Unified knowledge-based content management for digital archives in museums

Tien-Yu Hsu
Department of Information Science, National Museum of Natural Science, Taipei, Taiwan

Hao-Ren Ke
Library and Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan, and
Wei-Pang Yang
Department of Computer & Information Science, National Chiao Tung University, Hsinchu, Taiwan

Abstract
Purpose – This paper sets out to present a new model to avoid the content silo trap, satisfy the knowledge management requirement and support the long-term perspective of developing academic, exhibition, and education applications among various domains for museums.

Design/methodology/approach – This paper presents a unified knowledge-based content management (UKCM) model, which comprises the unified knowledge content processes, multi-layer reusable knowledge content structures and an integrated knowledge-based content management system to solve the content silo trap problem. The extended entity-relationship (EER) conceptual model is applied to design a global view of the integrated knowledge system and completely represent multi-layer reusable knowledge content structures for the spectrum of various knowledge assets for all domains and applications in a museum.

Findings – A practical case of a large-scale digital archives project that includes various domains of a natural science museum has been successfully implemented to demonstrate the feasibility of the proposed model.

Originality/value – This paper integrates content management and knowledge management. Digital archives programs in museums can apply the model presented in this study to satisfy the knowledge management requirement and support the long-term perspective of developing academic, exhibition, and education applications among various domains.

Keywords Digital storage, Museums, Content management, Knowledge management, Taiwan

Paper type Research paper

1. Introduction
Numerous digital archives programs in museums are conducted worldwide to preserve and sustain mankind’s cultural heritage. Such programs aim to preserve cultural heritage and collections; popularize fine cultural landmarks; encourage information/knowledge sharing; invigorating cultural content and value-added services, and improve literacy, creativity and quality of life. They are considered as

The authors would like to thank the National Science Council, Taiwan, ROC, for financially supporting this research under Contract No. NSC93-1201-34-05-00-00-00-16.
the prerequisite and foundation for developing digital museums from which a museum’s conventional functions in academic research, exhibition, education, and entertainment can be extended and developed through information technologies (MacDonald, 2000; Alonzo, 2001; Chen, 2003; Shindo et al., 2003; Hemminger, 2004; Mei, 2004).

Like many countries around the world, many museums in Taiwan have been conducting digital archives programs. Furthermore, the National Science Council (NSC) of Taiwan sponsors the National Digital Archives Program (NDAP) (www.ndap.org.tw). The NDAP aims at promoting and coordinating content digitization and preservation at leading museums, archives, universities, research institutes, and other content holders in Taiwan (www.ndap.org.tw/index_en.php). The predecessor of NDAP is the NSC Digital Museum Program (Ke and Hwang, 2000).

Most digital archives programs in museums incur some of the following problems:

- Content and resources created by diverse individual, groups, departments, projects, or applications cannot be shared and reused among each other.
- Content cannot be accumulated and managed centrally for long term utilization and applications development.
- Digital content is managed at a data or information level for preservation, but not in a knowledge level for creation, organization, sharing and reusing.
- Systems are developed separately and repetitiously, and lack integration, leading to redundant overhead costs and spent resources.
- Users cannot access related and integrated content from a single entry point.

The content silo trap (Rockley, 2003) largely accounts for the above problems, owing to the fact that content is created by individuals working in isolation from other individuals within a project. Walls are erected among content domains and even within the same content domain. Rockley proposed the unified content strategy (UCS) to avoid the content silo trap. UCS focuses on the need for effective content acquisition, representation, organization, publication and sharing in a global view among projects, applications, users, contents and organizations in an institute. The system framework of UCS described by Rockley comprises three components: unified process, content management system (CMS), and reusable content.

The UCS was developed to fit a general enterprise needs, but it needs be extended to consider the characteristics of knowledge management and application for museums. For museums as knowledge creators, the most important tasks and challenges are to integrate tremendous and increasing knowledge content acquisition, creation, organization, reusing, and publication from all specialists, departments and projects among domains, and to develop specific applications in academic research, exhibition, and education for users. Our previous works in some digital archives projects applied CMS (Hong, 2000, 2001) to exhibition applications. Many museums have implemented CMS to manage amassed content and to enhance functionality and accessibility of a Web site (Vitto, 2003; Honeysett, 2002). These projects only handle partial participants, processes, content, technologies, domains and applications. They not only lacked the application of knowledge management techniques, but also neglected the integration and long-term prospective of content management considerations for all domains and applications from entire museum’s viewpoint.
This study extends Rockley’s UCS to fit museums’ global needs for various domains and applications through knowledge management, and proposes a knowledge-based content management system (UKCM) model. The framework of the UKCM model contains unified knowledge content processes, the multi-layer reusable knowledge content structures and the integrated knowledge-based content management system. The model aims to satisfy the knowledge management issues among various domains and academic, exhibition, and education applications in museums.

This rests of this paper is organized as follows. Section 2 describes the knowledge-based content management system framework. Sections 3, 4 and 5 describe the three components of the proposed framework – the unified knowledge content processes, the multi-layer reusable knowledge content structures and the integrated knowledge-based content management system respectively. Section 6 details the application of the extended entity-relationship (EER) modeling tool to conceptually design the integrated knowledge system across the domains of a museum, and the multi-layer reusable knowledge content structures. Section 7 demonstrates a practical implementation of our approach. In section 8 future research directions are finally pointed out.

2. System framework
Knowledge management framework in museums not only aims to manage these knowledge assets but also to manage the processes that act upon the assets. These processes include developing, preserving, using and sharing knowledge (Macintosh et al., 1999; Marwick, 2001). This study proposes a system framework (Figure 1) for supporting the unified knowledge-based content management (UKCM) model to
extend the UCS and achieve knowledge management among various domains and for long-term digital museum applications developing perspective. The three components in this framework are unified knowledge content processes, an integrated knowledge-based content management system and multi-layer reusable knowledge content structures. The three components are described briefly below and are detailed in the following sections.

The unified knowledge content process functions as a common workflow among participants and projects that includes knowledge content collection, digitization, editing, organizing, publishing and accessing stages. The multi-layer reusable knowledge content structures define the spectrum of knowledge content for all participants to follow, from core knowledge elements to advanced and innovative elements. A core knowledge element is the basis of knowledge content and comprises a multimedia object and semantic metadata. Advanced and innovative elements are further manually authored or automatically inferred from existing content. The integrated knowledge-based content management system works to integrate the entire system. This system comprises the creation subsystem for constructing vocabulary, metadata, content and the classification hierarchy; the management sub-system for managing the entire knowledge content and resources for creation and publishing, and the publishing sub-system to transfer the authored content into the publishing structure and web pages for all departments and projects.

Aside from those advantages shown by Rockley such as faster time to publish, better use of resources, reduced costs, improved quality and usability of content, increased opportunities to innovate, improved workplace satisfaction and increased customer satisfaction, additional benefits from the proposed framework for digital archives projects in museums include the following:

- Establishing a knowledge-based content creation, management and publication process to closely connect the collaboration among knowledge and information technology (IT) specialists for all projects.
- Constructing an integrated, formal, explicit domain knowledge system to unify and integrate all domain specialists under a common conceptual framework.
- Providing standard and consistent multi-layer knowledge content structures for specialists to fully express and create their spectrum of knowledge content.
- Integrating and managing knowledge content created from projects, specialists and applications centrally for others to share and to reuse.
- Improving the effectiveness of business for museums and satisfying the needs of audiences.

3. Unified knowledge content processes
In Rockley’s UCS, the unified processes eliminate the silo walls and create a collaborative environment to ensure that authors are aware of reusable content and for all departments to follow a repeatable and transparent workflow. For museums, unified knowledge content processes must be defined providing common and collaborative workflows among knowledge specialists, IT specialists and museum users. Knowledge specialists require a standard, consistent, repeatable, sharable and transparent environment to collect, digitize, edit and organize
knowledge content. IT specialists require an integrated and collaborative system development environment to design an efficient, automatic, scalable, and interoperable system for supporting content creation, management, and publication. Users require organized, categorized, integrated and systematic knowledge content via classification hierarchy browsing and metadata searching across related domains.

To satisfy the needs of knowledge specialists, IT specialists and users, all specialists must follow a standard process to create and maintain knowledge content with well-defined knowledge content structures autonomously. The unified knowledge content processes can be regarded as a life cycle that comprises knowledge content collecting, digitizing, editing, organizing, publishing, and accessing phases. The major tasks of each phase are summarized as follows:

(1) *Collecting stage:*

- to express the knowledge concepts for a particular application and user group, the variety of knowledge element must be decided in advance; and
- to collect original materials such as slides, photographs, audio, video, or documents for the target content type and prepared under a standard knowledge content structure.

(2) *Digitizing stage:*

- to digitize the original materials into digital objects using standard formats; and
- to assign a meaningful and unique object identifier for each digital object.

(3) *Editing stage:*

- to interpret each digital object with semantic metadata to make it a core knowledge element; and
- to store core knowledge elements in the core knowledge element repository for further organization and reuse.

(4) *Organizing stage:*

- to reuse and organize core knowledge elements into advanced knowledge elements, which can be a multimedia document, knowledge unit, knowledge group, and knowledge network;
- to use a multimedia document composed of a set of digital objects for interpreting a subject relating to an artifact;
- to organize a set of multimedia documents into a knowledge unit for interpreting an artifact;
- to arrange a set of knowledge units to create a knowledge group with the same characteristics for a particular exhibition topic;
- to link the above elements to other related elements to form a knowledge network; and
- to store multimedia documents, knowledge units and knowledge groups in advanced knowledge repositories with their semantic metadata to be reused and shared among specialists and applications.
(5) Publishing stage:
- to justify knowledge content before delivering to users;
- to convert each knowledge element into XML-based content structure by the system;
- to allow specialists to assign a presentation template to each knowledge element; and
- to generate web pages from combining the XML-based content and the specified presentation template.

(6) Accessing stage:
- to allow users to access knowledge content through an integrated semantic classification hierarchy-based browsing interface to share the common knowledge concepts with specialists;
- to design a metadata query interface for each knowledge repository for each domain; and
- to publish knowledge content from various domains through a unified knowledge portal.

4. Multi-layer reusable knowledge content structures
In Rockley’s UCS, the reusable content is limited in documents that can be broken down into the smallest reusable object (section, paragraph, and sentence). For a museum, the content can be defined as any type or unit of digital information, such as text, image, graphics, video and sound, or anything that will probably be organized and published across an inter-, intra- and/or extranet. Two major considerations must be made when constructing knowledge content for museums. One of these is how to create a versatile and complete structural representation; the other is how knowledge can be reused among specialists and various applications. To provide a content creation and organization model that enables the sharing and reuse of content among specialists of coexistent domains, knowledge content must be expressed in formal and consistent structures, by which the entire spectrum of knowledge content from basic to complex can be expressed and organized efficiently and completely. Therefore, this study defines a multi-layer reusable knowledge content structure to facilitate specialists to organize knowledge content from a core knowledge element into advanced or innovative knowledge elements. Core knowledge elements are created during the editing phase; advanced knowledge elements are organized in the organizing phase manually; however, the system, exploiting data mining techniques, dynamically and automatically discovers, organizes and classifies innovation knowledge elements. From the requirements of content recreation and maintenance, authorized specialists can change and reorganize any knowledge content at will. The structures of core, advanced and innovative knowledge element are described below:

- **Core knowledge element.** A core knowledge element is a basic and individual multimedia object (image, audio, video, text, animation or 3D object) associated with semantic metadata that interprets the context of a multimedia object.
- **Advanced knowledge element.** An advanced knowledge element is further organized from a set of core knowledge elements. An advanced knowledge element can be a multimedia document, a knowledge unit, a knowledge group or
a knowledge network. A multimedia document is set of core knowledge elements for describing a topic relating to an artifact. A knowledge unit possesses hierarchical structures and is employed to organize all relating topics that are organized as multimedia documents. A knowledge group is formed by a set of knowledge units having the same characteristics for presenting a research, education, or exhibition topic. The cross-relationships between any pair of the above various elements within intra domain or inter domains can be specified by specialists with links to form a knowledge network.

- **Innovative knowledge element.** A large amount of implicit and correlating knowledge exists among coexistent domains. Innovation knowledge elements can be built automatically from core and advanced knowledge elements. Innovative knowledge elements have the same structures as advanced knowledge elements. The difference is that innovative knowledge elements are created, classified, and organized automatically, but for advanced knowledge elements the process is manual.

5. Integrated knowledge-based content management system
Content management attempts to follow a given framework for effectively creating, editing, managing, and publishing content (Robertson, 2002). Knowledge can be regarded as a kind of content to be managed. Content management can be considered as an infrastructure to amass and distribute knowledge (Boiko, 2002). No single list of the best requirements exists for a content management system. Every organization has its own needs. A classification scheme comprising content creation, content management, and publication and presentation has worked well for museums. The major requirements of content creation include integrated authoring environment, separation of content and presentation, multi-layer content structure, multi-user authoring, content reusing, metadata creation and powerful cross-linking. The key requirements of content management include version control, effective indexing, manage diverse content, integration with co-existent domains, adequate security and pro-active reports. The major requirements of publication and presentation include page templates, exchange support, and effective accessibility.

Considering the unique content management and knowledge management viewpoint of museums, the knowledge-based content management system includes the content creation sub-system, the content management sub-system, and the design of the publishing and presentation sub-system. The functionalities of each sub-system are described as follows:

1. **The content creation sub-system:**
   - editing the global classification hierarchy for various domains;
   - editing core knowledge content elements with semantic metadata; and
   - organizing advanced knowledge content elements.

2. **The content management sub-system:**
   - managing indexing structures of knowledge content;
   - managing metadata repositories;
   - managing vocabulary repositories;
   - managing core and advanced knowledge content repositories;
managing a global view of the classification hierarchy for all domains; and
managing authoring template databases.

(3) The publication and presentation sub-system:
• converting knowledge elements into XML-based content structure;
• publishing web pages from XML-based content and assigned XSL templates;
• building the classification hierarchy-based browsing structure; and
• creating the metadata-based query interfaces.

6. Conceptual modeling
Conceptual modeling means identifying relevant concepts of the real world with an abstract model. Conceptual modeling intends to integrate different views of an organization into one global and consistent model where entities and relationships are explicitly defined. The conceptual modeling of a unified knowledge-based content management system for digital archives in the proposed model is designed to construct both enterprise domain knowledge systems and multi-layer reusable knowledge content structures by adopting a thorough syntax, a semantic tool, and models that concretely express and interpret them.

The ERM can be considered to be the ancestor of all modern modeling methods (Chen, 1976). Since its original inception ERM has derived many descendents aimed at enhancing the conceptual design of relational databases. Due to the popularity of the object-oriented programming concept, the object-oriented model (OOM) (Elmasri, 2000) and extended entity-relationship model (EERM) (Shoval and Frumermann, 1994) have been proposed. EERM possesses the features of ERM and OOM, including aggregation abstraction, generalization abstraction, and association abstraction. Aggregation abstraction defines a PART-OF relationship between an entity and its components. Generalization abstraction defines a subset or IS-A relationship between entities, and establishes a hierarchy from a generic entity to its subsets. Association abstraction defines a multi-valued feature of attributes. Due to the many advantages of EERM, it has been applied in the conceptual modeling of database applications (Batini, 1992; Engles et al., 1993). Huang and Hsu (1999) also successfully applied EERM to conceptually model the multimedia databases of museum applications. This study also uses EERM as a conceptual modeling tool in this paper.

6.1 Conceptual modeling of enterprise domain knowledge system
The conceptual modeling of a knowledge system across domains and applications in a museum can be summarized as having the following features:

• A top-down approach is used to construct the global knowledge system across domains.
• The relationships and constraints can be constructed between entities within a domain or across domains.
• The attributes of knowledge content can be specified as metadata annotated by specialists of each domain.
• To support efficient administration and personalized services, profiles of specialists and users must be specified.
After completing the system conceptual design, the knowledge hierarchy and relationships in a domain or between domains can be organized to construct the global knowledge system. The knowledge system can be constructed for application requirements and can also be viewed as the knowledge classification hierarchy system (Guarino, 1995) to represent them. The knowledge classification hierarchy system plays the common view among specialists and users. The knowledge element entities contain a set of core and advanced knowledge elements specified in section 4. The next section reveals their conceptual modeling process.

6.2 Conceptual modeling of multi-layer reusable knowledge content structures
The multi-layer reusable knowledge content structures shown in section 4 provides a set of formal structures from elementary to complex for specialists to express and organize knowledge content for particular concepts. The knowledge element entity denotes the superclass of all knowledge content, and comprises the core knowledge elements, the advanced knowledge elements and the innovative elements.

The core entity denotes the class of core knowledge elements, each of which is a set of multimedia objects with semantic metadata. The multimedia object entity encompasses the entities of image, audio, video, text and animation. An instance of the advanced entity comprises a set of instances in the core entity. The advanced entity comprises of the multimedia document sub-class, the knowledge unit sub-class and the knowledge group sub-class. The instance of the latter subclass is organized from a set of instances of the former sub-class. Section 4 details the usage of the above three sub-classes. Any instance in the core entity, advanced entity, or innovative entity may have a set of related link to instances of itself or to others to form a set of knowledge networks. The innovative entity with the same types of instances is inferred from the core and advanced entity.

7. A practical implementation
The unified knowledge content management approach for a digital archives project has been implemented in the National Museum of Natural Science (NMNS) in Taiwan (www.ndap.org.tw/index_en.php). NMNS is one of the nine primary participants of the Taiwan NDAP introduced in Section 1. A total of fifteen domains in zoology, botany, geology and anthropology participate in the digital archives project of NMNS. All domains are coordinated and integrated by the information technology integration project to achieve unified processes, content structures, and knowledge-based management system development.

The unified process shown in section 3, including the collecting, digitizing, editing, organizing, publishing, and accessing phases has been specified through discussions. All specialists in each domain must follow the standards and specifications of each stage. A single and integrated knowledge-based content management system (KCMS) was developed, by which the content creation, management, and publication described in section 5 was fulfilled collaboratively among all content specialists and IT specialists. A multi-layer reusable knowledge content structures including core, advanced, and innovative knowledge elements defined in section 4 was designed. These structures are managed and maintained by KCMS. All specialists applied them to edit, organize, and maintain content under a standard and consistent process. All the finished and verified content created by specialists was converted into XML-based
content structures for publication. All knowledge content constructed under a single
global classification hierarchy-based system and the interchange formats among
institutes was also converted into XML-based structures. The XML-based content with
assigned XSL templates was transformed into web pages during accessing. Users
could access content through the integrated knowledge portal through a classification
hierarchy-based browsing and metadata search query interface. Figure 2 shows the
entire system architecture, while Figure 3 shows an example of knowledge unit for
interpreting the knowledge content of a species of insect.

8. Conclusions and future work
This study presents a unified knowledge-based content management (UKCM) model to
avoid the content silo trap, satisfy the knowledge management requirement and
support the long-term perspective of developing academic, exhibition, and education
applications among various domains for museums. The proposed model improves on
previous works and other museums’ web sites on content management, and extends to
the Rockley’s UCS (Unify content strategy) and knowledge management. The
framework of UKCM model comprises a common knowledge content process, an
integrated knowledge-based content management system and multi-layer reusable
knowledge content structures. The extended entity-relationship (EER) conceptual
model was applied to design a global view of the integrated knowledge system and
fully represent multi-layer reusable knowledge content structures for the spectrum of
knowledge assets among domains and applications in a museum. A practical case of a
large-scale digital archives project that includes various domains of a natural science
museum has been successfully implemented to demonstrate the feasibility of the
proposed model.
Our research on knowledge discovery, classification, and organization is underway. Currently, a classification hierarchy is exploited; in the future, it will be developed into a fully-fledged ontology. The amassed innovative implicit knowledge existing in core and advanced knowledge elements will be generated and inferred inside the same domain or among related domains automatically. The ontology will be dynamically extended, maintained, and mapped between specialists and users during content creation. Finally, the intelligent personalization services and web mining mechanism will be developed for providing an adapted and on-demand knowledge content communication environment between specialists and users.

References


Further reading

About the authors
Tien-Yu Hsu is an Associate Research Fellow in the Department of Information Science, National Museum of Natural Science (NMNS), and is also a PhD candidate in the Department of Computer and Information Science, National Chiao Tung University (NCTU). His research interests include digital library, digital museum, multimedia database, mobile learning and knowledge management.

Hao-Ren Ke is a Professor of the Library, and Institute of Information Management, National Chiao Tung University. He is also the Associate Director of the NCTU Library. His research interests include digital library, digital museum, web service, and data mining. Hao-Ren Ke is the corresponding author and can be contacted at: claven@lib.nctu.edu.tw

Wei-Pang Yang is a Professor of the Department of Computer and Information Science, National Chiao Tung University. He is also a Professor of the Department of Information Management, National Dong Hwa University. His research interests include database theory, information retrieval, data mining, and digital library/museum.