_____

METADATA POLICIES FOR THE DESCRIPTION OF DIGITAL FOLKLORE COLLECTIONS

Irene Lourdi
Libraries Computer Centre
National & Kapodistrian University
Campus University, Ilisia
Athens
Greece
elourdi@lib.uoa.gr
www.lib.uoa.gr

Christos Papatheodorou
Department of Archive and Library Sciences
Ionian University
Palea Anaktora, Eleftherias Sq.
Corfu
Greece
papatheodor@ionio.gr
www.ionio.gr/~papatheodor

Abstract
Information retrieval from heterogeneous resources is quite difficult and the requirement to follow a common policy permitting cross-domain searching in a variety is apparent. Since each holding institution (museum, library, archives) follows different material administration and metadata generation policies, it is obvious that information integration confronts several difficulties. This paper proposes a set of guidelines for the design of a metadata model for digital folklore collections. It focuses on the combination of the description policies from the museum and library perspective so that to fulfil both community´s requirements without losing information about the collection nature and unique characteristics.

1. THE INFORMATION LANDSCAPE

The past few years there has been an increasing demand for research and study of cultural heritage material. Many digitization projects have taken place making cultural content available to a wide audience and offering high quality services to users. Digital cultural content is a broad concept that includes multimedia surrogates for cultural material owned by the world's museums, libraries and archives, and their associated descriptive and contextual information (Gill et al. 2002). Besides material preservation, another goal of these projects has been to bring people from various places close to tradition and contribute to their education by helping them to search and find information about their cultural heritage. In a high degree, digitization projects accomplish their institutional targets for information discovery, material preservation and education, but the information landscape has been changed recently.

Citizens navigating various resources desire to find intellectual and cultural materials easily, without concerning about institutional or national boundaries (Dempsey, 1999). Users that need to study a specific subject, like "the development of agriculture in a country" prefer to collect data from all the institutions and sources that are holders of this kind of information. So nowadays, memory institutions like archives, museums and libraries are obliged to plan the adequate strategies to manage their cultural resources in a unified manner, ensuring that users will follow unified routes to their goals. The need for cross-domain searching is obvious and in front of this challenge cultural heritage institutions are exploring possible solutions.

_____

In this paper we deal with the arising problems of how libraries and museums are capable of managing their documents and artifacts to generate integrated access to them through electronic services and applications. Our study focuses mainly on how to describe digital folklore collections from the library and museum perspective preserving simultaneously the authenticity and integrity of folklore material. In particular, we present the main administrative issues referring to a folklore collection that belongs to a library but is exposed as a museum collection and we propose policies and tools to confront them theoretically and practically, by comparing the libraries and museums collection management policies.

The rest of the paper is organized as follows: in section 2 we define the problem and we discuss about other works that have dealt with similar matters. In section 3 we present as a case study a folklore collection that belongs to a library and contains museum objects and we compare the existing policies of cultural content description in library and museum communities. Further, we present the requirements for describing compound digital folklore collections both from museum and library perspective, without losing its nature and unique characteristics. In section 4 we refer to metadata standards that are used for cultural heritage material and in section 5 we investigate ways for achieving metadata interoperability, while in the last one we drew to conclusions and we make some suggestions.

2. RELATED WORK

The great need for libraries and museums to offer integrated access to qualitative cultural data, demands solutions able to satisfy various and sometimes different information seeking behaviors. However user requirements have a common line even the nature of material management traditionally differs across libraries and museums. Library users tend to seek and find information that will support their education and research needs, while museum users like to visit exhibitions not only for educational purposes but also for their own entertainment.

Moreover, memory organizations handle various kinds of material, for example a library may contain museum objects, or a museum may have archival material or may have a library of documents and books. This phenomenon imposes to plan a common strategy and material administration policy without worrying if the institution is a library or museum. In general all categories of users desire to refer to cultural material flexibly and to be able to find information related to a specific subject either from library or museum collections.

The impetus to make interoperable cultural content and accessible via Internet, has provoked many institutions to propose practices for content creators and to explore what it means to develop cultural institutions in a digital environment (Dempsey, 1999). There have been many projects and studies dealing with methods to integrate collection information from different descriptions. One project is called MOSC (Roel, 2005) and examines how the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) can bridge metadata standards and practices across museums, archives, and libraries. Particularly, it states that OAI can function with any metadata format expressed by Extensible Markup Language (XML). Another similar project is MOAC (Rinehart, 2003), which it highly define that "cultural heritage community needs to continue research, testbeds, and production in the area of aggregating content

_____

for the end-user". Also COVAX "examines and creates the technical solutions required to provide access through the Internet to homogeneously-encoded document descriptions of archive, library, and museum collections based in the application of XML" (Hernandez et al. 2003).

In combination with these projects there are also a number of related studies and papers that deal with the integration of services from several cultural heritage institutions. Jane Barton (2005) explores several issues of virtual museums in comparison to digital libraries and concludes that they have much in common. In particular, she describes how important is to collaborate museums and libraries communities in order to have cross-domain searching and to *facilitate new knowledge creation*. In another study P. Edwards (2004) proposes guidelines for collaboration between memories institutions and suggests that focus must be given on institutional relationships and on the social role of each institution.

Most of the papers and projects accept the fact that collaboration between cultural institutions is important but little attention has been given to extract guidelines, practical specs, and models that will enable the individual museum or library to distribute content to users easily and cost-effectively (Rinehart, 2003). The problems to this collaboration are quite many but the main issue is that museums and libraries have developed their own content standards to describe their holding material (Gill, 2004). One real concern is the context of the original artifact or digital resource. The value of information about the objects that are in museums and archives is often related to the context and the provenance of them. Further, it is a matter of having interoperability both in syntactic and semantic level between the metadata models of each institution. The term syntactic interoperability implies to have metadata standards with common syntax, structure, format, while semantic interoperability investigates which elements of different metadata schemes express the same semantics, i.e. it means that elements from different metadata standards are able to "communicate" even if one says *author* and the other *creator*.

Our research will focus mainly on issues that should be taken into consideration from libraries and museums for the development of their own metadata model to describe a digital folklore collection. The great expectation is to apply models that will take into account most of the user categories, will preserve the specialties of each collection and simultaneously will be interoperable with others. The goal of this study is to propose guidelines concerning the metadata interoperability among different memory institutions.

3. COLLECTION MANAGEMENT POLICIES

Many harmonization attempts have shown that virtual museums and digital libraries have common targets and boundaries between them are quite blurred (Barton, 2005). However, any approach must take into account both the differences and similarities of two communities, since these differences are quite essential and defining. The current study is provoked by the folklore collection of University of Athens, Greece. The collection consists of heterogeneous material such as documents, written notebooks, cultural objects (jewels, dressings, tools etc). The noteworthy characteristic of the collection is that these 3d objects are located and exposed in the library of Folklore Department and so the library has to adapt its role to the museum context.

_____

_____

3.1 Comparison of library and museum collection management policies

- *Material:* libraries traditionally tend to collect items that have been produced in multiple copies for wide usage, while museums contain mostly unique objects. Both institutions may contain heterogeneous material and both can be a part of the other.

- *Description policy:* in library community "collections" do not stand as an entity since description tends to be more for bibliographic units (items). In an OPAC items are described one by one and material documentation is acceptable even if a basic level of cataloguing is applied. On the other hand, within museum community the concept of "collections" has many interpretations according to several criteria. For example objects can be considered as a separate collection if they have common subject or medium or if they refer to the same historical period or they consist works of a particular artist. From the administrative point, objects can be separated in collections if they are of a particular donor or they have been created for a specific purpose. So in museums the metadata for collection description are quite important than in library items. In museum material the collection-level metadata is a medium to access the item-level resources.

- *Structure:* even though, museums organize their material in collections the concept of structure is not so important since these collections are mostly flat and hierarchy does not exist like in archives. The same stands for library material with some exceptions when referring to a series or serials, in which the volumes or parts follow a strict hierarchy and the librarian is called to show the relation between them.

- *Access:* library's collection is to be used widely and free without any restrictions, except in cases that they hold old and rare material that is quite vulnerable to any physical touch. The place allows users to locate and browse the items on their own, in an autonomous manner. On the other hand, items in museum are surrounded by a small amount of text that can be supplemented by a guide and/or a printed catalog. Museum place with the appropriate lighting and signage manage to reinforce the user in a particular flow to see the collection. Further, even though museum material was not allowed in the past to be accessed directly, this policy has changed. Nowadays, museums are open to all users and fund enough money to digitize their holdings in order to make them widely available.

- *Preservation:* examining the motivating forces of digitizing cultural objects both communities wish to enhance our understanding of the past and promote the educational challenges for all kinds of users (Edwards, 2004). Another motivation for digitization projects is also the preservation of resources. This aspect stands mostly for libraries, where digitization is considered to be a solution for preserving rare and vulnerable material in the future. On the contrary in museum community, digitization is not always a substitute for preservation, since other methods can be also applied for it such as conservation, closed access material etc.

3.2 Designing digital collections

Libraries and museums that want to be interoperable ought to adjust their management policies according to user needs and nature of material. In our case that we want to combine the aspects of museum and library collections, any effort must be applied within the scope of preserving and making available to users the authentic characteristics of each object without ignoring any usage restrictions. It would be quite useful to emphasize some issues that could contribute to cross-domain searching and could help to describe folklore objects from library and museum perspective:

_____

Collection-level description
First of all, since cultural objects are heterogeneous it is proposed to separate them in collections according to criteria that would either help users to find them easily or would satisfy a specific purpose. Collection-level description allows users to search for information across library and museum domains, enabling them to identify appropriate collections to visit or item-level searching. Especially for folklore objects, we believe that the best approach is to separate them in groups according to time or place of provenance. Time and place dimension is quite important for cultural heritage collections for scientific and administrative reasons, as it is described in (Lilis et al. 2005). By representing the knowledge in a structural and systematic way, users are able to find and learn about their subject with relative ease.

Metadata requirements
The metadata model ought to contain features that will provide users with effective services to access and retrieve data about the objects either by browsing the collections or by searching in the contents of them using keywords. Metadata authoring practices must be compliant with community accepted standard schema(s). So first of all it has to be focused both on collection entity and items. Also the librarians' understanding of the descriptive and analytical needs of three-dimensional objects has to be expanded. The descriptive metadata needs to cover not only the subject coverage of the items (like in libraries) but also to (a) express the relationships that exist between the digital objects, (b) contain elements concerning the digitization process and the digital surrogates of the items (technical metadata) (Patel et al. 2005) and to (c) inform about the access rights and usage restrictions to protect tradition (administrative metadata).

Further, since the composition of collections changes constantly as materials may flow in and out of the collecting organization, another important point for libraries is to provide information about the provenance of the objects, where it comes from and how it came to the library. Users will not use provenance as an access point, but for administrative reasons the library is important to know how the objects were acquired and how they are related with the its own context.

Standards and protocols
Each organization that deals with digitization projects is proposed to implement description and preservation standards that are widely used from the international cultural community. Compatibility to general practices and protocols contributes to the effective resource discovery and to the maintenance of a high level of consistency. A source that can be used as reference is the "Descriptive Metadata Guidelines" for RLG Cultural Materials which negotiate the different practices in the communities by establishing a guide to description that can be applied in any case (RLG, 2005).

Digitization process – purpose
Digitization needs of the different institutions have to be considered. For example to produce digital images of a museum object, which would be used for scholarly research or commercial publishing, require different digitization standards and skills than the process of scanning photographs for a library collection. The characteristics of digitizing material shall be recorded by the appropriate metadata elements that will

_____

preserve for administrative reasons the whole process of each object digitization and the technical fields for their digital surrogates.

4. METADATA STANDARDS

Due to the diversity of museum and library management perspectives, it is impossible a single descriptive schema to be created meeting almost all the communities needs (Gill, 2004). Therefore, a plethora of general standards and local community-specific metadata models have evolved for the documentation of cultural collections. So parallel technologies for description have been developed – meaning that they employ different data structures, data content rules and (to some extent) data formats to encode their collections. It would not be possible to describe here all the existing metadata standards but we will mention the most widely accepted and used ones in cultural heritage domain.

4.1 Libraries

MARC (Machine Readable Cataloguing)**:** it is undoubtedly the oldest metadata format in existence. Originally developed to structure the attributes of bibliographic units and it has been modified over the years to provide a rich and flexible structure for the description of a wide range of information types (also for digital resources - 856 fields). It is the standard bibliographic format for libraries around the world.

MODS (Metadata Object Description Schema)**:** it is bibliographic element set expressed as an XML schema. It contains a subset of MARC 21 fields intended also to enable the creation of original resource description records.

METS (Metadata Encoding and Transmission Standard)**:** it is expressed as an XML schema aiming to encode descriptive, administrative, and structural metadata for a digital collection. Specifically it provides a vocabulary and syntax for identifying the digital pieces that together comprise a digital entity, for specifying the location of these pieces, and for expressing the relationships between these digital pieces. It permits the co-existence of various metadata standards offering a mean for the exchange of digital library objects between repositories and enabling the association of a digital object with behaviors or services.

4.2 Museums

CDWA (Categories for the Description of Works of Art): are guidelines for formulating the content of art databases. They articulate a hierarchical structure for descriptions of objects and images about works of art.

VRA (Visual Resources Association): it is applied to create records to describe works of visual culture as well as the surrogate images that document them. So it can be used as many times as it is needed to describe all the surrogates of a work. The core suggests which data elements are required in order to describe an item in a visual resources collection in a shared environment.

SPECTRUM: it represents a common understanding of good practice for museum documentation, established in partnership with the museum community. It contains data for documenting objects and identifies the "Units of Information" that need to be recorded to support the procedures that these objects undergo.

4.3 General metadata standards

_____

_____

Further, it would be right to mention that there have been some attempts to help effectively the cultural (not only) resource discovery and description without concerning about the type of holding institution. The most widely used standards are:

DCMI (Dublin Core Metadata Initiative): the metadata set contains 15 descriptive elements used to provide a simple means of describing resources and to aid more effectively information discovery and retrieval. DC is considered to be the minimal data set that would satisfy a description model.

RSLP (Research Support Libraries Program): this description model is focused on collection-level entity, while acknowledging that *collections* may mean different things in the different library, archival and other content models. So it has been used to facilitate cross-domain working and searching collections serving each collection management purposes. The model contains descriptive information about the collection, information about how to access the collection, including physical access, in the case of library or museum and the terms and conditions associated with access to the collection and individual items within it.

CIDOC (International Committee for Documentation): it represents a domain specific "ontology" for cultural heritage information. It describes the concepts and relations relevant to the documentation of cultural heritage in a formal language. It has been developed by the ICOM/CIDOC Documentation Standards Group to serve as a basis for mediation of cultural heritage information.

5. METADATA INTEROPERABILITY

Since memory institutions handle all types of material, the need of information integration is quite apparent. The goals of library and museum communities to provide *wide usefulness*, *portability* (across networks, systems, organizations) and *longevity* of digital cultural resources are encapsulated by having *interoperability* between the metadata standards and models they use (Gill, 2002). In recent years, numerous projects have been undertaken in the information community to achieve interoperability among different metadata schemas. Some of the mechanisms that have addressed to metadata interoperability are the following: *Crosswalks/ mappings, Application profiles/modification, Metadata framework/container, Protocols.*

Crosswalks – Mappings: crosswalks sometimes are called "mappings". They are used to "translate" between different metadata element sets. The fields in one metadata set are correlated with the fields of another that have the same or similar meanings. This process is also sometimes called "semantic mapping". In NISO white paper (1998) crosswalk is differentiated from mapping, by saying that a fully specified crosswalk must also provide rules for element to element mappings, hierarchy and object resolution, and metadata content conversions. Many papers have written about the challenging issues during making crosswalks between standards but they are still considered to be a widely applied solution. There have been a substantial number of crosswalks like: MARC21 to Dublin Core, VRA to Dublin Core, EAD to ISAD etc. (MIT).

Application profiles: the concept of application profiles is based on the idea that in the heterogeneous information environment each community has different characteristics; consequently metadata standards are necessarily localised and optimized for specific requirements. To accommodate individual needs, often it is required to mix elements

_____

_____

from various metadata schemas and to customize them according to local requirements. Also an application profile can be an existing schema used as the basis for description in a particular digital system that is enriched with other elements for application by a particular interest group or user community. An application profile has been proposed in (Lourdi and Papatheodorou, 2004) to describe the Folklore collection of University of Athens.

Metadata framework/container: a metadata framework is used as a container within which fields from multiple metadata schemas can be accommodated. Two prominent examples are: (a) **RDF** (Resource Description Framework), it is a model developed for the description of resources on the Web. It provides a mechanism for integrating multiple metadata schemes and expressing their vocabularies and semantics and is considered to be the foundation of semantic web (b) **METS,** it is a standard that provides a method to encapsulate all the information about an object (digital or not).

Protocols: the most public protocol for data share between various digital services is the OAI-PMH that supports standardized exchange of metadata describing items in disparate collections such as those held by museums and libraries. The basic feature of OAI is that it can be used not only with DC but also with every metadata schema that is in XML format.

6. SUGGESTIONS – CONCLUSION

The value of a usable search system that represents texts, manuscripts, images, digital objects, and artefacts simultaneously is a challenge that has to be overcome from all memory institutions, if they want to provide cross-collection repositories. Current user needs for access to all available information across the board of digital libraries and virtual museums, address information professionals to give priority to the task of creating the highest feasible level of interoperability among metadata models. Maintaining interoperable collection-level and object-level metadata enhance the facility to aggregate and disaggregate content from multiple heterogeneous resources, without thinking the type of holding institutions.

Many studies and projects have shown that the traditional compromise for providing access across heterogeneous information sources is to map everything to a core schema. These simple descriptions — sometimes called *resource discovery metadata* — help the researcher discover and evaluate resources. Also the last years it is quite common to generate mappings to specific-domain ontology, since the mapping of metadata standards to ontology contributes essentially to overcome the difficult part of semantic interoperability (Tudhope and Binding, 2004). Specifically in cultural heritage domain, it is proposed to map metadata models to CIDOC, since it is an ontology that refers to cultural content and can be applied both from libraries and museums (Doerr, 2003; Doerr et al. 2003).

In general, a policy for achieving cross-collection repositories must follow some principles that integrate the special characteristics of both types of institutions: (a) preserve the unique context of collections and items (b) metadata authoring practices must be compliant with community accepted standard schema(s) (c) techniques for presenting and combining item-level and collection-level records need to be developed (d) all the types of metadata (descriptive, technical, administrative) should

_____

_____

be captured either for physical and digital items and (e) maintain the role and identity of each institution and develop practices that aid education and "rich" information discovery. There is no doubt that museums and libraries have much in common and so any collaboration between them appears to be essential for the whole scientific and research community.

7. BIBLIOGRAPHY

Barton, J. 2005. Digital libraries, virtual museums: same difference? *Library Review* [online] 54(3), pp. 149-154. Available from: http://www.emeraldinsight.com/Insight/ [cited 26 April 2006].

Categories for the Description of Works of Art (CDWA) [internet]. (Updated 27 June 2005). Available from: http://www.getty.edu/research/institute/standards/cdwa/ [cited 26 April 2006].

Dempsey, L. 1999. Scientific, Industrial, and Cultural Heritage: a shared approach. *Ariadne,* Dec, 22. Available from: http://www.ariadne.ac.uk/ [cited 26 April 2006].

Doerr, M. 2003. The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* [online]. 24(3) [cited 26 April 2006].

Doerr, M. et al. 2003. Towards a Core Ontology for Information Integration. *Journal of Digital Information* [online]. April, 4(1) [cited 26 April 2006].

Dublin Core Metadata Initiative (DCMI) [online]. (Updated 10 April 2006). Available from: http://dublincore.org/ [cited 26 April 2006].

Edwards, M. P. 2004 Collection Development and maintenance across libraries, archives and museums: A novel collaborative approach. *Library Resources and* Technical Services, Jan, 48 (1), pp. 26-33.

Gill, T. et al. 2002. Re-inventing the Wheel? Standards, Interoperability and Digital Cultural Content. *D-Lib Magazine* [online]. Jan., 8(1). Available from: http://www.dlib.org/dlib/january02/01contents.html [cited 26 April 2006].

Gill, T. 2004. Building semantic bridges between museums, libraries and archives: the CIDOC Conceptual Reference Model. *First Monday* [online]. 3 May, 9(5). Available from: http://www.firstmonday.org/ [cited 26 April 2006].

Hernandez, F. et al. 2003. XML for libraries, archives, and museums: The project COVAX. *Applied Artificial Intelligence* [online]. Sep. 17(8-9), pp. 797-816**.** Available from: http://taylorandfrancis.metapress.com [cited 26 April 2006].

ICOM. *The CIDOC Conceptual Reference Model* [online]. Available from: http://cidoc.ics.forth.gr/ [cited 26 April 2006].

Library of Congress. 2005. Machine Readable Cataloging 21 (MARC 21) [online]. (Updated 15 Sept 2005) Available from: http://www.loc.gov/marc/ [cited 26 April 2006].

Library of Congress. 2006. Metadata Object Description Schema (MODS) [online]. (Updated 5 April 2006) Available from: http://www.loc.gov/standards/mods/ [cited 27 April 2006].

_____

Library of Congress. 2004. Metadata Encoding and Transmission Standard (METS) [online] (Updated 8 April 2006). Available from: http://www.loc.gov/standards/mets/ [cited 27 April 2006].

Lilis, P. et al. (2005), A metadata model for representing time-dependent information in cultural collections, MTSR, First online metadata and semantics research conference, Conference Proceedings, Nov. 21- 30, 2005.

Lourdi, I., Papatheodorou, C. (2004), A metadata application profile for collection-level description of digital folklore resources, IEEE Computer Society, PEH 2004: Proceedings of 3rd International Workshop on Presenting and Exploring Heritage on the Web, August, 2004.

MIT Libraries. Metadata mappings [online]. (Updated 2 June 2004) Available from: http://libraries.mit.edu/guides/subjects/metadata/mappings.html [cited 27 April 2006].

NISO. 1998. Issues in Crosswalking Content Metadata Standards. [online] Available from: http://www.niso.org/press/whitepapers/crsswalk.html [cited 27 April 2006].

Patel, M. et al. 2005. Metadata requirements for digital museum environments. International Journal on Digital Libraries [online], May, 5 (3), pp. 179-192**.** Available from: http://www.springerlink.com [cited 26 April 2006].

Research Libraries Group. 2005. Descriptive metadata guidelines for RLG cultural materials [online] RLG (Published 2005) Available from: http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf [cited 26 April 2006].

Research Support Libraries Program (RSLP). Collection Description Schema [online] (Updated 15 Oct. 2001) Available from: http://www.ukoln.ac.uk/metadata/rslp/ [cited 7 April 2004].

Rinehart, R. 2003. MOAC - A report on integrating museum and archive access in the online archive of California. *D-Lib Magazine* [online], Jan., 9(1). Available from: http://www.dlib.org/dlib/january03/rinehart/01rinehart.html [cited 26 April 2006].

Roel, E. The MOSC project: Using the OAI-PMH to bridge metadata cultural differences across museums, archives, and libraries. *Information Technology and Libraries* [online]. March, 24(1), pp. 22-24 [cited 26 April 2006].

MDA. SPECTRUM Units of Information [online]. Available from: http://www.mda.org.uk/spectrum.htm/ [cited 26 April 2006].

Tudhope, D.,Binding, C. 2004. A Case Study of a Faceted Approach to Knowledge Organisation and Retrieval in the Cultural Heritage Sector. *Digicult,* Thematic Issue 6 - Resource Discovery Technologies for the Heritage Sector pp. 28-33.

The Open Archives Initiative Protocol for Metadata Harvesting [online]. Available from: http://www.openarchives.org/ [cited 7 April 2006].

VRA Core Categories [online] (Updated 20 February 2002) Available from: http://www.vraweb.org/vracore3.htm/ [cited 26 April 2006].

W3C. *Resource Description Framework* (RDF) [online] Available from: www.w3.org/RDF/ [cited 26 April 2006].