

Cultural Heritage Online:

Information Access across Heterogeneous Cultural Heritage in Japan

Noriko Kando, Jun Adachi

National Institute of Informatics, Japan

{kando|adachi}@nii.ac.jp

Abstract

This paper discusses the metadata schema for Japan's *Cultural Heritage Online* project. The purpose of the project is to set up a portal site providing seamless access to heterogeneous digitized cultural heritage objects across a wide variety of digital collections prepared by archives, museums, national, regional and local cultural heritage centres, and other related organizations in Japan for both Japanese and international users. It covers both tangible objects such as paintings, buildings and other artefacts, and intangible objects including theatre performances and dance, as well as art that creates artefacts. The key issues in system design are mechanisms for continuous search-and-navigation through a combination of content- and structure-based retrieval. Metadata and community-oriented ontology are the main components on the structure-based side, together with an associative search engine on the content-based side. In conclusion, problems and future directions in design of structure and search functionality are discussed.

Keywords: *Cultural Heritage, Metadata, Heterogeneous Metadata*

1 Introduction

Cultural Heritage Online is a portal site to provide access to various digitized cultural heritage objects in the collections of museums, archives and related organizations in Japan. It was proposed in an interim report [1] of the Committee on Cultural Heritage Digitization Strategy, under Japan's Agency of Culture. The report also emphasized the necessity to encourage the digitization of cultural heritage objects in the various organizations. It strongly recommended that *Cultural Heritage Online* should cover 1000 sites of museums, archives and other related organizations by the end of fiscal year 2006.

In response to the report, the committee's Technical Subcommittee has discussed detailed problems to establish a roadmap of the project by April 2004, and a pilot system was implemented to encourage and mediate discussion and dialogue among the various communities related to the creation and use of cultural heritage objects. The roadmap will pay special attention to metadata

schemes and rights management. The discussion and dialogue include, but are not limited to, the implications of digitization and building a portal site, the site's search functionalities and usability, information architectures including the metadata schema, types and formats of the digitized objects, and the current status of the information management of the cultural heritage objects in each site.

1.1 Digitization of Cultural Heritage

Unlike library materials, which are basically published objects, and other tokens of the same type that are available elsewhere, cultural heritage objects are basically *unique* and their usage and accessibility are quite limited in their original physical form. The implications of the digitisation of cultural heritage objects are tremendous and include the following aspects:

1. Enhanced usage and accessibility;
2. Multiple versions for different user groups or purposes;
3. Independent from the collection or context;
4. Virtual combination, comparison, or restoration;
5. Preservation.

The digitized objects are accessible regardless of geographic location. Especially this enlarges the opportunities for educational purpose use and benefit for enhance the mutual understanding between different cultures in the world. Multiple versions of images are often available for different purposes, for example, ultra high resolution images for publication, broadcasting, virtual exhibition or other content industries; thumbnails to quickly identify the relevance of the objects, typically in search systems; and mid-level resolution images for classroom use. By digitization, any object can easily be moved to a different location or context from the original collection, a new collection or comparison with other objects can be virtually constructed, and objects can even be restored virtually.

1.2 Digitization and Metadata

The importance of the metadata is increased when cultural heritage objects are digitized. For example, the metadata can improve the search effectiveness and usability of the search system by providing

multiple access points and preserving the semantics and context of the objects. The metadata is also critical in linking the multiple versions of the same object and objects from the same collections. It can provide detailed description frameworks appropriate for each community as well as more general frameworks for resource discovery across different communities. Information for preservation and rights management can be recorded as metadata.

The scope of *Cultural Heritage Online* is introduced in the next section. Section 3 describes the pilot system and its functionality, and Section 4 discusses the difficulties and problems of metadata for cultural heritage objects. Finally, some thoughts on future directions are presented.

2 Scope

The scope of the cultural heritage objects included is quite wide and heterogeneous. As shown in Figure 2, we plan to include such types of cultural heritage object as tangible objects such as paintings, sculptures, crafts, archaeological objects, historic sites, architecture and buildings, scenery, natural monuments and protection intangible objects like performance and dance, and the arts of creating the artefacts and crafts. Each community related to a genre has its own culture and ontology to describe the objects. To provide access across these varieties of communities is one of the challenges of *Cultural Heritage Online*.

The categorization of the object types itself is a matter for discussion. For example, should an object be categorized by the materials and techniques used in its creation (e.g., porcelain), or by usage (e.g., tea wares)? Such questions are deeply related to the ontology of each community and school.

In the database, the formats of the digitized objects are basically metadata and thumbnails of the digitized objects in each digital archive collection, with links to these objects for further information and more detailed images. The pilot system also accepts multiple sizes and resolutions of images, and some videos.

3 Pilot System

A pilot system was implemented as a tool or medium to encourage discussion and dialogue among the various communities related to the creation and use of cultural heritage objects.

The primary target users are non-specialist ordinary citizens without any technical or professional background in cultural heritage. School students and teachers are one group likely to use the system heavily. The users can initiate the search process

without defining their information needs as specific queries, and enjoy the interaction.

Currently the pilot system contains about 5,000 records, provided by 35 museums, archives and other related organizations for experimental purposes. We are grateful for their cooperation and quick responses to our requests.

The current version of the pilot system was implemented by Marukawa and Takano by modifying their experimental system called *Mozume* [2].

Figure 1 shows an overview of the search mechanism of the pilot system. The basic design concept is “search and navigation”. It combines content-based modern information retrieval technology using statistical features of the objects with metadata-based navigation.

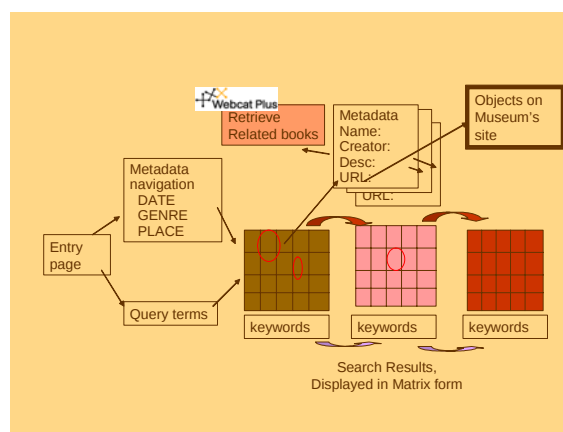


Figure 1. Search mechanism of the pilot system

On the pilot system’s top page, shown in Figure 3, a user can enter query terms in a window or select some values from the pictorial pull-down menus for each facet of metadata, Time, Type or Place. The retrieved records are then displayed in the matrix format so that the user can see as many images as possible at a glance.¹ Figure 4 shows an example of the retrieval results when the user selected “porcelain” from the genre facet metadata pictorial menu shown in Figure 2.

Among the displayed images, if the user is interested in *a round dish with red or pink flowers*, then the two are selected from the displayed results, and searched again. As shown in Figure 5, objects similar to the selected items are retrieved and displayed. In this way, users can continue the search and navigation as far as they wish until they are satisfied. In such a system, the total experience and

¹ The matrix format display is commonly used by online shopping sites.

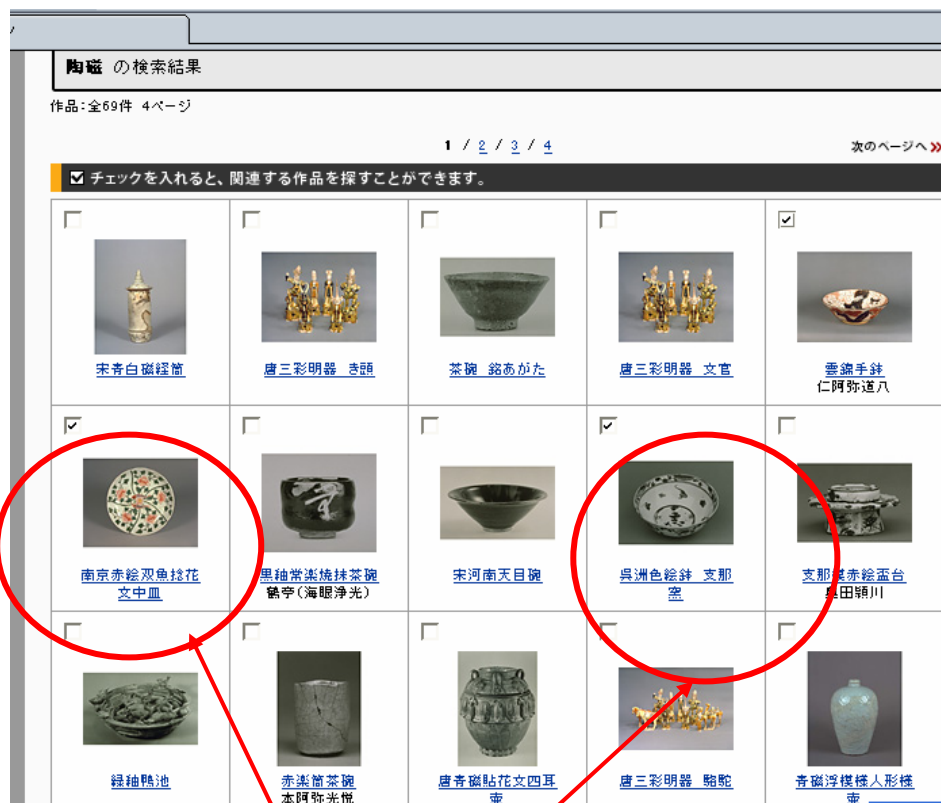


Architecture	Historic Villages	Pictures/Paintings	Prints
Religious	Samurai towns	Japanese	Wooden
Castles	Post towns	Oil paintings	Etchings
Houses	Ports	Water paintings	Lithograph
Modern	Farm/Mt Villages	Asia (Non-Japanese)	Silkscreen
Pre-modern	Others	Others	Others
Sculptures	Craft/Artefacts	Archaeology	History
Wooden	Metal	Stoneware	Documents/Books
Metal	Lacquer	Earthenware	Maps
Stone	Dyeing/Textile	Metalware	Others
Bone	Porcelain	Bone/teeth	
Others	Glass	Others	
Other Arts	Folk Art	Traditional	Restoration
Photographs	Tangible	Performing Arts	Architecture
Design	Intangible	Noh	Paintings/Sculptures
Handwriting		Bunraku puppets	Historic sites
Others		Kazuki	Traditional
		Music	Artefacts
		Others	
Historic Sites	Scenic Beauty	Natural Monuments/Protection	
Ancient tombs	Gardens	Protected animals	
Temples/Shrines	Ravines/Rivers	Protected plants	
Castles	Castles	Geological features/Minerals	
Villages		Protected areas	
Others			

Figure 2. Genres of Cultural Heritage Objects



Figure 3. Top Page of the Pilot System



Use these selected records as a query

Figure 4. Two Round Dishes with Red Flowers are Selected from the Search Results of “Porcelain/China”



Figure 5. Search Results of the Figure 3.

everything that the user learns through interaction with the system are the results of the retrieval.

For the content-based retrieval, *Mozume* and the pilot system use a search engine called GETA [3]. It is a content-based *text* retrieval system, and therefore the pilot system currently does not utilize any content-based information from images, but uses textual description in the metadata. In GETA, documents are usable as queries, and the system retrieves related documents from the user-selected documents and provides a list of highly associated keywords that can be used to enhance further retrieval. Using this associative search function, users can progressively search for similar objects.

This is also similar to the concept of the “Ostensive Search”, search without query, proposed by Ian Campbell [4], and thought to be effective and useful for users who do not have clear search requests prior to the search.

From the textual description of the retrieved metadata, information about related books can also be retrieved using NII’s WebCAT Plus [5], an Online Union Catalog Database freely available on the web, which also incorporates associative search functions using GETA.

3.1 Simple Metadata

The current version of the pilot system barely utilizes metadata, because it was implemented before the detailed discussion of the metadata schema. The metadata submitted from 35 museums and other related organizations contained such fields as title, title.yomi (pronunciation of title), description, number, size, designation (for instance, as national treasure), materials, structure and technique, creator, publisher, contributor, date.created, date.published, date.collected, subject.local-classification, URL, object id, place.produced, place.collected, place.used, place.found, and place.archeological-site-name. The pilot system uses only the very simple facets of date, genre and place in the search interface for navigation using pictorial menus. Institutions were asked to provide descriptions at least 300 characters long (about 150 words in English) to allow effective associative search.

Figures 5 and 6 show the pictorial menus for the facets of DATE and PLACE. An interesting point is that DATE and PLACE are integrated with each other. The era is defined by each country or area, and PLACE name can vary according to the time period.



Figure 5. Example of Pictorial Menu on DATE/ERA (for Japan)



Figure 6. Example of Pictorial Menu on PLACE

3.2 Iterative Improvement and Redesign

The technical subcommittee has discussed the metadata schema, and according to the roadmap we have set, specialized working groups consisting of curators and information architects from each community will decide the metadata schema to be used in *Cultural Heritage Online*. Ontology development is also included in the task of this working group.

A combination of content-based retrieval and structure-based search using metadata features theoretically has a good possibility of working well, as the two complement each other and are especially effective for a large-scale database with a rather small controlled vocabulary or less-controlled metadata descriptions. Algorithms to combine the two approaches more effectively will also be investigated.

The search functionality and metadata schema will be gradually improved through an iterative process of usability tests or evaluation by users and creators, and redesigning.

The next section discusses some of the issues related to metadata raised by the discussions we have had so far. These may depict some of the problems and challenges regarding metadata of cultural heritage objects.

4 Discussion of Cultural Heritage Metadata and Further Functionalities

To utilize all the advantages of cultural object digitization mentioned above, effective and easy-to-use search functionality and appropriate metadata schemas are required.

4.1 Standards for Cultural Heritage Metadata

For the cultural heritage objects and related areas, there are several well-used standards for metadata. As a basis for discussion, we have surveyed these standards and their inter-relationships. They include: SPECTRUM, by the Museum Documentation Association (MDA); CIDOC CRM, by the International Council of Museums Documentation Committee; the Simple Dublin Core and CIMI's Guide to the Best Practice Dublin Core; Categories for the Description of Works of Art (CDWA); Encoded Archival Description (EAD); and other work by some major museums and archives.

Based on the survey, we are reviewing the objects currently included and to be included in *Cultural Heritage Online*. Because of its wide variety, including intangible objects and scenery, and with

consideration of the workforce in each member museum to creating the metadata, we are discussing metadata requirements carefully.

The primary aim of the metadata for *Cultural Heritage Online* is to provide access across heterogeneous objects, i.e., metadata for resource discovery and interoperability. Therefore it will basically be rather simple, but we hope that the mapping and conversion from each community's metadata and ontology can be done comfortably for all communities.

4.2 Discussion of Cultural Heritage Metadata

Below are some examples that we have considered so far. The point here is that the design of the metadata schema is deeply related to the design of search functionalities, especially the user interface to multifaceted metadata for navigation.

Titles

In some types of cultural objects, the title is not clearly defined, or more precisely, only a rather small number of objects types have "titles", which they have often been given in recent times. For example, the titles of archaeological objects are usually object names. The titles may be also changed for each exhibition. Naming is sometimes a right of the owner; owners of objects may name them according to their preferences.

Owners and History

Related to the above observation, the owners of the objects and the histories of the owners are often critical attributes to differentiate one object from others. Relations such as "who created this for whom?" and "who gave this to whom?" are useful information to differentiate objects as well as to envisage the value of the objects.

Value

Cultural heritage objects are generally valuable, and users often wish to search and enjoy "valuable objects" without a clear definition of how valuable they are. However, as a metadata record to describe the object, it is not appropriate to say, for instance, "it is valuable". Instead, descriptions of attributes that indicate the value of the object are often useful for this purpose. Awards, designation as National Treasures, signatures of creators, or the signature of the eminent people who owned it before or to whom it was dedicated are examples of attributes indicating value.

Relationships

Cultural objects do not usually exist alone, and are often part of a collection or have relationships to other objects. This provides the context of the object, and without these contexts, the value and indication of the objects cannot be assessed correctly.

Collections vs. Single Objects

The choice of collection-based description or item-based description is often an issue.

Community-based Metadata or Ontology

Each community has its own metadata schema and ontology. In particular, intangible objects including Japanese traditional theatre performances such as *Kabuki* have very strong traditions. More detailed analysis of this domain is necessary. Scenery is also a characteristic object included with cultural heritage objects, but is often closely related to architecture and religious objects.

Scaling and Fuzzy Matching

Numeric descriptions of size or year are often too rigid and strict. In the historical record, there is substantial confusion about the time periods of dynasties. It will thus be useful and more practical to specify these numeric values in more tolerant or vague ways.

Place Names and GIS

Place names and the places indicated by a given name are not stable over time.

Exhaustivity vs. Selectivity

What objects should be included in *Cultural Heritage Online*? Should we try to include the whole collection of every museum, or should each museum select the “good” or “valuable” objects that they wish to show many people all over the world?

Isolated Objects vs. Systematic Knowledge

The current pilot system has search functionality only over isolated objects. We can enjoy finding unexpected relationships or similarities between objects through associative search. Often, however, we would like to gain more-systematic knowledge about objects and their relationships, and understand the value and meanings of an object by relating the object to systematic knowledge. How to construct the data for systematic knowledge and how to implement it on the system is an interesting challenge.

Currently, the pilot system has a link to NII's WebCAT Plus, a web-based union catalogue search service of Japanese university and research libraries, which is also powered by the same search engine with the pilot system. The users can retrieve the related books using the retrieved cultural heritage object as a query. We also solicit museum curators or other specialists to contribute “virtual exhibitions” to connect and relate isolated objects in systematic ways. This is an example of an attempt to overcome the problem of simple aggregation of single object and providing some systematic view among them by human effort.

Rights

We tried to restrict the scope of the system to description and discovery of the resources. Rights management is not the primary target of the project and will be done elsewhere. However, we must still consider rights management to some extent.

Paradigms and Viewpoints

Description of cultural heritage objects may differ in the principles, paradigms, viewpoints and interpretation of each creator of the metadata and users. When aggregating metadata from various sites, there can be conflicts between descriptions and values.

Other Issues to be Investigated

Further possible research and investigation for better information access for cultural heritage objects includes: cross-lingual information access, especially for Asian communities; content-based retrieval using image content information combined with textual information in metadata and metadata; and automatic metadata enrichment using natural language processing techniques.

6 Summary

This paper offers a brief overview of the *Cultural Heritage Online* project and discusses the issues relating to the metadata for cultural heritage. The project itself is under way, or more precisely will start from this coming April. From April we are organizing a working group to discuss in detail about metadata scheme in each of the communities related to cultural heritage, then finalize the metadata scheme used for the Cultural Heritage Online, as well as organizing various attempts and effort to digitizing the cultural heritage and enhancing the access to them. Any comments, leads or suggestions are always more than welcome.

References

- [1] “*Bunka Isan Johoka Suishin Senryaku Chukan Matome* (Interim Report. Strategy of Cultural Heritage Objects Digitization, translated by the authors)”. Bunka Isan Johoka Suishin Senryaku Kaigi, 26 August 2003, 21pp. (in Japanese)
- [2] Marukawa, Y., Takano, A., Takamizawa, A., “Mozume: An Associative Information System for Cultural Heritage”, in: Proceedings of the Second NII International Symposium, Nara Symposium for Digital Silk Roads, December 2003, Nara, Japan
- [3] GETA. The Source code is available at: <http://geta.ex.nii.ac.jp/e/index.html>
- [4] Campbell, I., Reisbergen, K. “Ostensive Retrieval of Image Documents”, in: Proceedings of the Second CoLIS, October 1996, Copenhagen, Denmark
- [5] WebCAT Plus: <http://geta.ex.nii.ac.jp/e/index.html>