DIGITAL ARCHIVE SYSTEMS USING CMS AND GALLERY TOOLS – IMPLEMENTATION OF ANTHROPOLOGICAL MUSEUM –

H. Kawano[*]

Dept. of Systems Design and Engineering, Nanzan University, Aichi, Japan, 4890863 - kawano@nanzan-u.ac.jp

KEY WORDS: Digital Archive, Digital Museum, Anthropological Museum, Metadata schema, Contents Management Systems

ABSTRACT:

Recently, we have good opportunities to integrate various digital collections and archive systems using loosely coupled and meaningfully connected hyperlinks from various view points of cultural, social and technical aspects. In this paper, from technical view points, we present our implementation of digital anthropological museum using CMSs (Contents Management System) and gallery tools. Firstly, we have a brief comparison of digital archive software including popular CMSs. Recent years, Museums, Libraries and Archives (MLA) are making efforts to provide various digital collections and born-digital information by using digital archive systems. For example, major national libraries and long-term preserving organisations, such as IIPC (International Internet Preservation Consortium), gather and preserve huge amount of web pages, and curate them using web curator tools. Secondary, we propose a metadata schema for digital collections in "*the Anthropological Museum of Nanzan University*", which is based on the guideline of museum objects. Our proposed schema based on XML metadata formats like URI/RDF/MODS. Finally, we show the system architecture of digital anthropological museum and implementation of three different prototype systems based on different CMSs add-on modules and other related software. From 2005 to 2012, we have 1776 digital collections with 74 metadata attributes and 40,000 ethnographic photographs without metadata, at present we store 277 digital contents having detail values into our prototype of digital museum system. Using interfaces of prototype systems, we also introduce workflows of museum collections including content rights management.

1. INTRODUCTION

Recent years, we have good opportunities to integrate various digital collections and archive systems using loosely coupled and meaningfully connected hyperlinks from various view points of cultural, social and technical aspects. Especially, Museums, Libraries and Archives (MLA) are making efforts to provide various digital collections and born-digital information by using softwares of digital archive systems in (Yoshimura, 2012).

For example, major national libraries and long-term preserving organisations, including IIPC (International Internet Preservation Consortium, http://netpreserve.org/), collect and curate huge amount of web resources, and preserve contents in web archive systems. Various types of digital contents, such as text, images, movies and other data types, are stored into digital archive systems, and key technologies are long-term preserving storage, database management systems, information retrieval, web services and others (Kawano, 2008; Kawano 2010, Kawano 2011, Kawano 2012).

In this paper, from technical view points, we focus on how to develop the digital archive systems in order to organise various digital collections with metadata, documents, pictures and movies. We propose a metadata schema for digital collections in "*the Anthropological Museum of Nanzan University*" and present our implementation of digital museum using CMSs (Contents Management Systems) , gallery tools and other related software.

In Section 2, we make a short summary of digital archive systems, CMSs, related software and curator tools. We have a brief comparison of various digital archive software including popular CMSs.

In order to administrate and search digital collections, various CMSs of database applications play an important role. CMS supports an archiving work flow in order to index, search, manage and archive various kinds of digital contents totally. Dspace (http://www.dspace.org) is one of famous software in order to construct open digital repositories. The mashup technology is also important in order to integrate different information resources in CMSs by using standard APIs.

In Section 3, firstly we design our metadata schema for collections in "*the Anthropological Museum of Nanzan University*", which is based on the conceptual reference model defined in "International Guidelines for Museum Object Information" (http://icom.museum/guide.html) and (International Council of Museum, 2011). We design XML metadata like URI/RDF/MODS (MODS Editorial Committee, 2010). Secondly, we present the workflow and system architecture of digital museum. We validate proposed schema by using the MSXML parser provided by (Microsoft, 2011).

In Section 4, we introduce several prototypes implemented from 2005 to 2012. First prototype digital museum is based on Joomla! and some other modules, later we reconstruct second and third prototype systems of digital museums based on Drupal and related packages of album/gallery and annotation tools. In 2011, we implement digital anthropological museum by using Drupal 7.8 on Ubuntu 10.04 LTS. At present, we store 2,054 pictures of museum collections and also associated pictures with csv file of metadata. In our previous paper (Kawano 2012), we shortly introduce the prototype, but we focus on the interfaces for management of museum collections in this paper.

---

[*] H. Kawano, hiroyuki.kawano@univie.ac.at, is a visiting researcher in Faculty of Informatics, University of Vienna from March 2012 to March 2013.

2. DIGITAL ARCHIVE TOOLS AND CMS

In this section, we introduce a brief summary of various digital archive tools for museums, libraries, archives and web archives. We also present a simple comparison of archiving systems including CMSs.

2.1 Digital Library and Digital Museum Software

From the view point of a software license, several digital library systems are useful, such as *Greenstone*, which is introduced in a book *greenstone*. In the Greenstone tutorial exercise (http://wiki.greenstone.org/gsdoc/tutorial/en/greenstone_to_dsp ace.htm), there is how to export a Greenstone collection in a form suitable for DSpace. DSpace is also a well-known digital library management tool using various types of academic repositories (Duraspace, 2012). The web site introduces *Digital Collection Builder* (Canadiana, 2012). Generally speaking, these software systems provide same kinds of functions:

· Management of digital collections
· Multimedia data, including common document formats, pictures, movie files and commonly used document formats
· OAI (Open Archives Initiative Protocol for Metadata Harvesting） metadata harvesting protocol
· Standard metadata formats, such as Dublin Core, MODS, and EAD
· Functions of importing and exporting metadata from/to different archive systems

In order to meaningfully integrate information resources stored in different archive systems, it is a first important step to utilize standard metadata formats, document formats, information exchange protocol and other functions.

In the web site of BeanWorks, a brief list includes "Comparing Digital Library Systems" (http://beanworks.clbean.com/category/technology/software/open-source/) such as digital asset management systems for digital libraries and open source software. "The Museum Association of New York" (http://manyonline.org/resources/collections-management-software/) also provides the list of management software.

2.2 Web Archive and Curation Tools

National libraries, national archives and consortia of organizations construct and maintain web archiving systems in order to preserve various web information, which are collected from both of surface webs and hidden webs. Internet Archive is a well-known service, it provides open source software for crawling, archiving and maintenance of web collections.

Firstly, *web crawlers* such as Heritrix, HTTrack and Wget, are the most basic tools for collecting web pages. Web crawlers simply collect HTML/XML web pages sequentially instead of browsing. They also gather metadata such as accessed/modified times, MIME/encoding types, and content length/structure, style sheets, JavaScript, word/pdf documents, pictures, and movies. Moreover, in order to reduce and optimize the load balance of web servers globally, advanced web crawlers have a function of distributed scheduler for gathering web contents from multiple web servers.

However, it is not possible to preserve time consistency of web contents, hyper-links inside a web server and hyper-links between different web servers. Therefore, in order to collect contents from deep web and hidden web, which are provided database systems, various techniques of database archiving and transactional archiving have been developed.

Secondly, *web curation tools* are important for constructing web archive systems. A suite of web curation tools and WARC (Web ARChive) format as ISO 28500 standard are accessed from IIPC (http://netpreserve.org/software/downloads.php). The NDIIPP Partner Tools and Services Inventory (http://www.digitalpreservation.gov/tools/) is also useful lists.

· Heritrix: an open-source, extensible, Web-scale, archiving quality Web crawler
· NutchWAX: a tool for indexing and searching Web archives using the Nutch search engine and extensions for searching Web archives
· Wayback: a tool that allows users to see archived versions of web pages across time
· Web Curator Tool: a tool for managing the selective Webharvesting process is designed for use in libraries and other collecting organisations
· Xinq (XML INQuire), a search and browse tool for accessing an XML database

2.3 Comparison of CMSs

On constructing suitable digital museum, library and archive systems, it is effective to utilize CMS and to integrate various add-on modules and related software. There is a research paper of performance analysis of CMSs in (Patel, 2011, Doulamis 2008), we make a brief comparison of major CMSs and additional modules. Using feature of web site CMSmatrix (http://www.cmsmatrix.org/), we compare popular CMSs, including Drupal, Joomla!, Wordpress, XOOPS and Zikula, from several view points of system requirement, security, support functions, ease of use, performance, management, interoperability, flexibility, build-in applications and commerce.

	Drupal 7.12 ⊗	Joomla! 2.5.4 ⊗	WordPress 3.3.2 ⊗	Xoops 2.0.18 ⊗	Zikula 1.2.0 ⊗
Last Updated	2/16/2012	5/2/2012	5/29/2012	8/5/2008	11/9/2009
System Requirements	Drupal 7.12	Joomla! 2.5.4	WordPress 3.3.2	Xoops 2.0.18	Zikula 1.2.0
☑ *Application Server*	Apache	CGI	blank	Apache	blank
☑ *Approximate Cost*	Free	Free	Free	Free	$0
☑ *Database*	MySQL	MySQL	MySQL	MySQL	MySQL
☑ *License*	Open Source	Open Source	Open Source	Open Source	Open Source
☑ *Operating System*	Platform Independent	Platform Independent	Platform Independent	Platform Independent	Platform Independent
☑ *Programming Language*	PHP	PHP	PHP	PHP	PHP
☑ *Web Server*	Any	Any	blank	Apache	Apache

Figure1: Comparison of CMS functions

3. METADATA AND ARCHITECTURE OF DIGITAL MUSEUM

In this section, we propose metadata schema for digital collections in *"the Anthropological Museum of Nanzan University"* and discuss difficulties to edit appropriate and accurate values for metadata schema. Our schema is based on the guideline of museum objects, and we propose XML metadata formats like URI/RDF/MODS. We also propose the system architecture of digital museum and implement prototype systems based on different CMSs, related modules and software.

3.1 Metadata Schema

In order to design a metadata schema of museum contents in the anthropological museum, we mainly focus on the "International Guidelines for Museum Object Information" (http:// icom.museum/guide.html) and "CIDOC Conceptual Reference Model CRM models" (ICOM/CIDD CRM Special Interest Group, 2008).

CRM model defines the relationships of various entities, such as "Man-Made Object", "Production", "Time Primitives" and others. On the other hand, in different metadata of Dublin Core, fifteen basic elements are selected and standardized, we choose following elements in our digital museum system; *"description"*, *"date"*, *"type"*, *"format"*, *"language"*, *"coverage"* and *"rights"*.

Furthermore, the guideline only provides the conceptual framework for managing museum objects. Actually, we have to define metadata attributes of information management and technical attributes of photographs and movies, which are based on other standard metadata formats.

Firstly, we have a brief survey of other standard metadata formats, such as Dublin Core of Dublin Core Metadata Initiative in (Dublin Core Metadata Initiative, 2008), MPEG-7 and MPEG-21 for multimedia contents in (ISO/IEC JTC1/SC29/WG11, 2004). We also consider metadata attributes of photographs based on MPEG and JPEG standards, after that we design our XML metadata formats.

Secondly, in our museum system, we also need multilingual attributes, so we introduce multilingual fields based on MODS metadata schema. A <Object_Name_Information> has child elements <Title_Information>, which have <lang> and <type> attributes. The description in the attribute <lang>, we use ISO 639-2b, *jp*: Japanese, *eng*: English and so on.

Finally, we design following 22 information groups and create detail metadata schema based on some examples of the museum collections.

- ・ Acquisition Information
- ・ Condition Information
- ・ Deaccession and Disposal Information
- ・ Description Information
- ・ Image Information
- ・ Institution Information
- ・ Location Information
- ・ Mark and Inscription Information
- ・ Material and Technique Information
- ・ Measurement Information
- ・ Object Association Information
- ・ Object Collection Information
- ・ Object Entry Information
- ・ Object Name Information
- ・ Object Number Information
- ・ Object Production Information
- ・ Object Title Information
- ・ Part and Component Information
- ・ Recorder Information
- ・ Reference Information
- ・ Reproduction Rights Information
- ・ Subject Depicted Information

For example, "Condition Information" has categories of "Condition", "Condition Summary" and "Condition date", we extend these categories for administrators and add several attributes, "Receiving Date", "Transportation", "Receiver" and "Return Date". In the information group of "Image Information", we need more detail attributes, such as Title, Image Reference, Keyword, Comment, Medium, Film Instrument, Artist, Film Date, Width, Height, Size, Size Unit, Film Place, Film Latitude Longitude, Photo (including horizontal and vertical resolutions and bit depth etc.) and so on. Initially, we propose and design a metadata schema with 102 attributes presented in Figure 2, and we validate our prototype schema by using the MSXML parser (Microsoft, 2011), which provides extension functions of DOM, SAX, XSLT and others.

Figure 2: Interface of importing CSV file with metadata

3.2 System Architecture

In Figure 3, we show the conceptual architecture of the museum CMS, the user send a query to the server and the contents management system retrieves various data, such as text, photograph, metadata and layout from different databases, the system dynamically creates a web page based on a user's preference. The granularity of information depends on the grant of users, such as administrators, researchers, guests and so on.
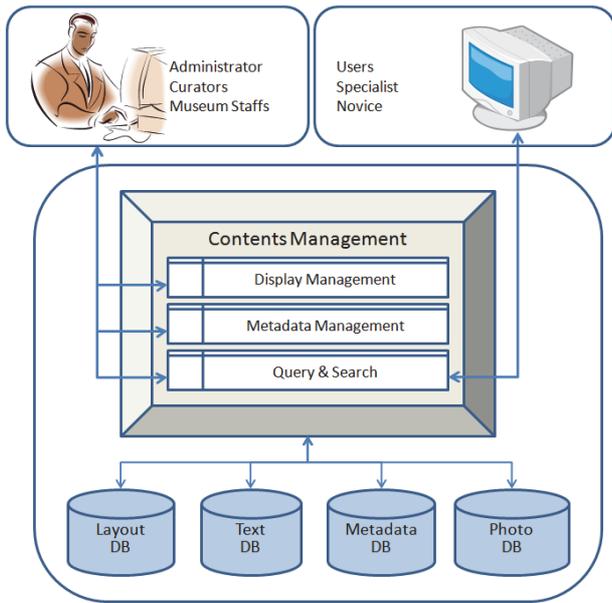
Figure 3: Conceptual framework of museum archive

4. IMPLEMENTATION OF DIGITAL MUSEUM

In this section, we present three different prototype systems of digital museum developed from 2006 to 2012. At present, we enter 2054 digital contents and a metadata CSV file in our system. As we discussed in previous section, we propose 102 attributes, but it is not easy to investigate several attributes by curators in the museum. Furthermore, 1776 contents don't have sufficient metadata values. As a result, we store 277 collections having detail metadata values in 74 attribute fields into our present prototype system.

4.1 Prototype Systems based on Joomla! and Drupal 5

As we presented in Section 3.3, there are various CMSs, such as Drupal, Joomla!, WordPress, XOOPS and many others.In 2006, we develop a first prototype museum system having photo gallery by Joomla! 1.0.11JP and related module software presented in Figure 4. We had a comparison of accessibility between museum CMS and the original museum web site, the performance of CMS is better than the original web site from various view points, such as contrast of colours, stylesheet of layout design and control, design of link navigations, W3C standard descriptions, consistency of hyperlinks and etc. However, our system does not satisfy the conditions of "Double-A" of "Web Content Accessibility Guidelines 1.0" (W3C, 1999).

In 2008, in order to improve the usability of our previous prototype system, we reconstruct digital archive by using Drupal 5.10. The top page of gallery shows a few collections having a part of attribute values.

In this second prototype system, in order to provide the annotation function, we compare several tools and web sites, such as PhotoStuff, M-OntoMat-Annotizer, Taggify and others, but it is so difficult to integrate other different modules in Joomla!. Then we move from Joomla! to Drupal in order to construct the museum archive based on the combination of Drupal and related packages of Gallery2 and Fotonotes annotation module.
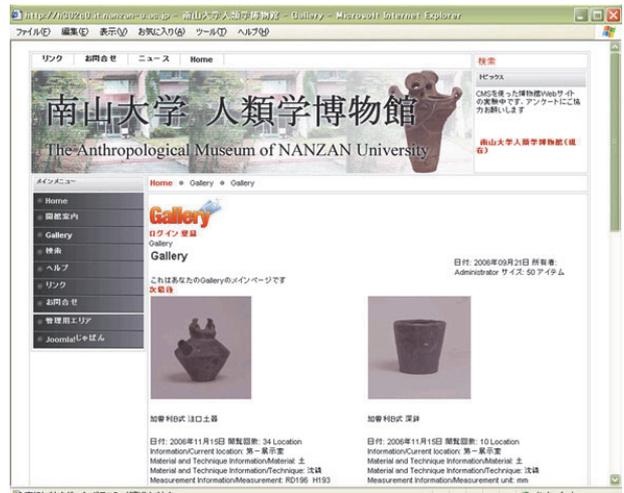

Figure4: Top page of museum archive based on Joomla! and Gallery 2

When we construct the museum archive, we consider the stability and extensibility of the system, and we select Drupal (drupal-5.10.tar.gz) and Japanese module (D_ja_resource _kit_5__7_009.tgz). In order to support annotation functions of photographs, videos and other content, we select several modules of "Gallery2" (gallery-2.2.6-developer.tar.gz), "Fotonotes" (fotonotes-5.x-1.1-beta6.tar.gz) and other tools into the Drupal. The detail environment is as follows:

- ・ Ubuntu 7.04
- ・ Apache 2.2.3
- ・ Database (MySQL 5.0.38)
- ・ PHP 5.2.1
- ・ Drupal 5.10
- ・ Gallery Developer 2.2.6
- ・ Fotonotes 5
- ・ Image module 5

We stored 173 digital collections into our digital museum system, and we use 74 metadata attributes, detail attributes are presented in Figure 2.

4.2 Prototype System based on Drupal and modules

In the second prototype system, we constructed digital museum based on Drupal 5 and Gallery2, but we have newest version of Drupal 7 in 2011. It provides the function of image management using standard add-on module of "*Views*" in Figure 5.

Figure 5: Views module in Drupal 7

Therefore, in third prototype system, we adopt "*Views*" module for TBIR (Text Based Information Retrieval) instead of Gallery2, which is presented in Table 6.

functions	Views	Gallery2
Display with thumbnails	yes	yes
Additional fields	With CCk module	yes
Drupal version	7 or 8	Under 6

Table 6: Comparison of Views and Gallery 2

However, it is not possible to create additional attribute fields using only *Views* module, then we furthermore need to implement a core module of "*CCk*" in order to create metadata scheme for pictures.

By the way, staffs in the museum are continuing to have surveys of museum collections in detail, however they are not so familiar with XML database, we provide two importing interfaces "import_museum" and "import_no_metadata", which are presented in Figure 7. A simple sketch of a workflow in digital museum is presented in Figure 8.

インポート	説明
import_museum	人類学博物館資料を取り込みます
import_no_metadata	メタデータが無い画像資料の取り込み用
Node import	Import nodes from CSV file.
User import	Import users from CSV file.

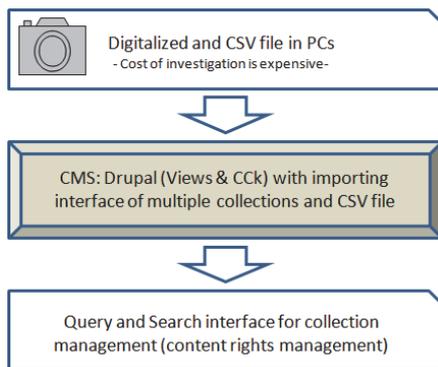Figure 7: Two importing functions with/without metadata

Figure 8: Simple workflow of contents management

Next, when administrators select "import_museum" in Figure 7, they can upload CSV formatted metadata file with detail attributes presented in Figure 2. As we stated in the previous section, we define 74 attributes in the schema of metadata in order to management museum collections.

Administrators can import CSV formatted metadata using "import_museum", and import photo data without metadata using "import_no_metadata". Then, the master file of museum collections are created using CSV format at present.

Actually, administrators can control processes in the museum using search interface, for example, museum management information exist in the database presented in Figure 9. After entering a keyword "need to repair" (in Japanese) into the query box, results of several collections including"1-0393 J-180 IIC-263" and others are displayed.

On the other hand, one of results with "not need to repair" (in Japanese) in an attribute of "Condition Information" is presented in Figure 10. After clicking "Edit" tab an editing interface is displayed in Figure 11. Attributes for contents right management such as "Acquisition Information" are also included in this interface. Finally, in 2011, we try to reconstruct digital museum by using Drupal 7.8 on Ubuntu

10.04 LTS. At present, we have 2,054 digital collections and metadata csv files. However, 1776 contents don't have enough metadata, we store 277 collections having detail attribute values in 74 fields. As a result, during several years, we define metadata for 104 collections, but it is very hard to create appropriate and accurate metadata attribute values. This kind of cost for investigation and management is usually very expensive.

Figure 9: Search for management of museum collections

5. CONCLUSION AND FUTURE RESEARCHES

In this paper, firstly we have a brief survey of various digital archive systems, including digital museums, digital libraries and web archiving systems. Secondly, we have short introduction of prototype systems of digital gallery in "the Anthropological Museum of Nanzan University". From 2005 to 2012, we have developed three different prototype systems using Joomla, Drupal and related modules.

Next, in order to organize various collections with metadata and photographs, we propose metadata attributes and design XML metadata formats and system architecture. Furthermore, we present our prototype systems using Joomla! and Drupal, which has 277 digital contents with 74 metadata attributes, for curators and research staffs in the museum. During several years, we tried to implement advanced functions of annotation function and image retrieval functions in the digital museum.

Finally, the museum collection consists of more than 40,000 ethnographic photographs including the landscape, portraits, ceremonies and others. In near future, we are planning to store various photographs and implementing the similar image search function of those photographs stored in the system.

Figure 10: Detail results including all metadata attributes

Figure 11: Edit and modification interface

REFERENCES

Canadiana, 2012. "What is the Canadiana.org Digital Collection Builder (DCB)?". http://dcb-gcn.canadiana.org/ (accessed 12 June 2012)

Doulamis, N. (et al.) 2008. Technical infrastructure of the COSMOS Portal, International Journal of Emerging Technologies in Learning (iJET), Vol. 3, pp.21-32.

Dublin Core Metadata Initiative, 2008. "Dublin Core Metadata Element Set, Version 1.1".

DuraSpace, 2012. "DSpace 1.8.2". http://www.dspace.org/ (accessed 12 June 2012)

ICOM/CIDD CRM Special Interest Group, 2008. "Definition of the CIDOC Object-oriented Conceptual Reference Model, Version 5.0". http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3. 2. rtf. (accessed 12 June 2012)

International Council of Museum, 2011. "Definition of the CIDOC Conceptual Reference Model, version 5.0.4". http://www.cidoc-crm.org/official_release_cidoc.html. (accessed 12 June 2012)

ISO/IEC JTC1/SC29/WG11, 2004. "MPEG-7 Overview". http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm (accessed 12 June 2012)

Kawano, H., 2008. Reputation-based contents crawling in web archiving system, Lecture Notes in Operations Research, World Publishing Corporation, Vol.8, pp.317-324.

Kawano, H., 2010. Management of storage devices and file formats in web archive systems, The Ninth International Symposium ISORA'10, Lecture Notes in Operations Research, World Publishing Corporation, Vol.12, pp.356-361.

Kawano, H., 2011. Hierarchical storage systems and file formats for web archiving, 21st International Conference on Systems Engineering, IEEE Computer Society, pp.217-220.

Kawano, H., 2012. Implementation of digital museum using CMS and gallery tools, 22nd International Conference on Systems Engineering, (to appear)

Microsoft, 2011. "MSXML 6.0 (Microsoft XML Core Services)". http://support.microsoft.com/kb/269238/ja. (accessed 12 June 2012)

MODS Editorial Committee, 2010. "Metadata Object Description Schema (MODS), version 3.4", The Library of Congress. http://www.loc.gov/standards/mods/ (accessed 12 June 2012)

Patel, S.K. (et al.), 2011. Performance analysis of content management systems - joomla, drupal and wordpress. International Journal of Computer Applications, 21(4), pp.39-43

W3C, 1999. "Web Content Accessibility Guidelines 1.0".

Witten, I. H. (et al.), 2002. How to Build a Digital Library, The Morgan Kaufmann Series in Multimedia Information and Systems, Morgan Kaufmann.

Yoshimura, K. S. (et al.), 2012. Technical report, A publication of OCLC Research "Social metadata for libraries, archives, and museums; part 3: Recommendations and readings". http://www.oclc.org/research/publications/library/2012/ 2012-01.pdf. (accessed 12 June 2012)