, computer program code, algorithm or mathematical formulae, that have an objectively recognizable structure and are documented as single units[45].
	Administrative	dm:Admin	All the administrative information

---

[43]http://www.cidoc-crm.org/Entity/e28-conceptual-object/version-6.2
[44]http://www.cidoc-crm.org/Entity/e18-physical-thing/version-6.2
[45]http://www.cidoc-crm.org/Entity/e73-information-object/version-6.2

Entity Type	Description Module	Description Module Class	Definition
			related to an object falls under this category. Acquisition, identification, ownership and copyright information, etc. can be categorized under this class.
	External Resource	crm:E51_Contact_ Point	This class comprises identifiers employed, or understood, by communication services to direct communications to an instance of E39 Actor. These include E-mail addresses, telephone numbers, post office boxes, Fax numbers, URLs, etc.[46]

Table 7 (b). Classes Defined for Description Modules

Entity Type	Description Module	Description Module Class	Definition
Description Modules	Content Description Module	dm:Description	This class contains all textual descriptions related to tangible/ intangible heritage objects, digital surrogates and their administrative data which gives contextual meanings. (e.g., title, description, subject, category, acquisition notes, registration number, etc.)
	Agent Module	dm:Agent	All the person/s, organizations responsible of providing or creating an original object related to tangible/ intangible heritage, digital surrogates and their administrative data (e.g., painter, sculptor, actor, dancer or data provider, etc.)
	Location Module	dm:Spatial	All locational details related to original objects their digital

---

[46]http://www.cidoc-crm.org/Entity/e51-contact-point/version-6.2

			surrogates or administrative data (e.g., country, coordinates, finds-spot, gallery location, etc.)
	Timeline Module	dm:Temporal	Time-line information related to tangible/ intangible heritage, digital surrogates and their administrative data (e.g., period, date, era or acquisition date)
	Technical Description Module	dm:Tech	Measurable characteristics of a tangible/ intangible heritage object, digital surrogate (e.g., dimension, disc size, herbal quantities use to make traditional medicine)
	Provenance Module	dm:Prov	The history of the ownership, transmission of an object, and any change history related to a digital or original asset which belongs to administrative category (e.g., exhibition history of an object)
	Rights Module	dm:Rights	All copyright and ownership information associated with original objects or digital surrogates (e.g., copyright of a photo, owner of an artifact)
	External Link Module	dm:Links	URIs directed to external resources which give further information to the objects (e.g., links, maps, additional images, etc.)
	Bibliographic Module	dm: Biblio	All bibliographical resources related to original objects or digital surrogates (e.g., references, publications, etc.)

Table 7 (c). Classes Defined for Facets			
Entity Type	**Facet**	**Facet Class**	**Definition**
Facets	Details	F:1_Details	All contextual information related to an object (which may be physical or digital)
	Language	F:2_Language	Language used to describe an object
	Person	F:3_Person	People involved in creating an object
	Group	F:4_Group	Group or organization/s involved creating an object
	Location	F:5_Location	Place information related an object
	Date	F:6_Date	Specific date/s related an object
	Period	F:7_Period	Specific period/s related to an object
	Material	F:8_Material	Materials used to create an object
	Dimension	F:9_Dimension	Measurements related to an object
	Format	F:10_Format	Format of an object
	Type	F:11_Type	Type or category of an object
	Tech Details	F:12_Tech Details	Additional technical details
	Object No	F:13_Object No	Identification details of an object
	Provenance	F:14_Provenance	Provenance details related
	Rights	F:15_Rights	rights details related
	Related Links	F:16_Related Links	External links related to an object
	Bibliographic	F:17_Bibliographic	Bibliographical]: information related to an object
Key: crm: CIDOC-CRM Schema dm: Description Modules Schema F: Facets			

Also, we identified the utilization of *Description Modules* and *Facet* and how they can be used to identify the metadata and object of cultural heritage. If we consider the metadata creation point of view, metadata creators can create CHI based on the *Application Schema View* of an object mainly based on the schemas such as *Description Modules* and *Object View* of an entity via *Facets*. These two levels of standing points are important in the heritage

information retrieval domain as well. Figure 22 depicts the *Object View* and *Application Schema View* of a CHI entity via *Description Modules* and *Facets*.
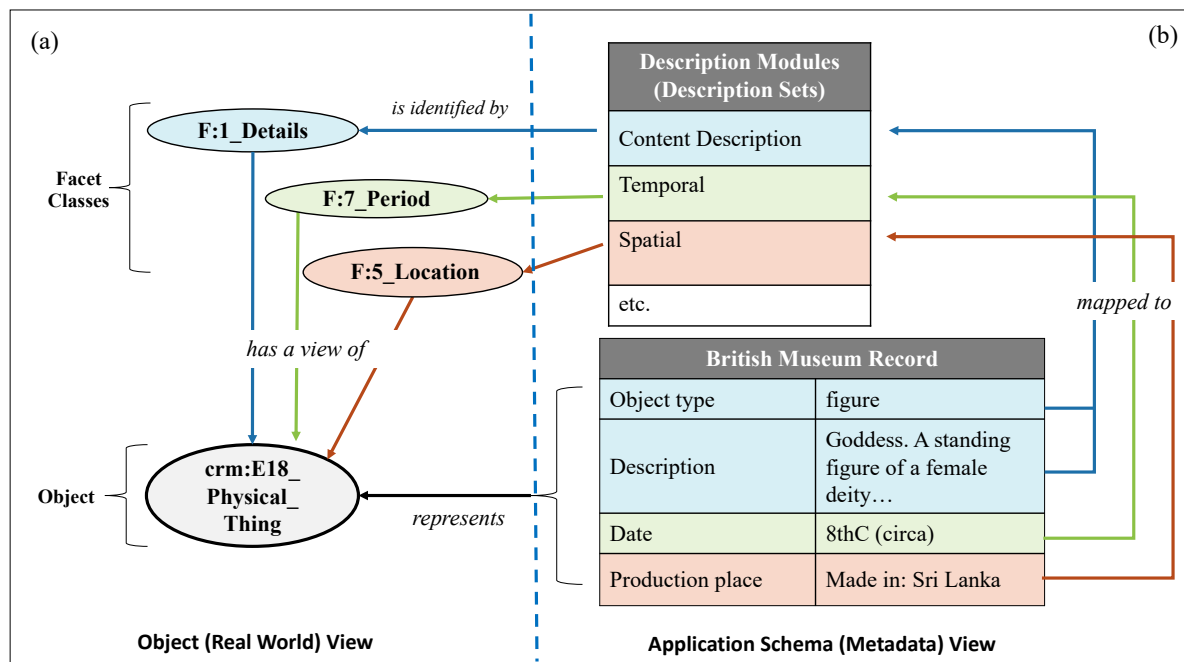

Figure 22. Object View and Application Schema View of a CHI Entity via Description Modules and Facets

Section (a) of Figure 22 shows a few *Facet Classes* (*F:1_Details, F:7_Period, F:5_Location*) related to an *Object* (*crm:E18_Physical_Thing*) which can be linked through various viewpoints. The *crm:E18_Physical_Thing* class was taken from the CIDOC-CRM ontology as it is used to define objects in the real world and the *Facet* classes are defined by the author (Table 7). Section (b) of Figure 22 shows the *Application Schema* view of a CHI record[47] instance related to the same Object in the real world. The metadata descriptions of that CHI record can be mapped to the *Description Modules* as it acts as a structure to identify metadata. Therefore, the study identifies the metadata descriptions of the CHI record and the schema-like structure of the *Description Modules* as the implementation-based, *Application Schema View* related to the same *Object*. Therefore, the Description Modules model provides these two types of viewpoints to identify an Object while mapping these two ends enabling aggregation.

47

https://www.britishmuseum.org/research/collection_online/collection_object_details.aspx?objectId=251954&partId=1

70

The relationships between the objects, classes and the metadata descriptions from three different Web services are presented in the following figure (Figure 23).
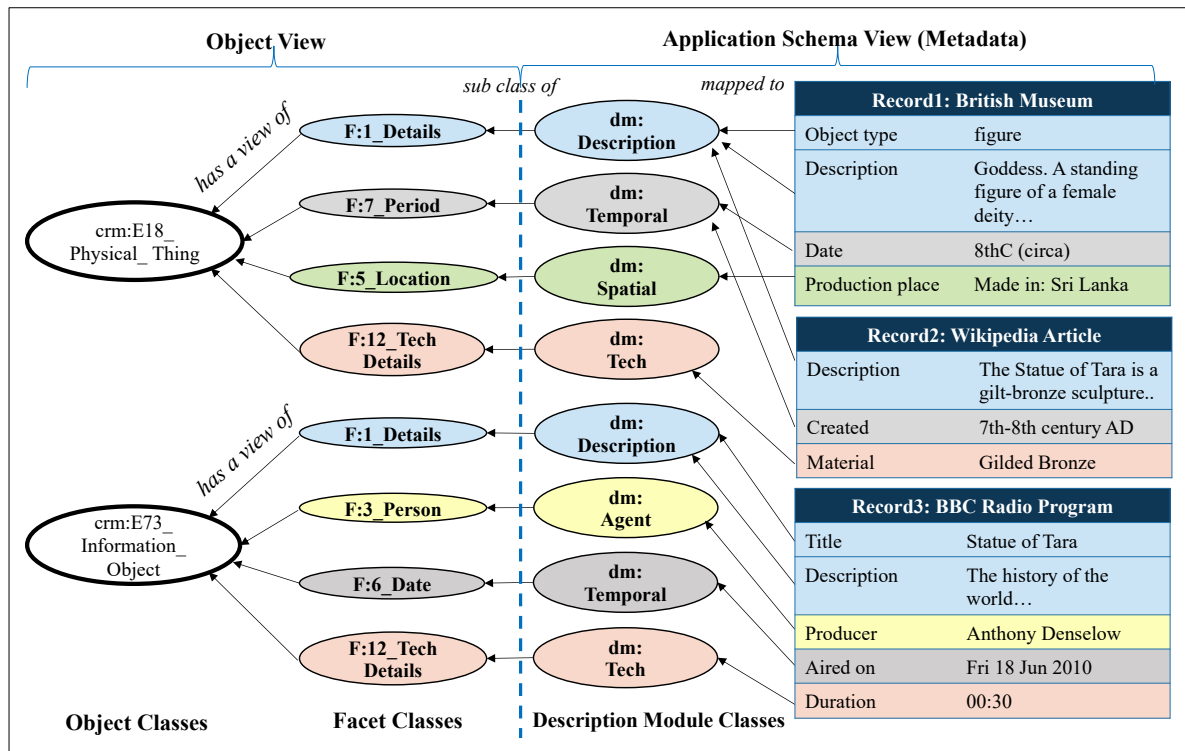

Figure 23. Mapping Diverse Schemas via Classes

According to Figure 23, we can clearly identify the *Object View* and *Application Schema View* of a cultural heritage object and its data separately. These two views are like two faces of a single coin. Somehow, bridging these two ends could be achieved through the *Facets* and *Description Module Classes*. Starting from the left-hand side of Figure 23, we can identify two objects related to the physical artifact (*crm:E18_Physical_ Thing*) and its digital object/surrogate *(crm:E73_Information_Object)*. So, these objects in the real world represented via *Object Classes* can be connected to the *Facet Classes* (*F:1_Details, F:5_Location,* etc.). *Facet Classes* are derived from the *Facets* categories created previously (Table 7 and Appendix 3) and they describe the attributes of an *Object* from different angles. The *Facet Classes* have a relationship between the *Description Module Classes*. As identified in Figure 22, the *Description Module* can be identified by a *Facet*. When it comes to *Classes* (Figure 23), *Facet Class* might be a "*subclass of*" the *Description Module Class* and it can be used to bridge the gap between the *Object View* and *Application Schema View* ends. In addition, *Description Module Classes* can be used to map the metadata descriptions provided by different Web services and act as the main aggregator in the Description Modules model.

As a result, the *Application Schema View* of the Description Modules model always represents the implementation-level of metadata.

The idea of viewing objects via *Facet* and class-based mapping will be further discussed under Chapter 7 (Section 7.5).

6.3. Discussion

This chapter overviewed a few use case examples based on the proposed CHDE model and the Description Modules model.

Figure 18 depicts two use cases related to tangible and intangible cultural heritage aggregation of the CHDE model. The two examples are based on real cultural heritage entities and all the resources are extracted from the Web one way or another. The conceptual aggregation examples give some insight into how this CHDE model works in real situations.

Section 6.1.2 explained the main aggregation scenario of the CHDE model. CDI, which works as the main aggregated entity, aggregates many resources varying from digital media resources to metadata descriptions. Every resource consists of metadata and the CDI aggregates them all into a single platform.

Next, the author explained the usage of RDF in the CHDE model using a fragment of the CHDE model described using RDF triples (Figure 20). As CHDE entities are general and flexible it can be realized in RDF format easily and utilizing these techniques is essential to realizing the CHI in LOD environment.

Finally, this chapter discussed how the Description Modules model and its entities work to aggregate CHI. To do so, the study defined some classes related to the Description Modules model (Figure 21 and Table 7). The *Description Modules* and the *Facets* identified two different perspectives of an *Object* and the classes are used to bridge these two ends by connecting the *Object View* and the *Application Schema View* of an *Object* (Figure 22 and 23). The aggregation example (Figure 23) clearly showed these two perspectives of the *Object* and how these spaces can be connected with the help of the *Facet Classes* and *Description Module Classes*.

Chapter 7: Results and Discussion

Organizing and linking institutional and non-institutional CHI for better access to and for richer context about cultural heritage was the main goal of this study. This thesis proposed two metadata models to organize and aggregate CHI about heterogeneous cultural heritage entities in the networked information environments.

Chapter 7 discusses the whole study presented in this thesis in accordance with the two main research questions and requirements for the metadata models presented in Section 2.4. In addition, this chapter addresses the issues encountered and the limitations of the study.

7.1. Metadata Models to Aggregate and Organize Institutional and Non-Institutional CHI

The first research question of this thesis was "How can we model metadata for digital objects to be created by aggregation of fragments extracted from existing digital archives and other Web resources?" This question is connected to the requirement of providing users with more contextual information by linking institutional and non-institutional CHI. From the analysis about the current CHI aggregation models, platforms and related researches shown in Chapter 3, the author recognized that South and Southeast Asia need a novel metadata model for CHI aggregation to organize cultural digital archives collecting CHI from the Web.

7.1.1. A Model to Define Entities for the Development of Digital Archives

The study proposed a generalized metadata model called CHDE to help develop a digital archive of cultural heritage in Chapter 4.

The CHDE model gives some insights into aggregating any kind of information corresponding to cultural heritage in the physical space (Figure 6). The author found that existing metadata models for digital archives are usually designed to aggregate information resources about tangible cultural heritage objects. In contrast, CHDE is designed to aggregate both tangible and intangible cultural objects. CHDE consists of more broad and general entities (Table 3) which enable absorption of any type of CHI without difficulty. The physical and digital spaces of a cultural heritage object and its information are precisely identified in the CHDE model. This kind of differentiation is important when identifying the boundaries of

information resources during the digital curation process which is essential for building a digital archive.

CDI is the main aggregated instance in the CHDE model, and it consists of several digital resources, metadata descriptions related to a single cultural heritage object in the physical environment. As shown in Figures 8 and 19, CDI aggregates individual items, their original metadata descriptions plus their related external resources such as Wikipedia articles. A CDI entity aggregates all these resources while providing users a more enriched form of CHI by linking institutional and non-institutional metadata descriptions. Therefore, CDI is a set of digital resources connected to a single tangible or intangible cultural heritage instance which enables item-level aggregation. Collection of these CDIs leads to collection-level data aggregation forming a digital archive. A CDI is composed of not only digital surrogates of the cultural heritage instance, e.g., photographs and videos, but also contextual information about the cultural heritage instance, e.g., links to an external object and metadata descriptions. A CDI is defined as a structural entity which solves the requirement of this study "identification of facets required for aggregation."

Another requirement of this model is "identification of the objective of metadata aggregation, using the One-to-One Principle of Metadata". Since CHDE collects and aggregate diverse information from institutional and non-institutional sources, it may have to handle various metadata schemas and mixed forms of metadata. For the aggregation process, the author used the One-to-One Principle of Metadata as the key concept of building up the CHDE model and each level of the CHDE model adheres to the above One-to-One Principle. Since a single cultural heritage object can be associated with many digital surrogates, CHDE used the One-to-One Principle as a core rule to identify the digital surrogates (resources) and their original resource descriptions separately. As in many cases, metadata of a cultural heritage digital archive is hybrid (Figure 4), the study needed to identify objects described by source metadata which should be aggregated into a CDI. In other words, we have to identify an object corresponding to a description component in the source metadata. Figure 17, 18 and 19 conceptually show this metadata aggregation scheme.

7.1.2. A Model to Define Metadata Mapping to Aggregate Metadata Across Diverse Schemas

While CHDE gives an abstract level aggregation idea for CHI resources, finding avenues specifically for aggregation of CHI metadata is a crucial aspect in this research. As the

identification of target objects of component descriptions in a hybrid record used for metadata aggregation is a key issue in this research, this thesis introduced the Description Modules model which is designed specifically for identification and aggregation of components of hybrid record metadata (Chapter 5). The author considered that the DCMI Application Profiles model, in particular, its *Description Set* Profiles model works well to build the model of metadata aggregation for CDI.

Section 2.2.2 of this thesis discussed a few issues and challenges associated with metadata aggregation based on property level mapping. One such critical issue is the "risk of losing the context of the properties". In addition, a large amount of property-to-property level mapping of metadata is very complicated and time-consuming. When the number of schemas grows, the mapping combinations also grow exponentially, causing unmanageable and complex mapping (Table 1). Because of these factors, the author proposed finding some semantic groups of descriptive elements in each application metadata schema called *Description Modules* (Table 6). Module based mapping is identified as an effective way to integrate information. Duval et al. say, "in a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way" (Duval, Hodgins, Sutton, & Weibel, 2002).

The entities that are shown in the CHDE model (Figure 6) are the objects of higher-level classes such as *Digital Resource* and *Curated Digital Instance* which can be expressed as a *Description Set*. Therefore, classes of the objects described by components of the *Description Set*, i.e., Descriptions, can be used in metadata matching for aggregation. For example, in Figure 6, metadata of every entity of a *"selected set of digital resources"* has to have one or more Descriptions which can be consistently aggregated with corresponding Descriptions of other entities. As a CHI record is a set of descriptions, the author applied the idea of DCAP's *Description Set* and created these *Object Categories* and *Description Modules* to help support the aggregation.

The Same idea gives a solution for the hybrid record issue as well (Figure 4). Description Modules model clusters the metadata descriptions according to their objectives which is based on the One-to-One Principle of Metadata and it gives some insights to fulfil the requirement of "identifying objectives of metadata aggregation, using the One-to-One Principle of Metadata as a foundation".

However, *Description Modules* and *Object Categories* are not sufficient to describe an object as a whole. From the database design perspective, *Description Modules* gives a high-level abstraction of information integration via their schema. Information retrieval would,

however, need a different view as they need more specific and finer descriptions about objects to be retrieved.

The *Facets* introduced in Section 5.4 (Appendix 3) helps the linkage between Descriptions across different schemas. A *Facet* of a cultural object in the model has a one-to-one relation with a *Description Module* of application metadata and it shows an aspect for Descriptions. The *Application Schema View* and the *Object View*s (Figure 22) are essential for database creation, as they consist of two different viewpoints. *Facets* represent the *Object View*s while the *Description Modules* represent the *Application Schema View* of the same. For instance, the involvement of ShEx (Shape Expressions)[48] and SHACL (Shapes Constraint Language)[49] are directly dealing with the data structure and it is something to do with the implementation schemas. Nevertheless, a schema which describes both ends (implemented depending on the *Application Schema* or data structure and implementation neutral *Object View*) is not visible in the existing information domains. Perhaps this kind of conception is important for accessibility of CHI over the Web and for many other activities as well.

To support this idea, the author has created classes and definitions to represent each entity in the *Object Categories, Description Modules* and *Facets* in the Description Modules model (Table 7(a), 7(b),7(c)). The *Facets* helped to identify the classes which are useful to describe the connection between entities in the *Description Modules* (Figures 23). Giving a class to a *Facet* means giving meaning to an *Object.* Also, these *Facets* are intentionally based on the existing CHI instances and schemas and it cannot be considered as a complete set. Therefore, the number of *Facets* can be changed based on the diverse schemas used in the model (Appendix 3).

Figure 22 presents the use of *Facet* and *Description Modules* as two different ways to view an *Object*. Basically, each *Facet* has a *is part of* relationship with a Description module and *Facets* has its own classes (e.g., F:1_Details, F:7_Period, F:5_Location) which can be used to connect and view the original object (crm:E18_ Physical_Thing).

Figure 22 is also connected to the same discussion. In Figure 22, we can identify that the *Description Module* classes (e.g., dm:Description, dm:Temporal, dm:Spatial etc.) are directly connected with the metadata of the record. However, identification of individual descriptions based on the one-to-one relationship can be performed through *Facet* classes (F:1_Details, F:7_Period, F:5_Location, etc.) only. Also, *Facet* classes can simply connect to

---

[48]http://shexspec.github.io/primer/
[49]https://www.w3.org/TR/shacl/

the *Object* classes (crm:E18_ Physical_ Thing and crm:E73_ Information_ Object) providing a filter to view the *Object* via *Facets.*

7.2.   CHDE Curation Process

The requirement of "Identification of "Entities" in the process of organizing CHI" is fulfilled by the CHDE curation process in the CHDE model as follows. Figure 9 describes the CHI creation, organization, and dissemination as a straightforward process. This section does not discuss the digital curation process as a whole. For instance, groundwork such as data collection or creation, data preservation, and maintenance are out of the scope of this discussion. The CHDE curation process presented in this thesis identifies three main tasks known as *Ingestion, Description,* and *Organization* which are mainly involved in the CHDE.

Each step of the CHDE curation process was described in Section 4.2 and the final output of the entire process is a *Digital Archive*. The CHDE curation process exists behind the CHDE model and it shows the step-by-step process of making CDI and finally a *Digital Archive*. The CHDE process has two types of aggregations. The first one is the CDI aggregation, and this is the main aggregation phase of the CHDE Model which we have discussed previously (Section 6.1.2). The second one is the *Digital Archive* level aggregation. This aggregation was formed by a collection of *Curated Digital Instances* corresponding to a set of cultural heritage entities (Figure 9) which are finally aggregated as a composite *Digital Archive*.

This *Digital Archive* may act as a digital portal of CHI with more user-friendly, linked, contextualized information. Moreover, this kind of *Digital Archive* and CDI can provide collection-level CHI to the users and it may level up current conventional item-centric digital archival systems. However, the CHDE specifically presented the creation of CDI which is the foundation of creating a digital archive. The second level CDI aggregation which results in a digital archive is out of the scope of the current study.

7.3.   Semantic Relationship Between the CHDE and Existing Schema Classes

This study made a semantic mapping between CHDE classes and existing related schema classes (Table 8). The author defined nine CHDE classes (namespace= chde) and mapped these classes to the schema classes selected from OAI-ORE, EDM, CIDOC-CRM, FRBRoo, and DCMI. Moreover, AAT vocabulary was used as an additional resource to make this

formalization more meaningful. Many CHDE classes were *subclasses* (rdfs:subClassOf) of existing schema classes (e.g., *ore:Aggregation, dcmiType:Collection, crm:E7_Activity* etc.).

	CHDE Classes		Relationship with Existing Schema Classes	
	Class Label	**Class Name**	**Relationship**	**Schema Class**
1	Curated Digital Instance	chde:CDI	rdfs:subClassOf	ore:Aggregation
2	Collected Set of Digital Resources	chde:DigitalSet	rdfs:subClassOf	dcmiType:Collection
			rdfs:seeAlso	aat:Collection
3	Digital Resource	chde:Digital	rdfs:subClassOf	crm:E73_Information_Object
			rdfs:seeAlso	edm:WebResource
			rdfs:seeAlso	aat:Recording
			rdfs:seeAlso	aat:Digital
4	Collected Set of Offline Resources	chde:OfflineSet	rdfs:subClassOf	dcmiType:Collection
			rdfs:seeAlso	aat:Collection
5	Offline Resource	chde:Offline	rdfs:subClassOf	FRBRoo:F26_Recording
			rdfs:subClassOf	dcmi:PhysicalResource
			rdfs:seeAlso	aat:Recording
6	Instantiation	chde:Instantiation	dcterms:hasPart	edm:Event
			rdfs:subClassOf	crm:E7_Activity
			rdfs:seeAlso	aat:Event
			rdfs:seeAlso	aat:TimeBasedWorks
			rdfs:seeAlso	aat:PerformanceArt
7	Agent	chde:Agent	rdfs:subClassOf	crm:E39_Actor
			rdfs:subClassOf	dcmi:Agent
			rdfs:seeAlso	aat:Agent
8	Intangible Cultural Heritage	chde:ICH	rdfs:subClassOf	crm:E28_Conceptual_Object
			rdfs:seeAlso	aat:IntangibleCulturalHeritage
9	Tangible Cultural Heritage Object	chde:TCH	rdfs:subClassOf	crm:E18_Physical_Thing
			rdfs:seeAlso	aat:TangibleCulturalHeritage

Table 8: Semantic Relationship Between the CHDE and Existing Schema Classes

It was very difficult to find an appropriate schema class for *chde:Digital, chde:DigitalSet, chde:Offline* and *chde:OfflineSet*, therefore, the author had to select some general classes such as *dcmiType:Collection*.

Also, *chde:Instantiation* could not be fully or partially related to an existing class as it is something novel in the CHDE model. This study has selected classes related to the *Instantiation* (*chde:Instantiation*) from existing standard vocabularies as far as possible. Table 8 shows how these CHDE classes are semantically connected to the existing schemas. The definitions of these CHDE entities/labels were described previously (Table 3).

Meanwhile, Table 4 represented some formalization between *Instantiation* classes (discussed in Section 7.1.1) and Table 7 created classes and definitions related to the

Descriptions Modules model. However, in the *Description Modules* discussion, the author used only a few classes from the CIDOC-CRM schema (e.g., crm:E28_Conceptual_Object, crm:E18_Physical_ Thing, crm:E73_Information _Object, and crm:E51_Contact_ Point) and the rest of the other classes are created according to the requirement of the model (Table 7). The author identified that existing schema classes are not sufficient to express entities in the CHDE (e.g., *Digital Surrogate, Instantiation*) and Description Modules model (*Content Description Module*). As a result, the study defined its own classes to represent a few entities precisely.

The CHDE model is defined based on the RDF data model. Representing CHDE in RDF helps define the CHDE entities and their relationships in a formal scheme. This was discussed previously (Figure 20) and the author tried to present a fragment of the CHDE through RDF. This kind of RDF realization is essential as we need to link the CHI resources and realize them in the LOD environment.

7.4. Instantiation as a Physical Resource for Building Digital Archives of Intangible Cultural Heritage

The second research question asked in this research was, "how can we describe intangible cultural heritage for digital archives?" A digital archive of cultural heritage is primarily a collection of digital surrogates of cultural heritage objects. Those digital surrogates are mostly realized as visual and/or audio data. A fundamental requirement for creating a digital surrogate is that the original cultural heritage has to be presented in a form recordable into audio-visual media (Figure 7). However, capturing tangible and intangible heritage are slightly different as intangible heritage has no physical existence. Therefore, "identification of the "Object" of intangible cultural heritage" has become another requirement of this study.

Intangible cultural heritage is primarily composed of knowledge and skills inherited from our ancestors and can be presented as a performance by people or a group of people who own the knowledge and skills. Basically, intangible cultural heritage is associated with objects (tangible objects/byproducts of an intangible cultural heritage activity, e.g., handmade Japanese paper-washi papers), skills (skills associated with an intangible cultural heritage, e.g., dancing/singing skills), events (particular performance performed on a given time and location) and contexts (contextual/background information associated with intangible cultural heritage). Integration of these four factors would create a complete intangible cultural heritage entity. However, in reality, memory institutions usually present an intangible cultural

heritage as one or a few records, which is not quite correct for understanding the whole picture of the heritage.

A digital surrogate of intangible cultural heritage in a digital archive is not a surrogate of the intangible cultural heritage but a surrogate of a single performance. This distinction seems trivial but is important to properly organize digital archives of intangible cultural heritage. For instance, a memory institution might recognize a record taken at *Kandy Esala Perahera* in the year 2016 as an intangible cultural heritage entity of Sri Lanka (Figure 18). There is no way to digitally record *Kandy Esala Perahera* as an intangible cultural heritage entity, as *Kandy Esala Perahera* is an event which takes place every year. In principle, intangible cultural heritage cannot be digitized but only its *Instantiations* can, and this can be identified as the "Object" of an intangible cultural heritage entity. Aggregation of digital surrogates of *Instantiations* may be a quasi-surrogate for intangible cultural heritage: it has a significant advantage of showing historical and contextual information which is a crucial for digital archives of intangible cultural heritage.

Further, *Instantiation* may be used to model artifacts which physically exist only in a particular time period and place, that is, dynamic artifacts such as fireworks, installations, theater plays, etc. These artifacts may have one or more instantiations. Here, a clear distinction between the dynamic artifacts and their instantiations is useful to build a digital archive of the dynamic artifacts.

Digital archives have been created to record events such as natural and man-made disasters and activities such as sports and game plays. Unlike the dynamic artifacts referred to above, *Instantiation* may not apply to events and activities even if we use recording media to archive the events and activities because they are physical entities that existed at some point in time and location. CIDOC-CRM has *E5 Event* and *E7 Activity* which is a sub-class of *E5 Event*. As shown in Table 7, *Instantiation* is defined as a subclass of *E7 Activity* because *Instantiation* can be defined as an *"Activity"* specialized to present a physical entity from an artifact of intangible cultural heritage.

Table 4 presented the six *Instantiation* classes that were identified based on 5W1H questions (Section 4.1.1.). All these entities are important to understand an intangible heritage and *Instantiation* may be created based on one or more of these entities. The author tried to map *Instantiation* classes to existing schemas such as CIDOC-CRM, FRBRoo and Getty AAT (Table 4). For example, classes, such as E53 Place, E21 Person, E39 Actor, and E7 Activity correctly match with the meanings of the *Instantiation* classes developed for the CHDE model. FRBRoo had a handful of classes which can be mapped directly to the same

instance classes (e.g., F8 Event and F9 Place). Some classes can be mapped but they do not convey the same meanings as the instantiation classes. For instance, E55 Type can be mapped to category instance class. Also, F3 Manifestation Product Type is a subclass of the CIDOC E55 Type class. Nevertheless, F3 Manifestation Product Type real meaning is not compatible with the intangible instance class category.

7.5. Limitations

This section discusses the limitation of this study based on the research questions of this thesis.

Regarding the first research question "How can we model metadata for digital objects to be created by aggregation of fragments extracted from existing digital archives and other Web resources?", this study answers mainly in the abstract entity level defined in Chapter 4 and 5, but not in the levels of metadata creation or implementation. The CHDE model has the entities named *Collected Set of Digital Resources* and *Collected Set of Offline Resources*. These entities explicitly show that a *Curated Digital Instance* is an aggregation of existing archived resources in the entity level. On the other hand, the Description Modules model's *Description Modules* and *Facet* give metadata-centric approaches but still, they are describing a high-level abstraction only. Therefore, this study does not implement or create an application/ system based on the proposed metadata models.

The *Facet* view of an object is an interesting finding of this study. The *Facets* in the Description Modules model are created based on the properties extracted from ten different institutional and non-institutional Web services only. However, if we add more properties that might change the current list of *Facets* (Appendix 3) (Table 7). Therefore, *Facets* are not a fixed component but can be changed according to the requirements and properties used in the schema. This idea was introduced as a hint to identify and aggregate metadata of cultural heritage objects in diverse schemas. The outcome of this discussion was the *different viewpoints of a cultural heritage object* which can be utilized when creating databases and later on to support information search and retrieval. This idea is still immature and needs to be investigated in more detail in the future.

One of the aims of this research is to aggregate institutional and non-institutional CHI. Aggregating institutional CHI is not a big challenge as the information is created in a structured manner. This study gave a generic approach to aggregate any kind of CHI. However, aggregation of non-institutional CHI should be explored more in the future. The

reason is that the non-institutional CHI (e.g., Wikipedia) consists of complex information structures and they are mostly not based on formal and well-structured schemes. The CHDE and Modular Metadata models can be utilized in that context but, the real situations of utilizing this CHI for aggregation may be challenging and should be investigated in more detail.

Besides, the aggregation and mapping examples in the feasibility study (Chapter 6) of this study was manually done. Therefore, the study could not create any mass data aggregation based on automated or semi-automated application.

Instantiation of intangible cultural heritage entity is the key entity as an answer to the first research question, "how can we describe intangible cultural heritage for digital archives? The author identified that we need to use physically existing entities but not abstract entities as an instance to be curated into a digital archive. CIDOC-CRM has class Activity, which is used as a super-class of *Instantiation* in the mapping table (Table 8) shown in section 7.2. However, the author realized that existing schema classes cannot fully represent *Instantiations* in the CHDE. CHDE *Instantiation* is intended to define activities which can be recorded into physical media whose contents can be collected into a digital archive. A single intangible cultural heritage may be presented as a collection of *Instantiations*. According to the CHDE, *Instantiation* is a physical entity and it is recorded into *Offline Resources*. Curators select these recorded *Instantiations* based on their institutional policies.

Nevertheless, the first research question "How can we describe intangible cultural heritage for digital archives?" would include contextual and historical descriptions about an intangible cultural heritage, which means that we would need to collect not only digital surrogates of *Instantiations* but also those resources that explain the intangible cultural heritage, such as Wikipedia articles and websites. On one hand, this point is related to the first research question as discussed in section 7.1. Moreover, there would be such cases that identification of a single intangible cultural heritage or a single *Instantiation* is not very clear because of the diverse nature of intangible cultural heritage. However, such identification and selection should be solved by domain specialists and is out of the scope of this study.

Finally, this study faced many incompatibilities to map the CHDE entities to the existing schemas such as CIDOC-CRM. This is due to the insufficiency of cultural heritage schemas and vocabularies which is a general CHI issue in the region. Therefore, the author sometimes had to create her own classes to describe entities in the CHDE model and Description Modules model. Creating our own classes might create problems in the long run,

but the author tried to minimize this by mapping many classes to the existing schemas and giving appropriate definitions to the newly created classes (Table 7) (Table 8).

Chapter 8: Conclusion and Future Direction

This research concentrated on organizing and connecting dispersed digital information and creating a digital archive for cultural heritage in the South and Southeast Asia. If we recall the background of this research, the study initiated as a Sri Lankan museum information aggregation effort and later on expanded to other regions aiming for boarder CHI aggregation. The preliminary survey of this research recognized that the user needs are diverse, and they need multiple information resources related to a single tangible or intangible heritage object. In addition, the author found that conventional institutional heritage metadata are more item-centric and non-institutional metadata are more comprehensive and context-rich. Currently, memory institutions such as libraries are shifting from their item-centric data to context-oriented data platforms enabling users more contextual information. For instance, the FRBR-WEMI model is one such intervention in the library domain. Cultural heritage institutions also need such perspective and metadata aggregation based on conceptual models to be a key methodology to achieve this intention.

During preliminary investigations of this study, the author identified some general issues associated with the South and Southeast Asian CHI domain. Specially the scarcity of web-based CHI and lack of standardized digital archives are the main issues identified during this investigation. The same problem applied to the authority/controlled vocabularies in cultural heritage. Therefore, the development of digital archives of tangible and intangible cultural heritage in the region has become a challenging effort. Additionally, the limited information delivered by existing digital archives is not enough to fulfill the diverse user requirements.

Apart from the regional issues, the author investigated a few renowned existing CHI aggregation and organization models plus related researches prior to developing the proposed metadata models. During these investigations, the author found some shortcomings of those models and realized that they cannot be fully adapted to aggregate CHI of the intended study area.

Therefore, as a solution to these regional CHI issues and existing model related issues, the thesis proposed abstract-level metadata models to organize and aggregate CHI in networked information environments. This main aim was supported with the intention of creating a context-oriented data platform which is beyond the conventional item-centric CHI perspective. The resource identification and integration were done along with the One-to-One Principle of metadata and it makes a clear distinction between the CHI and its original object.

The final outcome of the CHDE is an enriched set of CHIs related to cultural heritage in the physical environment which can be used later to form a larger digital archive of cultural heritage.

Through the crosswalk done between the CHDE model and CIDOC-CRM, FRBRoo and other vocabularies, the author tried to formalize the CHDE entities to understand them precisely. The deviation of tangible and intangible cultural heritage and their physical, digital resources is not expressed entirely through existing cultural heritage schemas or models. Therefore, developing a generalized model such as the CHDE model can be a solution to distinguish physical and digital entities of a cultural heritage asset in a diverse environment.

Primarily, digital archives related to cultural heritage materials are common and they possess many digital surrogates related to various cultural heritage objects, e.g., Europeana. However, adding contextual information into the same archive is out of the scope of most of those archives. The CHDE model tries to aggregate digital media resources, related contextual information and related external resources via an entity called *Curated Digital Instance* which is distinctive from conventional digital archives.

The CHDE model relies on the existing information on the Web provided by various standardized services which may be institutional or non-institutional. This is a solution to the scarcity of heritage information provided by institutions in the study area.

The CHDE model gives a novel idea to organize intangible cultural heritage which is not clearly visible in most of the existing cultural heritage schemas and models. Identifying the "Object" of the intangible cultural heritage via *Instantiation* is a promising approach to organize abstract-level intangible cultural heritage into a physical entity within a digital archive. *Instantiation* based CHI organization proposed through the CHDE model can be utilized by institutions to organize their intangible cultural heritage resources based on aggregation according to their requirements. Since this research does not provide any criteria to create or select *Instantiations,* database creators and digital curators can freely use the *Instantiation* idea and organize their information as they want.

The latter part of the study specifically focused on the metadata descriptions of CHI instances provided by both institutions and non-institutional services. One such common issue was hybrid records in the digital archives. Therefore, this study tried to give a solution to identify and categorize CHI related metadata descriptions collected from institutional and non-institutional services to develop a data model for metadata aggregation. The author has found the One-to-One Principle and DCAP have crucial roles for metadata aggregation. As a result, the study proposed the Description Modules model which explicitly identifies

metadata components based on their objectives. *Description Modules* are designed as common modules among different schemas.

Although the Description Module model might be seen as "another common schema for metadata identification", the study introduced a concept termed as *Facets*. *Facets* grouped the attributes and it does not represent a structure like *Description Modules* but, it bridges the gap between the metadata and the Object by reducing the space for mapping. Also, *Facet* identifies a view of an Object based on the One-to-One principle of metadata concept, which is the key concept in this research. Even though the *Facets* are derived from *Description Modules*, *Description Modules* and *Facets* play discrete roles in this study. While *Facet* represents the Surface View of an object, *Description Modules* deals with the metadata of the same object. Current schemas have no such clear separation of "Object" and "Data" but, this kind of perspective is essential for database creation, metadata aggregation and information retrieval.

Both models proposed in this research are intended for CHI aggregation. However, since the models are abstract and general they can be used in other domains as well. For instance, to aggregate resources related to pop-culture (Kiryakos, Sugimoto, Nagamori, & Mihara, 2017) we can utilize similar methods. The Agency for Cultural Affairs' Media Arts Database[50] in Japan is one such database which aggregates manga, animation and game information resources into one platform. Hence, the author believes that the CHDE model also can be utilized and extended to aggregate resources in domains other than cultural heritage.

In addition, the *Instantiation* is the bridge between the real object or concept in the *Physical Space* and the *Digital Surrogate* in the *Digital Space*. This idea of *Instantiation* based information organization can be extended to other domains and applications that deals with "Event" like activities which exist temporarily such as disaster archival information etc. The models proposed in this thesis are developed to aggregate South and Southeast Asian CHI, yet, these models can be identified as region-neutral models, as they can be used to aggregated CHI in other countries as well.

The models proposed in this research enrich the existing digital collections and enable creating digital archives of cultural heritage with more contextual information. When aggregating multiple resources belonging to multiple schemas, we need a common model

---

[50]https://mediaarts-db.bunka.go.jp/

based on defined principles. Here the One-to-One Principle of Metadata acts as the defined principle to construct these models while allowing consistency in the models.

The thesis presented an abstract level of CHI aggregation through metadata models. An implementation of a system based on the above metadata models will be a future direction of this study. The abstract level model becomes operational to the real world if it is implemented as a usable system. As the author specified a few limitations of this study such as aggregation of non-institutional CHI, they should be investigated more in depth. In addition, issues such as rights description, provenance description and long-term maintenance of aggregated metadata are left as future works of this research.

Acknowledgements

Undertaking this Ph.D. has been a truly life-changing experience for me, and it would not have been possible to do without the support and guidance that I received from many people and organizations.

First and foremost, I would like to express my sincere gratitude to my main advisor, Professor Shigeo Sugimoto. He has been the shadow behind me throughout my research life at the University of Tsukuba. From the very first definition of METADATA to lifetime experiences, what I have learnt from Prof. Sugimoto is exceptional. The opportunities he provided me to participate in top LIS conferences around the globe and introducing me to professionals, scholars, and peers in the field is unforgettable. I am indebted to him for letting me know those professional networks and I am sure that it will be a great opportunity for my future profession as well. Further, I would like to thank him for his continuous support throughout my masters and Ph.D. researches and for all the guidance, ideas, advices, comments that he rendered me without any hesitation. Professor Sugimoto's dedication towards teaching, patience, and kindness means a lot to me. Thank you for being a wonderful professor and helping me with every step to accomplish my Ph.D. dream in Japan. 先生、どうもありがとうございます.

I would also like to show gratitude to my supervisor Prof. Atsuyuki Morishima for his insightful comments and suggestions. His comments and discussions always helped me to clarify certain aspects of this research. Though his arguments and questions during the lab meetings made me nervous at the beginning, later on I realized the true value and advantage of those activities. I am thankful to him for dedicating his time for me and for all the help and encouragement that he has given me throughout my studies at the University of Tsukuba.

Thank you very much, Prof. Tetsuo Sakaguchi, Prof. Yasuaki Tsuji, Prof. Eiji Mizushima (previous supervisor) and the members of the research committee Prof. Tomoo Inoue, Prof. Masao Takaku, Shoichiro Hara (Kyoto University) for all their valuable comments, insights and involvement in this research process.

Thank you very much Prof. Mitsuharu Nagamori, Prof. Tetsuya Mihara at the Metadata Lab for the advices and kindness in numerous ways. I very much appreciate their support in this research. Their advices always made my life comfortable and thank you very much for everything.

References

Amin, R., Baker, O.F., Deraman, A., Yatim, N.F.M. (2012). Transforming model to meta model for knowledge repository of Malay intangible culture heritage of Malaysia. *Int. J. Electr. Comput. Eng.* 2(2), 231–238. doi: 10.11591/ijece.v2i2.205

Berners-Lee, T. (2006). Linked data. Design issues. W3C. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data- the story so far. Linked Data. *International Journal on Semantic Web and Information Systems (IJSWIS).* Retrieved from http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world.* Cambridge, MA: MIT Press.

Cameron, F., & Kenderdine, S. (2007). *Theorizing digital cultural heritage: A critical discourse.* MIT Press, Cambridge.

Carboni, N., de Luca, L. (2016). Towards a conceptual foundation for documenting tangible and intangible elements of a cultural object. *Digital Applications in Archaeology and Cultural Heritage,* 3(4), 108–116. doi:10.1016/j.daach.2016.11.001

Chen, Y.N., Ke, H.R. (2013). FRBRoo-based approach to heterogeneous metadata integration. *Journal of Documentation.* 69(5), 623–637.

Chryssoula, B., Doerr, M., Le Bœuf, P., & Riva, P. (eds.). (2015). Definition of FRBRoo a conceptual model for bibliographic information in object-oriented formalism. International Working Group on FRBR and CIDOC CRM Harmonisation, version 2.4.

Constantopoulos, P., et al. (2009). DCC & U: An extended digital curation lifecycle model. *International Journal of Digital Curation*, 4(1).

Coyle, K., Baker, T. (2009). Guidelines for Dublin core application profiles. Dublin Core Metadata Initiative. Retrieved from http://dublincore.org/documents/profile-guidelines/

Doerr, M., Fundulaki, I., and Christophidis, V. (1997). The specialist seeks expert views: Managing digital folders in the AQUARELLE project. *Museums and the Web: An International Conference*, Los Angeles, CA, March 16 - 19, 1997. Retrieved from http://www.museumsandtheweb.com/mw97/speak/doerr.html

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine, 8*(4).

Europeana data model primer. (2013). Isaac, A. (ed.). Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requ irements/EDM_Documentation/EDM_Primer_130714.pdf

Freire, N., Robson, G., Howard, J. B., Manguinhas, H., & Isaac, A. (2017). Metadata aggregation: assessing the application of IIIF and Sitemaps within cultural heritage. *International Conference on Theory and Practice of Digital Libraries*, pp. 220-232. Springer, Cham.

Freire, N., Robson, G., Isaac, A., Manguinhas, H., & Howard, J. B. (2017). Web technologies: a survey of their applicability to metadata aggregation in cultural heritage. *ELPUB*, pp. 235-244.

Geisler, G., Giersch, S., McArthur, D., and McCelland, M. (2002). Creating virtual collections in digital libraries: benefits and implementation issues. *Joint Conference on Digital Libraries 2002*, Portland, OR, pp. 210-218.

Giannoulakis, S., Tsapatsoulis, N., & Grammalidis, N. (2018). Metadata for intangible cultural heritage: the case of folk dances. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2018)*, pp. 634-645. doi: 10.5220/0006760906340645

Gill, T. (2004). Building semantic bridges between museums, libraries and archives: The CIDOC Conceptual Reference Model. *First Monday*, 9 (5). doi: http://dx.doi.org/10.5210/fm.v9i5.1145

Gilliland, A. J. (2008). Setting the stage. *Introduction to metadata.* M. Baca (Ed.). Online Edition, Version 2.0. Getty Publications. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/settin g.html

Gogina, I. (2016). The world digital library: Metadata crosswalks. *INFO 281-11. Seminar in Contemporary Issues - Metadata.* San José State University.

Greenberg, J. (2017). Big metadata, smart metadata, and metadata capital: toward greater synergy between data science and metadata. *Journal of Data and Information Science*, 2(3), 19-36.

Guthro, C. (2013). Digital Public Library of America. Maine Policy Review. *Libraries and Information*, 22(1), 126-129. Retrieved from http://web.simmons.edu/~weigle/DIGCOLL/Guthro_DPLA.pdf

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1-136.

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1).

Hillmann, D. I., Dushay, N., & Phipps, J. (2004). Improving metadata quality: augmentation and recombination. *International Conference on Dublin Core and Metadata Applications.* Shanghai: DCMI.

Hjørland, B. (2013). Facet analysis: The logical approach to knowledge organization. *Information processing & management*, 49(2), 545-557.

Hu, J., Lv, Y., Zhang, M. (2014). The ontology design of intangible cultural heritage based on CIDOC CRM. *Int. J. U-and E-Serv*. 7(1), 261–274.

Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the semantic web. *Synth. Lect. Seman. Web: Theor. Technol.* 2(1), 1–159.

Hyvönen, E. (2016). *Cultural heritage linked data on the semantic web: Three case studies using the Sampo model.* VIII Encounter of documentation centres of contemporary art: open linked data and integral management of information in cultural centres. Artium, Vitoria-Gasteiz, Spain. Retrieved from https://seco.cs.aalto.fi/publications/submitted/hyvonen-vitoria-2017.pdf

Isaac, A., Manguinhas, H., Stiller, J., & Charles, V. (2015). *Report on enrichment and evaluation*. The Hague, Netherlands: Europeana Task Force on Enrichment and Evaluation. Retrieved from http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf

Kiryakos, S., Sugimoto, S., Nagamori, M., & Mihara, T. (2017). Aggregating metadata from heterogeneous pop culture resources on the web. *International Conference on Dublin Core and Metadata Applications,* pp. 65-74.

Lagoze, C. (1996). The Warwick framework: A container architecture for diverse sets of metadata. *D-Lib Magazine*. Retrieved from http://www.dlib.org/dlib/july96/lagoze/07lagoze.html

Lagoze, C. (2001). Keeping Dublin Core Simple Cross-Domain Discovery or Resource Description? *D-Lib Magazine*, 7(1).

Lanzi, E. (1998). *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information*. Getty Publications, Los Angeles.

Le Boeuf, P., Doerr, M., Ore, C.E., Stead, S. (eds.). (2015). Definition of the CIDOC conceptual reference model. ICOM/CIDOC Documentation Standards Group and CIDOC CRM Special Interest Group, version 6.2.1.

Lourdi, I., & Nikolaidou, M. (2009). Guidelines for developing digital cultural collections. *Handbook of research on digital libraries: design, development, and impact*. IGI Global.

Low, B. (2006). *Machine Services for Metadata Discovery and Aggregation - metadata+*. UK: JISC Development Programmes.

Low, J. T. and Doerr, M. (2010). A Postcard is Not a Building - Why we Need Museum Information Curators. *Proc. of the CIDOC 2010 Conference*.

METS. (2010). *Metadata encoding and transmission standard: primer and reference manual*. USA: Library of Congress.

Miller, S. J. (2010). The one-to-one principle: challenges in current practice. *Proceedings on the International Conference on Dublin Core and Metadata Applications*, pp. 150-164.

Miller, S. J. (2011). *Metadata for digital collections: a how-to-do-it manual*. New York, NY: Neal-Schuman Publishers.

NISO. (2004). *Understanding metadata*. National Information Standards Organization.

NISO. (2007). *A framework of guidance for building good digital collections*. Baltimore, MD: National Information Standards Organization (NISO).

OAI-ORE. (2008). C. Lagoze, , H. Van de Sompel, P. Johnston, M. Nelso, S. Robert , & W. Simeon (Producers). *ORE user guide-Primer*. Retrieved from https://www.openarchives.org/ore/1.0/primer

Orgel, T., Höffernig, M., Bailer, W., & Russegger, S. (2015). A metadata model and mapping approach for facilitating access to heterogeneous cultural heritage assets. *International Journal on Digital Libraries,* 15(2-4), 189-207.

Papatheodorou, C., Dallas, C., Ertmann-Christiansen, C., Fernie, K., Gavrilis, D., Masci, M. E., & Angelis, S. (2011). A new architecture and approach to asset representation for europeana aggregation: The CARARE way. *In Research Conference on Metadata and Semantic Research*, pp. 412-423. Springer, Berlin, Heidelberg.

Peroni, S., Tomasi, F., & Vitali, F. (2012). Reflecting on the Europeana data model. *In Italian Research Conference on Digital Libraries,* pp. 228-240. Springer, Berlin.

Peroni, S., Tomasi, F., & Vitali, F. (2013). The aggregation of heterogeneous metadata in web-based cultural heritage collections: a case study. *International Journal of Web Engineering and Technology*. 8, 8(4), 412-432.

Ruthven, I., & Chowdhury, G. G. (Eds.). (2015). *Cultural heritage information: Access and management*, vol. , Facet Publishing.

Schulte-Albert, H. G. (1974). Cyprian Kinner and the idea of a faceted classification. *Libri*, 24(4), 324–337.

Signore, O. (2006). The Semantic Web and cultural heritage: ontologies and technologies help in accessing museum information. *Information Technology for the Virtual Museum*, Sønderborg, Denmark.

Smiraglia, R. P. (2005). Content metadata–An analysis of Etruscan artifacts in a museum of archeology. *Cataloging & classification quarterly*, 40(3-4), 135-151.

Sugimoto, S. (2014). Digital archives and metadata as critical infrastructure to keep community memory safe for the future lessons from Japanese activities, *Archives and Manuscripts*, 42:1, 61-72, doi: 10.1080/01576895.2014.893833

Sugimoto, S., Kiryakos, S., Wijesundara, C., Monika, W., Mihara, T., & Nagamori, M. (2018). Metadata models for organizing digital archives on the web: metadata-centric projects at Tsukuba and lessons learned. *In International Conference on Dublin Core and Metadata Applications*, pp. 95-105.

Swan, A., & Awre, C. (2006). Linking UK repositories: Technical and organisational models to support user-oriented services across institutional and other digital repositories. *JISC Study Report*. Retrieved from http://www.webarchive.org.uk/wayback/archive/20140614201504/http://www.jisc.ac.uk/publications/reports/2006/linkingukrepositoriesreport.aspx

Tan, G., Hao, T., Zhong, Z.: A knowledge modeling framework for intangible cultural heritage based on ontology. (2009). *In: Knowledge Acquisition and Modeling*, KAM 2009, vol. 1, pp. 304–307. IEEE. doi:10.1109/KAM.2009.17

Tomasi, F., Ciotti, F., Daquino, M., & Lana, M. (2015). Using Ontologies as a Faceted Browsing for Heterogeneous Cultural Heritage Collections. *In IT@LIA@AI* IA.*

Tunkelang, Daniel (2009). *Faceted Search*. Morgan & Claypool.

Urban, R. (2014). The 1: 1 principle in the age of linked data. Proc. *Int'l Conf. on Dublin Core and Metadata Applications.*

Van Hooland, S., & Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*. Facet publishing.

W3C. (2014). *RDF 1.1 Primer*. Retrieved from. https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). *Dublin core metadata for resource discovery (No. RFC 2413)*. Dublin Core. RFC Editor. USA. Retrieved from https://www.rfc-editor.org/rfc/rfc2413.txt

Weibel , S., & Miller, E. (2000). An introduction to dublin core. Retrieved from https://www.xml.com/pub/2000/10/25/dublincore/index.html

Wickett, K. M., Isaac, A., Doerr, M., Fenlon, K., Meghini, C., & Palmer, C. L. (2014). Representing cultural collections in digital aggregation and exchange environments. *D-Lib Magazine*, 20 (5/6).

Wickett, K. M., Isaac, A., Fenlon, K., Doerr, M., Meghini, C., Palmer, C. L., & Jett, J. (2013). *Modeling cultural collections for digital aggregation and exchange environments. (CIRSS Technical Report 201310-1).* Champaign, IL: Center for Informatics Research in Science and Scholarship, University of Illinois at Urbana-Champaign.

Wijesundara, C. (2019). Cultural heritage information aggregation in networked information environments. TCDL. *Bulletin of IEEE Technical Committee on Digital Libraries*. vol. 15, 2.

Wijesundara, C., Monika W., & Sugimoto, S. (2017). A metadata model to organize cultural heritage resources in heterogeneous information environments. Choemprayong S., Crestani F., Cunningham S. (eds). ICADL 2017. *Lecture Notes in Computer Science (LNCS)*, 10647, Heidelberg, Springer. pp. 81-94. doi: http://doi.org/10.1007/978-3-319-70232-2_7

Wijesundara, C., Sugimoto, S. (2018). Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space. *LIBRES*. vol. 28, 2. pp. 58-80.

Wijesundara, C., Sugimoto, S. & Narayan, B. (2015). Documenting spatial and temporal information for heritage preservation: A case study of Sri Lanka, *Proceedings from the Document Academy:* vol. 2: 1, Article 5. doi: http://doi.org/10.35492/docam/2/1/5. Retrieved from https://ideaexchange.uakron.edu/docam/vol2/iss1/5.

Wijesundara, C., Sugimoto, S., Narayan, B., & Tuamsuk, K. (2016). 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.

Woodley, M. (2009). DCMI glossary. Retrieved from http://dublincore.org/documents/usageguide/glossary.shtml

Woodley, M. S. (2016). Metadata matters: connecting people and information. Baca, M. (Ed.): *Introduction to metadata* (3rd ed.). Los Angeles, LA: Getty Publications. Retrieved from https://www.getty.edu/publications/intrometadata/metadata-matters/

Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El profesional de la información*, v. 28, n. 1, e280103.

doi: https://doi.org/10.3145/epi.2019.ene.03

Zeng, M. L., & Qin, J. (2016). *Metadata* (2nd edition). ALA. Chicago, IL: Neal-Schuman Publishers.

List of Publications (by the Author)

1.   Peer-reviewed academic journal papers

Wijesundara, C., Sugimoto, S. (2018). Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space. *LIBRES*. vol. 28, 2. pp. 58-80.

2.   Peer-reviewed proceeding papers at international conferences

Wijesundara, C., Monika W., & Sugimoto, S. (2017). A metadata model to organize cultural heritage resources in heterogeneous information environments. Choemprayong S., Crestani F., Cunningham S. (eds). ICADL 2017. *Lecture Notes in Computer Science (LNCS)*, 10647, Heidelberg, Springer. pp. 81-94. doi: http://doi.org/10.1007/978-3-319-70232-2_7

Wijesundara, C., Sugimoto, S., Narayan, B., & Tuamsuk, K. (2016). 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.

Wijesundara, C., Sugimoto, S. & Narayan, B. (2015). Documenting spatial and temporal information for heritage preservation: A case study of Sri Lanka, *Proceedings from the Document Academy:* vol. 2: 1, Article 5. doi: http://doi.org/10.35492/docam/2/1/5. Retrieved from https://ideaexchange.uakron.edu/docam/vol2/iss1/5

Monika, W., **Wijesundara, C.**, & Sugimoto, S. (2017). Modelling digital archives of intangible cultural heritage based on 1-to-1 principle of metadata. *A-LIEP 2017*, pp. 137-148.

Sugimoto, S., Kiryakos, S., **Wijesundara, C.**, Monika, W., Mihara, T., & Nagamori, M. (2018). Metadata models for organizing digital archives on the web: Metadata-centric projects at Tsukuba and lessons learned. *DCMI International Conference on Dublin Core and Metadata Applications*, pp. 95-105.

Appendices

Appendix 1. Crosswalk between Heritage classes with CIDOC-CRM and FRBRoo

	Category	Main Classes	AFS/ AAT Related Terms	CIDOC-CRM Classes	FRBRoo Classes
1	TCH	Movable cultural heritage	cultural artifacts (AFS) arfifacts (AFS/AAT) specimens (AAT) natural objects (AAT)	E22 Man-Made Object	
				E18 Physical Thing	F7 Object (=)
2	TCH	Immovable cultural heritage	archaeological sites (AFS/AAT)	E24 Physical Man-Made Thing	F4 Manifestation Singleton (*)
			monuments (AFS/AAT)	E25 Man-Made Feature	F53 Material Copy (*) (≠)
				E27 Site	
3	ICH	Oral traditions	oral traditions (AFS/AAT)	E28 Conceptual Object	F6 Concept (=)
4	ICH	Performing arts	performing arts (AFS/AAT)	E7 Activity	F31 Performance (*)
				E29 Design or Procedure	F25 Performance Plan (=)
5	ICH	Social practices, rituals, festive events	social ethics (AAT) civic rituals (AFS) rituals (events) (AAT) festivals (AFS/AAT)	E7 Activity	F31 Performance (*)
				E28 Conceptual Object	F6 Concept (=)
				E29 Design or Procedure	F25 Performance Plan (=)
6	ICH	Knowledge and skills to produce traditional crafts	traditional knowledge (AFS) knowledge (AAT) crafts (AFS) ability (AFS)	E28 Conceptual Object	F6 Concept (=)
				E29 Design or Procedure	F25 Performance Plan (=)
				E77 Persistent Item	
7	ICH	Knowledge and practices concerning nature and the universe	traditional ecological knowledge (AFS)	E28 Conceptual Object	F6 Concept (=)
			knowledge (AAT)	E56 Language	
			language (AAT/AFS)	E36 Visual Item	
			visual art (AAT)	E77 Persistent Item	

Key :   **TCH**: Tangible Cultural Heritage   **ICH**: Intangible Cultural Heritage

AAT: Art & Architecture Thesaurus (Getty)

AFS: American Folklore Society Ethnographic Thesaurus

(*) Subclass   **(=)** Equal Class   **(≠)** Not Equal

Appendix 2. Mapping Between Description Module and Properties Extracted from the Web Services

Object Categories	Description Modules	British Museum	Metropolitan Museum of Art	Asian Art Museum	Rijksmuseum	Europeana	Technical University of Cluj-Napoca	Wikipedia	UNESCO World Heritage Centre	Asia/Pacific Database for ICH	UNESCO- ICH
Original Object	Content Description Module	Title, Description, Subjects, Associated names, COMPASS Title	Title, Classification, Description, Culture	Title, Object Name, Object type, DESCRIPTI-ON	Title, Object type, Description, Subject	Title, Type, Subject, Description, Language,	Title, Description, Language	Title, Sub title, Description, Subjects	Title, Description	Title, Sub title, Essential elements, Detailed explanation	Title, Description, Concepts
	Agent Module		Artist		Artist						
	Location Module	Production place, Findspot, Location	Geographic Location	Place of Origin	Place	Location, Map	Geographic coverage	Coordinates, Countries	Map, Country, Location, Coordinates	Country, Area where performed	Country
	Timeline Module	Culture, period, Date	Period, Date, Timeline, Date, Era	Date	Dating	Date	Historical Date, Time period				
	Technical Description Module	Materials, Technique, Dimensions, Object Type	Medium, Dimensions, Material, Object Type	Materials, Dimensions	Material, Technique, Measurements	Medium, Size	Dimensions				
Digital Surrogate	Content Description Module			Images			ID, Language	Image title, More details	Images, Description	Audio documentation, Visual documentation	Images, Video
	Agent Module					Provider		Author	Author		Provider

	Location Module										
	Timeline Module								Date and time of data			Year
	Technical Description Module					format, file size	Scan Type, Scan Format, Digitization Method	File (image format), Metadata related to media (Camera manufacturer, Camera model, Created (image file) etc.	Medium size, Original size		
Administr-ative	**Content Description Module**	Museum number, Acquisition notes, Registration number	Accession No	Object Number, On Display	Object number	Identifier,			Criteria, Ref	Reasons for selection	Nomination file No, Decision, Periodic reporting
	Agent Module	Acquisition name	Credit Line Signature	Credit Line		Publisher, Institution, Provider	Data Provider			Data provider, Institution/ organization involved in preservation and promotion	Data providing institute
	Location Module	Department	Gallery location, Department	Department, Collection		Providing Country					
	Timeline Module	Acquisition date				First Published in Europeana, Last Updated in Europeana			Date of Inscription		
	Provenance Module	Exhibition history	Exhibition History								

	Rights Module	Copyright,	Public Domain (Copyright)		Credit line Copyright	Rights	Rights	Copyright	Copyright,		Copyright (image/ video)
External Resources	**External Resource Module**	Use image, Request new photograph, Open data for this object with SPARQL endpoint, Image view	Additional Images, Link, Related objects	Images, Image Request, Link	URL, Image URL	References And Relations, Similar items	Sources	Sources, External Links, More details (image)	Reusing this photo, Links, Map, Gallery, Permanent URL, Language (tabs)		Links
	Bibliographic Module	Bibliographic	Met Publications, Reference		Documentation			Bibliography, Sources	Documents	Publication and textual documentation	

Appendix 3. Alignment of the Object Categories, Description Modules with Properties from Web Services and Assigning Facets to the Same

Object Categories (Set of Description Sets)	Description Modules (Description Sets)	Related Attributes (Descriptions)	Facets
Original Object	**Content Description Module**	Title	Details
		Sub title	
		COMPASS Title	
		Associated names	
		Description	
		Detailed explanation	
		Essential elements	
		Subject/s	
		Classification	
		Type	
		Object Name	
		Object type	
		Concepts	
		Language	Language
		Culture	
	Agent Module	Artist	Person
	Location Module	Production place	Location
		Find-spot	
		Place	
		Location	
		Place of Origin	
		Geographic Location	
		Country/ Countries	
		Area where performed	
		Geographic coverage	
		Coordinates	
		Map	
	Timeline Module	Culture	Period
		Period	

		Date	
		Era	
		Time period	
		Timeline	
		Dating	Date
		Historical Date	
	Technical Description Module	Materials	Material
		Medium	
		Dimensions	Dimension
		Measurements	
		Size	
		Object type	Format/ Type
		Technique	Tech Details
Digital Surrogate	**Content Description Module**	Image Title	Details
		Image/s	
		Description	
		Visual Documentation	
		Audio documentation	
		Video	
		More Details	
		Language	Language
		ID	Object No.
	Agent Module	Author	Person
		Provider	Group
	Location Module		Location
	Timeline Module	Date and time of data	Date
		Year	
	Technical Description Module	Original size	Dimension
		file size	
		Medium Size	
		Scan Type	Format/ Type
		Scan Format	
		File (image format...)	
		format	
		Media type	

		Digitization Method	
		Metadata related to media (Camera manufacturer, Camera model, Created (image file) etc.	Tech Details
		File information	
Administrative	**Content Description Module**	Acquisition notes	Details
		On Display	
		Reasons for selection	
		Criteria	
		Decision	
		Periodic reporting	
		Nomination File	
		Object Number	Object No.
		Registration number	
		Accession Number	
		Museum number	
		Identifier	
	Agent Module	Acquisition name	Person
		Credit Line	
		Signature	
		Publisher	Group
		Data Provider/ Provider	
		Institution/organization involved in preservation and promotion	
		Data providing institute	
	Location Module	Department	Location
		Gallery location	
		Providing Country	
		Collection	
	Timeline Module	Acquisition date	Date
		First Published in Europeana	
		Last Updated in Europeana	
		Date of Inscription	
	Provenance Module	Exhibition history	Provenance

	Rights Module	Copyright	Rights
		Public Domain (Copyright)	
		Rights	
		Copyright (image/ video)	
External Resources	**External Resource Module**	Use image	Related Links
		Request new photography	
		Open data for this object with SPARQL endpoint	
		Additional Images	
		Link/s	
		Related objects	
		Images	
		Image Request	
		URL	
		Image URL	
		References And Relations	
		Similar items	
		Sources	
		External Links	
		More Details	
		Gallery	
		Reusing this photo	
		Map	
		Permanent URL	
		Language (tabs)	
	Bibliographic Module	Bibliography	Bibliographic
		MetPublications	
		Reference	
		Sources	
		Publication and textual documentation	
		Documents/ Documentation	