Metadata Models for Organizing Digital Archives on the Web: Metadata-Centric Projects at Tsukuba and Lessons Learned

Shigeo Sugimoto
University of Tsukuba, Japan
sugimoto@slis.tsukuba.ac.jp

Senan Kiryakos
University of Tsukuba, Japan
senank@gmail.com

Chiranthi Wijesundara
University of Tsukuba, Japan
chiranthis@gmail.com

Winda Monika
University of Tsukuba, Japan
windabi.wm@gmail.com

Tetsuya Mihara
University of Tsukuba, Japan
mihara@slis.tsukuba.ac.jp

Mitsuharu Nagamori
University of Tsukuba, Japan
nagamori@slis.tsukuba.ac.jp

Abstract

There exist many digital collections of cultural and historical resources, referred to as digital archives in this paper. Domains of digital archives are expanding from traditional cultural heritage objects to new areas such as popular culture and intangible entities. Though it is known that metadata models and authority records, such as subject vocabularies, are essential in building digital archives, they are not yet well established in these new domains. Another crucial issue is semantic linking among resources within a digital archive and across digital archives. Metadata aggregation is an essential aspect for the resource linking. This paper overviews three metadata-centric on-going research projects by the authors and discusses some lessons learned from them. The subject domains of these research projects are disaster records of the 2011 Great East Japan Earthquake, Japanese popular culture such as Manga, Anime and Games, and cultural heritage resources in South and Southeast Asia. The main goal of this paper is not to report on these projects as completed research, but to discuss issues of metadata models and aggregation which are important in organizing digital archives in the web-based information environment.

Keywords: digital archives; digital collections; digital curation; digital libraries; fan-created websites; Japanese popular culture (pop-culture); manga; linked open data; memory institutions; metadata aggregation; tangible and intangible cultural heritage

1. Introduction

Since the early 1990s, many digital collections of cultural, historical, and scholarly resources have been developed by many memory institutions, such as libraries, museums, and archives. Those collections were built to improve user accessibility to important cultural resources via the Internet while simultaneously preserving those resources. It is widely known that metadata is key to building those digital collections for both preservation and access, as metadata is needed to handle digital resources as well as cultural heritage objects in various aspects, i.e., search, select, organize, access, and preserve. Meanwhile, major activities to develop metadata standards oriented to the web environment also started in 1990s, e.g., Dublin Core Metadata Initiative (DCMI, http://dublincore.org/), Open Archival Initiative (OAI, http://www.openarchives.org/), and so on.

Metadata aggregation from multiple sources has been widely recognized as an important technology to create and organize a merged digital collection on the web. OAI developed a metadata harvesting protocol named OAI-PMH which is used to build value-added services. For example, National Science Digital Library (NSDL, https://nsdl.oercommons.org) collected and aggregated metadata from many sources. Europeana and Digital Public Library of America (DPLA), well known as large portals for digital collections of cultural resources in Europe and the United States, respectively, harvest metadata from their participating institutions. In Japan, the

https://doi.org/10.23106/dcmi.952139012

National Diet Library (NDL) has built a portal which collects metadata using OAI-PMH and provides unified access to the digital collections of Great East Japan Earthquake developed by regional public sectors, NPOs, universities, news and information companies, etc. Thus, metadata harvesting is a commonly used technology to add values to digital collections of cultural resources. In the current web environment, Linked Open Data (LOD) technologies provide information environments to develop more sophisticated services using various LOD datasets in addition to those digital collections.

Digital collections and services are sometimes referred to by different names, e.g., digital library, digital museum, digital archive and digital gallery. We use the term *digital archive* to refer to those digital collections and services for the rest of this paper. This is because the term *archive* has the meaning to collect important resources to provide access over time for future use. We call digital archives created by memory institutions using their holdings *institutional digital archives*.

Many digital archives developed by memory institutions provide digital copies of cultural heritage objects held by the institutions. They are built using the catalog data of the institution as the base metadata for the digital archives. In most cases of existing institutional digital archives, those digital copies are created by digitizing the original cultural heritage objects using devices such as digital cameras, scanners, analog-to-digital converters, etc. Even in cases where institutions cannot allow open access to the digitized copies of the objects, the metadata of the cultural objects plays important roles for users in finding and accessing the digital resources and the original cultural objects.

In general, while institutional metadata developed by memory institutions to describe their holdings is highly standardized, it often provides very limited information. A typical problem is that they do not provide contextual information of cultural heritage objects which would help users understand their values. On the other hand, websites such as Wikipedia and those created by fans and specialists provide in-depth descriptions about the objects, including contextual information of the objects, but they are less standardized. Aggregation of different types of cultural heritage object descriptions, which we call metadata aggregation, is important because we can obtain better description of cultural heritage objects by combining institutional metadata and other websites. Metadata in domains such as pop-culture and intangible cultural heritage are not well standardized like those of traditional cultural heritage objects. Metadata aggregation is very important in these domains. This viewpoint is shared among the three projects shown in this paper.

This paper describes three research projects that have been undertaken to address these issues. While not intended as a report of completed research projects, this paper seeks:

- to discuss challenges in metadata aggregation in domains which are not well covered by conventional institutional digital archives in the cultural domain,
- to discuss the development of underlying models of metadata and some lessons learned from the research in the three domains – disaster records, Japanese pop-culture and intangible cultural heritage, and
- to re-think a few general metadata models such as One-to-One Principle (Miller, 2010; Woody, Clement & Winn, 2005), FRBR (IFLA, 2009) and Metadata Application Profiles (Nilsson, Baker & Johnston, 2008).

The rest of this paper is organized as follows. In Section 2, we discuss metadata aggregation as a key issue to enhance usability of the digital archives. Section 3 shows the three projects followed by discussions on lessons learned from the projects and conclusion in Sections 4 and 5, respectively.

https://doi.org/10.23106/dcmi.952139012

2. Backgrounds

2.1. Metadata Aggregation for Digital Archives of Cultural and Historical Resources – Basic Concepts and Issues

Metadata aggregation is broadly used to collect metadata from multiple repositories and organize a virtually unified repository as mentioned earlier. OAI's metadata harvesting protocol (OAI-PMH) is a widely used protocol for collecting metadata.

Europeana Data Model (EDM) (Isaac, 2013) defines entities expressed in their metadata such as digital images, original cultural heritage objects, and relationships among those entities. EDM defines aggregation both within a single digital archive and across digital archives.

Ministry of Internal Affairs and Communication (MIC) of the Japanese government supported regional public sectors to digitally archive the records of the Great East Japan Earthquake occurred in March 2011. NDL worked as a national institution to help organize the regional digital archives and built a national portal for the disaster archives, which is named Hinagiku (http://kn.ndl.go.jp). Hinagiku collects metadata from participating archives. A crucial issue of Hinagiku is the quality of the metadata. The quality of the metadata of the participating archives are, in general, not high mainly because of the limitation of financial, human and time resources; the majority of the resources are photographs and the metadata was created by third parties who were not metadata specialists and within a very limited time. Another aspect is lack of standards to organize metadata, e.g., granularity of the archived resources.

Agency for Cultural Affairs (ACA, Bunka-cho in Japanese) has been hosting the Japan Media Arts Festival for since 1997, which covers four resource types: Art, Entertainment, Animation and Manga. The Media Arts Database (MADB, https://mediaarts-db.bunka.go.jp) hosted by ACA collects metadata in the four resource types Manga, Animation, Game and Media Art. (notes: Manga is a Japanese term meaning Japanese comics or graphic novels. Animation is written as Anime in this paper.)

Anime and Games are typically part of large franchises, e.g. *Dragon Ball*, *Gundam*, *Pokémon*, and so on. Metadata aggregation for members of these franchises from different mediums is a crucial requirement to enhance the usability of the MADB. Linking contents of different media types by shared franchise can improve features such as search and retrieval. To achieve this, however, we need to model the franchise as an entity like the Work and Item entities of FRBR, and relationships among franchise entities and other entities described by MADB.

2.2. Data Model Issues for Metadata Aggregation

Data models have crucial roles for defining metadata schemas interoperable with other schemas. EDM and OAI-ORE define structures of metadata collected and aggregated from multiple metadata sources. Metadata mapping is a key issue in the aggregation process in the case that metadata to be aggregated are created on different schemas. Metadata mapping is often done only based on properties. However, property-level mapping has risks of losing context of properties given in the schema in which the properties are included, such as mandatory levels and value types. Thus, we need to use contextual information for metadata mapping.

FRBR provides us with generalized object types for books and other published materials. MADB in part adopts FRBR to define the object types for their databases. In our project on Japanese pop-culture, we defined our data models based on those object types of MADB. In particular for Manga, we defined a three-layer model composed of *Superwork*, *Work* and *Volume*.

One-to-One principle of metadata is well-known as a basic model which says that relationships between metadata and its objective resource should be One-to-One. A typical One-to-One metadata example is a catalog record of a physical object at a memory institution. Non-One-to-One metadata is not uncommon; typical digital archive metadata has descriptions about the original cultural heritage objects, descriptions about a digital resource(s) created from the original object and some other related descriptions in the digital archives hosted by the memory institutions. This model which has an original cultural heritage object fits well to tangible cultural heritage objects.

https://doi.org/10.23106/dcmi.952139012

However, it does not work well for those cultural heritage which cannot be collected by memory institutions as a physical item, e.g., intangible cultural heritage (e.g., performance and craftsmanship), very large objects (e.g., monuments and large statues) and any objects which have short life-time (e.g., installation arts, ice arts and fireworks). A typical solution for this kind of cultural heritage is to use photographs or videos as a surrogate of the heritage object. In this case, we would need description not only about an original cultural heritage object but also about its surrogate. Identification of the cultural heritage objects as an objective of metadata description is crucial if the relationship between the heritage objects and their metadata is non-One-to-One.

In the case of the disaster archives of Great East Japan Earthquake, their metadata model is a simple One-to-One, i.e., one metadata record for one resource such as a document, photograph, video, or oral record. This simple model works well for metadata creation. However, it often brings us too many hits for a simple search query. The records collected were primarily about an event or a thing related to the disaster, e.g., photographs of damaged local harbors, photographs taken at local festivals, news documents issued by local governments, and so forth. So, it is a natural demand to make groups of the items based on the events, but metadata of a record item does not include description about the contexts of the item which is meaningful for end-users and useful for the grouping of those items. However, identifying those events was not an easy task for the third-party catalogers and metadata description scheme for those events was not well defined.

2.3. Building Digital Archives of Non- Conventional Domains: Research Questions

The access environments for general users to obtain cultural information on the web has been increasingly developed since 1990s, i.e., directory services, full text search engines, blogs, and SNS. Another significant factor of cultural information on the web is the development of cultural resources by non-memory institutions and by crowds. The value of institutional digital archives and their portals are well-recognized. On the other hand, quite a lot of cultural information resources are available in those non-institutional resources such as Wikipedia and those sites created by domain specialists and fans. Semantic Web and Linked Open Data technologies obviously have crucial roles for linking the resources across the boundaries between institutional and non-institutional resources.

There exist digital archives of various types of cultural objects and the boundaries of the archives are becoming fuzzier; from traditional tangible cultural heritage objects to intangible cultural heritage objects such as dance, festivals, cuisines, and craftsmanship; from printed books to electronic and web-based books; from traditional arts to media arts such as videogames and computer arts; from a movable objects to immovable objects such as large statues and buildingss; from simple photographic images to 3D and virtual reality images, and motion graphics.

These digital archives create metadata designed in accordance with their archived resources. Conventional institutional digital archives use metadata standards used by memory institutions as the basis for the digital archives. However, the metadata in the new domains need some new features which are not covered by metadata of those conventional digital archives. For example, in the case of Manga, metadata schemas specialized for describing Manga are not well developed.

Web resources such as Wikipedia, fan-created sites, and exhibition pages by museums often provide richer information for users than institutional metadata which is primarily bibliographic descriptions of their holdings. On the other hand, the institutional metadata provides authoritative information. Therefore, metadata aggregation across institutional digital archives and websites is a crucial issue to build better environment for users on the web to find and access digital resources of cultural domains. The challenge is to bridge the gap between the metadata of these different types of resources, i.e., well-structured institutional metadata and non- or semi-structured descriptions in webpages.

3. Research Projects on Metadata for Digital Archives at Tsukuba

This section briefly describes three research projects at the authors laboratory from which we have learned issues on metadata to improve usability of digital archives.

https://doi.org/10.23106/dcmi.952139012

3.1. Enhancing Usability of Digital Archives of Great East Japan Earthquake by Metadata Aggregation within and across Archives

(1) Project Background

As mentioned earlier in this paper, there are many digital archives of the Great East Japan Earthquake on March 11, 2011. NDL is running a portal named Hinagiku for those archives. Each of the archives collects various types of resources. Photographs taken by digital cameras are the largest portion of the archived resources. This feature is a significant difference from the archive for the earthquake in January 1995 in Kobe/Awaji because the major archived resources was still paper centric. The Japanese government defined a guideline to develop disaster archives in which they suggested a simple metadata scheme defined based on Simple Dublin Core[1] because, on one hand, the archive resources had to be collected and organized as a database in limited time and costs and, on the other hand, the metadata should be interoperable across digital archives.

We analyzed metadata from five disaster archives that participate in Hinagiku. As these archives were developed by different sectors, each archive had their own features. A feature common among the archives was that a metadata was basically created for every single item and, in general, the quality of those metadata was not high as mentioned earlier. As a result, they had common usability issues; for example, numerous results for a simple query, such as multiple photographs taken at a single location, and low-quality descriptions in the subject and title fields of metadata.

(2) Research Problems

A fundamental problem we have found is the need to create a set of items by aggregating metadata and providing users with set-level access functions in addition to item-level access. For example, photographs taken at a single event should be aggregated as a set of photographs of the event. We need to automatically create metadata for an aggregated instance. However, this aggregation is not a simple task because contextual information to semantically group the items is not given in most cases.

(3) Approaches and Some Findings

The followings are approaches in our project,

(a) metadata aggregation
- aggregating metadata of photographs by time and location (longitude/latitude) information
- aggregating metadata by co-occurrence of subject terms

(b) building regional ontology resources
- domain ontology for local service and sectors using terms extracted from metadata
- datasets to record provenance of geographic entity names such as villages and towns

In the approach (a), we did metadata aggregation by time-location information of photographs collected by the three archives, Aomori Archive, Kuji-Noda-Fudai (KNF) Archive and Michinoku Shinrokuden. Aomori and KNF are developed by regional governments and Shinrokuden is developed by Tohoku University and covers broader area than the former archives. The aggregation process: (1) collect photographs whose location information is within an area represented as a hexagon whose edge length is approximately 1600m, and then (2) sort photographs by time of creation of each of the collected photograph sets and split the photographs if an interval between two consecutive photographs are larger than 30 minutes. For subject-term based aggregation, we applied Levenshtein distance for clustering subject terms. For each term cluster, we aggregated metadata which included at least one subject term contained in the cluster to form an aggregated metadata set. Table 1 and 2 show the results of time-location and subject-terms approaches, respectively (Seki, 2018). The number of resources shown in the tables differs as not all metadata had time-location information or proper subject terms. This results show that we can aggregate resources fairly effectively, but neither of these approaches is perfect. Precise evaluation is still left for our future work.

[1] The metadata schema is defined based on DC-NDL, which is a metadata schema defined by NDL based on Simple Dublin Core.

https://doi.org/10.23106/dcmi.952139012

Table 1 Aggregation by Time and Location

Archives	Metadata	Sets	Size=1	Size>99
Aomori	48,338	6,571	1,599	102
KNF	72,894	8,531	2,023	189
Shinrokuden	96,441	8,188	962	59

Table 2 Aggregation by Subject Terms

Archives	Metadata	Sets	Size=1	Size>99
Aomori	68,032	3,596	1,060	1,332
KNF	127,383	8,290	2,240	609
Shinrokuden	124,552	7,910	2,047	4,612

Metadata: Number of Metadata Instances
Sets: Number of Sets created by Aggregation
Size: Number of Aggregated Sets of the given Size

We have not obtained aggregation results in approach (b) but we have learned from approach (a) that vocabularies to represent regional entities such as place names and organizations are crucial to link data within and across archives. In addition, we are currently developing a dataset which stores change history of place names, e.g. towns and villages, to link data in the disaster archives and some other resources such as Tsunami Digital Library (http://tsunami-dl.jp/) which collects many papers, reports and news articles about disasters caused by Tsunamis in Japan since the 1890s.

From our analysis on the metadata of the component archives, we have learned that automated aggregation of contents in a single archive and metadata creation for the aggregated contents is crucial. As each participating archive contains items related to the disaster and events directly or indirectly caused by an earthquake, a crucial issue to improve usability, in particular for the local communities, is to link the disaster archive to other resources in order to cover longer period of time and to help regional communities keep their memory safe and see their history in future.

3.2. Metadata Model for Aggregating Manga Resources

(1) Project Background

Manga, Anime and Video Games are very popular Japanese pop-culture types, each with numerous commercially published materials. However, there is no commonly accepted metadata standards defined for those materials. For example, libraries use MARC for bibliographic descriptions of items of these resource types, while MADB defines its own metadata schema and its underlying data model for each.

Interoperability across the metadata databases of different types was a crucial requirement for the database design of MADB. FRBR was used as an underlying framework in the design of the data models which define classes to represent component instances such as a monograph, a series of monographs, and a game package. Each class is given a set of properties to describe attributes of a component instance. The first author of this paper was involved in the data model design as an advisory group member with some of his co-authors. From the discussions on the data models, the authors learned many interesting features of Manga and other pop-culture resources, which were used in the research project described in the paragraph below.

(2) Research Problems

The primary research issue of the Manga metadata project described in this paper is to develop a metadata model for aggregating Manga metadata of MADB as an institutional metadata and metadata presented in web resources such as Wikipedia and fan-created sites e.g., AniDB and MyAnimeList. The fundamental problem for aggregation is the differences of their description levels; MADB is primarily developed based on Item/Manifestation and those web resources are Work/Expression based. This is quite natural because MADB is developed using bibliographic data of Items held by memory institutions such as NDL and Kyoto International Manga Museum and descriptions in the web resources are likely on Work and Expression levels. The research issue is how to bridge the gap. In our earlier study, we applied EDM for aggregation (Kiryakos & Sugimoto, 2015). In the current research, we are using OAI-ORE as the base aggregation model and have defined a hierarchical model to describe entities of Manga, which may be extended to other resource types (Kiryakos & Sugimoto, 2018). The hierarchical model has three levels – Superwork,

https://doi.org/10.23106/dcmi.952139012

Work and Volume. Work and Volumes roughly correspond to Work/ Expression and Manifestation/ Item (see FIG. 1.). Superwork is defined as an entity aggregating Works in different media created under a single franchise such as Gundam, Dragon Ball, and One Piece (Kiryakos et al., 2017) (Lee et al., 2018).

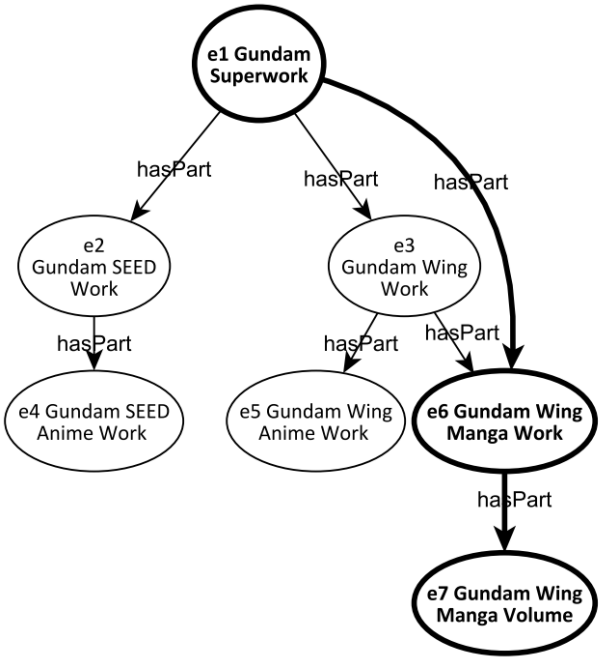

FIG.1. Three Layer Model (Superwork / Work / Volume)

(3) Approaches and Some Findings

Metadata mapping of MADB across resource types: From the advisory discussion for MADB, we have learned that class-level matching using the data models prior to property level mapping across resource types works well for the metadata mapping task. We chose a minimum set of classes defined in each resource type of MADB, and found that identifying corresponding core classes was a key issue in the matching across resource types. These classes may stand for Description Template of the DCMI application profiles.

Metadata aggregation for Manga and other resource types: The focus of our research is to define a simple data model for metadata aggregation using MADB and fan-created web resources. We found FRBR Group 1 entities were useful in their hierarchical structure, but few web resources described entities in the way FRBR had defined them, thus leading to the definition of a model composed of the three levels – Superwork, Work and Volume. We defined Superwork as an entity different from Complex Work of FRBR object oriented (FRBRoo). Note that a detailed discussion about the Superwork entity is not included in this paper.

3.3. Modeling Cultural Heritage Objects for Digital Archives

(1) Project Background

This on-going study involves the creation of a data model to describe cultural heritage objects and their metadata for digital archives with a focus on intangible cultural heritage objects. This project started from the viewpoint of digital archives for South and Southeast Asia (Wijesundara, Sugimoto, Narayan, & Tuamsuk, 2016) (Wijesundara, Sugimoto, & Narayan, 2015). Digital archives by memory institutions in the region are not well developed. However, on the web, we can find rich digital cultural heritage resources of the region developed by memory institutions in Europe and North America. Websites such as Wikipedia also provide detailed description on cultural heritage. Therefore, we think that aggregation of metadata taken from those sites is an

https://doi.org/10.23106/dcmi.952139012

important aspect to develop digital archives of cultural heritage in the regions.

(2) Research Problems

Institutional digital archives of cultural heritage are built on institutional metadata. A significant regional issue is the development levels of institutional metadata. A crucial domain-specific issue is the lack of well-developed and standardized metadata models for intangible cultural heritage. As mentioned above, we need to aggregate web resources and the institutional metadata to obtain detailed descriptions. This is the same issue discussed in the previous section on pop-culture metadata. Moreover, objectives of metadata description are not clear in the case of intangible cultural heritage because we can only archive records of a particular performance(s) inherited by a person or a community, e.g., dance, music, craftsmanship, etc. Those recorded materials may be a photograph, video, data captured by sensors, or documents. Thus, collecting performance records is essential to develop digital archives of intangible cultural heritage.

However, we need to archive many tangible objects such as costumes, instruments, and music scores, together with those records to preserve intangible cultural heritage. Moreover, we need to collect descriptions about the cultural heritage to link these tangible objects and recorded materials. "We need to collect information and aggregate them to answer these requirements" was our primary research question. Then, we had further questions "What data model is suitable for intangible cultural heritage?" and "Can we define digital archiving as an intellectual creation activity?"

(3) Approaches and Some Findings

We have defined a model named CHDE (Cultural Heritage in Digital Environment) to describe a process to organize digital archives for both tangible and intangible cultural heritage and to identify entities to be described in metadata about cultural heritage objects and their digital surrogates (Wijesundara, Monika, & Sugimoto, 2017). In the design process of CHDE (see FIG 2.), we applied the One-to-One Principle of Metadata to clearly identify the relationships between metadata and its objective of description in order to avoid confusion in the process of metadata aggregation. We analyzed several cultural heritage metadata taken from institutional and non-institutional data resources, e.g., British Museum and Wikipedia, to define a metadata aggregation scheme based on CHDE. We first categorized the properties of the metadata into four categories, which are original cultural heritage objects, their digital surrogates, administrative information and other miscellaneous entities named as external resources. Then, we grouped the properties in each category into sub-categories based on the classes of description objectives, e.g. agent, location, rights, and so on. The first level categories are useful to identify objectives in non-One-to-One situation, and the second level categories are useful to define metadata mappings across archives.

We proposed a model which uses FRBR Group 1 entities (FRBR WEMI) to identify intellectual creation by curators who organized digital archives of cultural heritage objects (Monika, Wijesundara, & Sugimoto, 2017). We are investigating the applicability of FRBR WEMI to digital curation and exhibition as intellectual creative activities and products by curators. Works of media arts are often dynamic objects, which have similar feature with intangible cultural heritage. The metadata model proposed in this project which is based on CHDE may be applied to those dynamic objects. And, exhibitions are intangible by nature. In general, metadata about exhibition is not a collection metadata because exhibitions are events as well as collections of cultural objects. The metadata model for exhibition will help find and re-use the intellectual creations by curators.

4. Discussions and Lessons Learned

This section discusses some lessons learned from the projects described above.

(1) Data Model Issues for Metadata Aggregation

https://doi.org/10.23106/dcmi.952139012

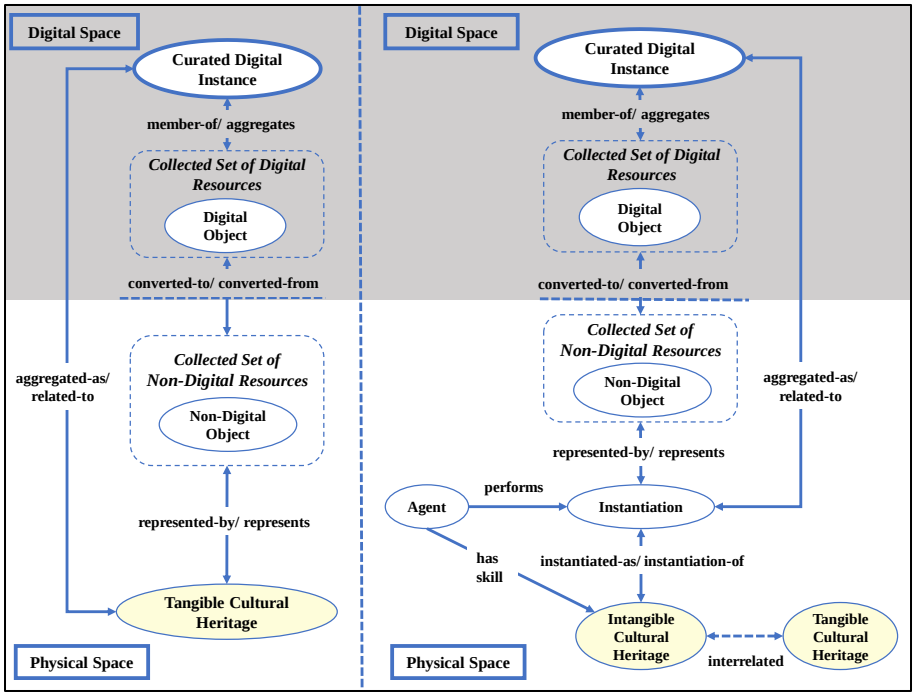

FIG.2. CHDE Model

Metadata aggregation is a key technology to improve usability of various cultural resources available on the web. There are several different types of metadata aggregation; for example, (A) metadata aggregation based on a standardized protocol among digital archives, (B) metadata aggregation among heterogeneous data sources such as institutional digital archives and websites, and (C) metadata aggregation within a single digital archive. Europeana is type A, the Japanese pop-culture project and cultural heritage project mentioned above are type B, and the metadata aggregation of the disaster archives and analysis shown in section 3.1 is type C. We have learned that the most fundamental issue common among these types is that objectives of metadata description should be identifiable because those objectives are often used as a key for aggregation. This issue is crucial in particular for aggregating metadata built on different schemes (type B). CHDE and the hierarchical model presented above are designed based on this aspect.

(2) Data Model for Digital Archives of Non-Conventional Objects

Conventional digital archives are mostly a collection of digital copies of cultural and historical resources. Those digital copies are surrogates of the original cultural heritage objects which have to be identifiable. Original cultural heritage objects may be easily identified in the case of tangible cultural heritage objects maintained by memory institutions. However, the same scheme may not work well in the case of intangible cultural heritage objects, natural and man-made events, and dynamic objects whose actions have values. In these cases, we need to archive the records of the original objects as their digital surrogates. The relationships among the original objects, their records and the digital surrogates of the records have to be clearly described. Thus, the underlying model for non-conventional digital archives and the model for conventional cultural digital archives are different. This difference may not be significant when developing a single digital archive in a single domain, though it becomes problematic when attempting to link multiple archives and websites. We have learned that we can use One-to-One Principle of Metadata as a simple and clear guideline to define data models to help linking objects across archives and to help clearly identify rights in accordance with each object. There are some arguments on One-to-One but we think it is a crucial concept for digital archives (Urban, 2014).

(3) Requirements for Domain Knowledge – a KOS-oriented View

A common issue for unconventional areas is the lack of controlled vocabularies for describing subjects, types and classes of resources in each area. We are lacking authority records in the non-

https://doi.org/10.23106/dcmi.952139012

conventional domain, e.g., descriptive subject headings for pop-culture resources, authority data of characters appear in Manga, Anime and Game, and local terms used in a certain region for disaster archives, and so forth; there is no easy answer to solve this issue. As Wikipedia and other websites are good sources of terms used by fans in the case of Manga and other pop-culture domains, an obvious task would be the extraction of terms from these resources. Automatically identifying semantic relationships between terms, however, would be difficult. Changes to the terms and vocabularies over time is also a crucial issue from the viewpoint of longevity of digital archives.

(4) Metadata Mapping across Different Resource Types

Property level mapping is frequently done for defining a mapping function between different metadata tables. This methodology works well in cases where the tables are not very different, or the tables have fewer numbers of attributes. In the MADB discussion, we have learned that this mapping scheme does not work. We found that we need to identify the entities in each resource type which are the objectives of metadata description, then find entity pairs for mapping across resource types prior to any property level mapping. We need underlying data models explaining entities of the resource types while also defining the classes of those entities. A FRBR-based model was used by the MADB project to identify entities and define their classes for mapping across the three resource types. We think that this class-based mapping framework, which is primarily similar to Description Template-based mapping across different schemas, could be used with various applications that require metadata schema mappings.

(5) Viewing Cultural Digital Archives as Intellectual Creation by Digital Curators

Digital curators who collect digital resources and organize them into a digital archive create various descriptions about the resources and archives, e.g., an exhibition webpage and catalogue. Those descriptions add significant value to the digital resources because they provide contextual information of the cultural resources. An exhibition program may be entirely or partly reused in other events as an intellectual Work. Organization of contents and their visual presentation in an exhibition are Expression of a Work. Digital curators add value to the original cultural objects in various aspects, e.g., selecting, organizing, describing, etc. Clear identification of their intellectual contributions in the organization of digital archive metadata is useful not only to reuse their products but also to create new intellectual products.

5. Concluding Remarks

The research goals of these three projects are not the same and the objectives and content of metadata descriptions are different. However we have discovered problems common across these projects, namely the importance of metadata aggregation for better usability of digital resources, demands to vocabularies based on the domain knowledge, and the importance of underlying data models to connect resources across digital archives of different resource types. We used general models for metadata such as One-to-One Principle and Application Profiles as well as domain specific models such as FRBR and EDM. While these models do not always work well with real-world metadata because of the metadata itself and its schema qualities, we have found that these models have nevertheless played crucial roles in our projects and have taught us valuable lessons.

Acknowledgements

We thank Yusuke Seki and all students who contributed in the research projects presented here. This study is supported by JSPS KAKENHI Grant Number 16H01754.

References

IFLA. (2009). *Functional Requirements for Bibliographic Records*, p. 136, Retrieved, May 2, 2018, from https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

Isaac, A. (ed.). (2013). *Europeana Data Model Primer*. Retrieved, May 14, 2018, from http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

https://doi.org/10.23106/dcmi.952139012

Kiryakos, S., & Sugimoto, S. (2018). Aggregating manga metadata from diverse data providers using the portrayal of bibliographic hierarchies on the Web based on OAI-ORE, Manuscripts in preparation

Kiryakos, S., & Sugimoto, S. (2015). A Linked Data Model to Aggregate Serialized Manga from Multiple Data Providers. *Lecture Notes in Computer Science*, 120–131. doi:10.1007/978-3-319-27974-9_12

Kiryakos, S., Sugimoto, S., Lee, J. H., Jett, J., Cheng, Y., & Downie, J. S. (2017). Towards a Conceptual Framework for Superworks. *Proceedings of JADH 2017*, pp.47-49.

Lee, J. H., Jett, J., Cho, H.R., Windleharth, T., Kiryakos, S., Disher, T., & Sugimoto, S. (2018). Reconceptualizing Superwork for Improved Access to Popular Cultural, Objects. *Proceedings of ASIS&T 2018*. Manuscript submitted for publication.

Miller, S. (2010). The One-To-One Principle: Challenges in Current Practice. *International Conference on Dublin Core and Metadata Applications*, pp. 150-164.

Monika, W., Wijesundara, C., & Sugimoto, S. (2017). Modeling Digital Archives of Intangible Cultural Heritage Based on One-to-One Principle of Metadata. *Proceedings of the 8th Asia-Pacific Conference on Library and Information Education and Practice (A-LIEP)*, pp. 137-148.

Nilsson, M., Baker, T., Johnston, P. (2008). *The Singapore Framework of Dublin Core Application Profiles*, Retrieved, May 2, 2018, from http://dublincore.org/documents/singapore-framework/

Seki Y. (2018). A Study on Improvement of Usability of Digital Archives of Great East Japan Earthquake using Resource Aggregation, Master Thesis, University of Tsukuba, 46p.

Urban, R. J. (2014). The 1:1 Principle in the Age of Linked Data, *International Conference on Dublin Core and Metadata Applications*, pp.119-128.

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). Digital Libraries: Data, Information, and Knowledge for Digital Lives. ICADL 2017. *Lecture Notes in Computer Science*, Vol 10647. pp. 81-94, Springer. doi:10.1007/978-3-319-70232-2_7

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*, 2(1), 5.

Wijesundara, C., Sugimoto, S., Narayan, B., & Tuamsuk, K.: Bringing Cultural Heritage Information from Developing Regions to the Global Information Space as Linked Open Data: An Exploratory Metadata Aggregation Model for Sri Lankan Heritage and its Extension. *7th Asia-Pacific Conference on Library and Information Education and Practice (A-LIEP)*, pp. 117–132 (2016).

Woodley, M. S., Clement, G., & Winn, P. (2005). *DCMI glossary*. Retrieved, May 2, 2018, from http://dublincore.org/documents/usageguide/glossary.shtml

https://doi.org/10.23106/dcmi.952139012