

Guidelines for Developing Digital Cultural Collections

Eirini Lourdi
Libraries Computer Centre, University of Athens
Panepistimiopolis, Ilisia, 15784
Athens, GREECE
elourdi@lib.uoa.gr

Mara Nikolaidou
Harokopio University of Athens
70 El. Venizelou St, Kallithea, 17671
Athens, GREECE
mara@di.uoa.gr

Topic: Digital Libraries: Case Studies and Applications

Keywords: Folklore collections, digital libraries, metadata interoperability

1. Introduction

Cultural heritage, especially in the digital era, is the cornerstone of the unique identity of a society or a nation, since cultural heritage sources enable people to anchor themselves in a historical period, create a reference framework for their way of living and provide a value base for their lives. Thus, cultural heritage can be perceived as a multidimensional and multiform representation of a nation or a society over time and place. In view of cultural material importance, the last few years many digitization projects related to cultural collections have taken place, trying to expose their material to a wider audience through the Web. Most of these digitization efforts are following guidelines published by international organizations and are based on general practices and experiences gained from similar projects.

In the following, a guidelines are discussed considering the requirements and issues concerning planning and organizing a digital cultural collection in order to be functional for all user categories and easily retrievable through the Web. The issues focused are: i) material organization and management policy ii) metadata standards and iii) data preservation practices. The great need for cultural institutions (library, museum or archives) to offer wide and easy access to qualitative cultural data demands to approach all the issues above from an integrated aspect without concerning about institutional or national boundaries. Proposals need to satisfy various and sometimes different information seeking behaviours by always keeping the authenticity and integrity of cultural material.

2. Material organization and management policy

Cultural heritage institutions should adjust their digital collection planning according to user needs and nature of material. In most cases, cultural collections contain heterogeneous material with unique characteristics, like written texts, photographs, physical objects, sound recordings, maps or even born-digital material. Consequently, it is justifiable to have collections and sub-collections with complex structure and rich semantics.

Since cultural objects are heterogeneous, they should be grouped according to criteria either facilitating users easy access or expressing internal standards following by the holding institution (owner of the collection). As such, criteria can be for instance the topic coverage, the specific usage or purpose that each resource has in the context of the collection, the provenance, the type of material or the geographic region and historical period the object covers. It is useful to create sub-collections (if they do not pre-exist), since it is easier to represent composite structures and accredit rich semantics to any level. By defining specific groups of objects, the whole collection and sub-collections can easily be manipulated as separate objects

with their characteristics and metadata elements like all the other digital items, the attributes inherited from the collection to sub-collections are identified and the overall collection can be effectively navigated by users (Lourdi, 2005).

In case cultural heritage collection consists of composite objects like texts with photographs or traditional dressings, it is a good practice to separate them into their parts and represent their structure. Composite objects must be de-composed into their disparate parts since it is possible to characterize them individually with the appropriate metadata elements, as it is proposed above.

Besides material organization, another crucial issue for a cultural collection is the management policy and more specifically the administration policy. In a digitization project, it is important to preserve the holding institution policy concerning the access restrictions and material protection rules. For instance, whether it is about unique and rare material with copyright issues or the purposes of the institution is not to provide the material freely to audience, digital collection plan shall respect these matters and shall not give full access to collection and its contents.

The most prevailing factor for a digitization project is material preservation. Most cultural heritage institutions consider digitization as the best solution for preserving rare and vulnerable material in the future. So for administration purposes, a digital collection shall follow the most accepted digitization practices and shall preserve all the information related with digitization process and devices.

The main requirements for cultural information management and dissemination process can be summarized the following: a) the material should preserve its unique characteristics, b) the semantic context of the collection and of its objects should be expressed clearly and c) cultural data should be retrievable at any structured information level. It is obvious that for manipulating and representing cultural information objects, which are characterized by complexity and intense heterogeneity, rich and flexible metadata models are required, capable of encapsulating and manipulating, contextual information.

3. Metadata model/ standards

In the current digital environment, users are called to find information among various resources without easily being able to criticize and define whether it is of their interest. For this reason the basic parameter in a digitization project is to characterize meaningfully the material with the metadata model. For this purpose, many metadata standards have been developed either by international committees or by local projects describing cultural heritage collections and providing semantically rich information to users.

Metadata model requirements

The metadata model for a cultural collection 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 those using keywords. Metadata authoring practices must be compliant with community accepted standard schema(s).

The metadata model needs to be focused both on collection entity and item level. Collection-level description allows users to search for information across all kinds of memory institutions and domains, enabling them to identify appropriate collections to visit or item-level searching. High-level collection description is important in order to help the navigation, discovery and selection of cultural content [Dempsey, L. 1999]. The collection-level description simplifies the retrieval of information because the user can decide whether the collection is of his interest without getting into details about the objects and also contributes to better administration of large collections. Thus, it is required to offer a detailed collection and sub-collection level description with the appropriate metadata elements, after specifying the structure of the collection.

Librarians' understanding of the descriptive and analytical needs of three-dimensional objects has to be expanded. The descriptive metadata needs not only to support the subject coverage of the items but also to:

(a) express the relationships between digital objects. It is necessary to represent all kinds of relations that exist inside and outside the collection throughout structural levels, in order to provide the users with all the

information that is hidden in the collection. For example, the relation between a photograph referenced at text page and the actual photograph belonging in the photographs sub-collection - probably in another format - should be identified (for example: “has format” or “is converted to” or “is the same with”),

(b) contain elements concerning the digitization process and the digital surrogates of the items (technical metadata). This is considered important mainly for preservation reasons and for protecting digital sources from future risks (data transfer, operational system changes, physical destructions etc.) (Patel et al. 2005) and

(c) inform about the copyrights and material usage restrictions (administrative metadata). This is a strong necessity in order to keep all the valuable information for preservation, authenticity and retrieval of information.

In general, the main metadata principles for a cultural collection are: a) to be appropriate for the described material, users, and future uses, b) to support interoperability, standard, controlled vocabulary employed to reflect content, clear statement of conditions/terms of use, c) support the long term management of objects in the collection (Distinctions: Administrative, Technical, and Preservation Metadata) (NISO, 2004), d) metadata records should be considered as objects themselves and so they should have the qualities of good objects and should be authoritative and verifiable.

Metadata standards / interoperability matters

Information retrieval from heterogeneous resources is quite difficult since each memory institution (museum, library, archives) follows different material administration and metadata generation policies. Due to the diversity of museum, archives and library management perspectives, it is impossible to create a single metadata schema 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 is not possible to describe all the existing metadata standards for cultural heritage collections but a general statement about them is documented in next paragraphs and emphasis is given in matters of interoperability.

The existing metadata standards for cultural collections and items can be separated in two categories: according to the coverage domain and according to the description level. For instance concerning the domain, there are metadata standards that are used by libraries (MARC and MODS), others by museum collections (CDWA VRA, SPECTRUM) and those by archives (EAD). Besides them, there also exist some metadata standards for general use like Dublin Core and RSLP. On the other hand, concerning the description level, some of the above standards cover item-level description while others cover collection-level description (RSLP, Dublin Core Collection Description Application Profile).

At this point, no one can claim which metadata schema is the most appropriate to apply. Nevertheless, the best practice for designing a “good” metadata model for a cultural collection is to be compatible with general practices and protocols that contribute 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). So, the metadata schema for a cultural collection needs to include all kinds of data: descriptive (describe the resources intellectual content), structural (document the structure of the objects and the relationships between them) and administrative (provide information about the digitization process and the collection preservation).

The above aspects are based on the expectation to assure the greatest interoperability with other projects and applications and to be able to exchange and transfer data from other systems. Let’s not forget that 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. For this reason 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.

Many papers have been written about the challenging issues in creating 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. 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. So in order to cover various needs, often it is required to mix elements from separate metadata schemas and to customize them according to local requirements. A *metadata framework* is used as a container within which fields from multiple metadata schemas can be accommodated. For instance RDF (Resource Description Framework) and METS are considered to be such a framework. At last *Protocols* for data share between various digital services ensure the data communication and transfer among digital library services. The most public protocol for metadata share is the OAI-PMH that supports standardized exchange of metadata describing items in diffused collections.

In general, for covering the needs of a cultural collection with heterogeneous material and composite structure, there are two prevailing aspects:

- to design a metadata model that will be based on a widely accepted standard (a core schema like DC) and further extend it to cover other required aspects as i) the technique of digitization and the technical requirements ii) meta-metadata information because of the heterogeneous material and iii) the educational character and the purpose of every resource. change it according to project needs (either by extending it with additional local elements or with elements from other metadata standards), or
- to combine or apply different metadata standards in order to cover all the formats and types of physical and digital objects (text, sound, image etc.) and create afterwards mappings between these standards or mappings from these standards to only one schema for reasons of interoperability, persistency and querying processes (Cornell Univ, 2001).

Since memory institutions handle all types of material, the need of information integration is quite apparent. The goals of cultural heritage 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 et al, 2002).

4. Data preservation policy

Each organization dealing with digitization projects should also implement preservation standards widely used from the international cultural community. The fact that information is increasingly stored in digital form makes data preservation policy a considerable issue for a cultural digital collection to manage. Preservation Metadata Set is intended to include the elements that are believed to facilitate digital information resources persistency. It is meant to be not only a data input model, but also a data output model. It indicates the information we want our metadata system to keep inside (what data should be entered, how it should be entered, by whom and at what time) and also the information we want to take of the system for a digital resource. In other words preservation metadata is the structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”.

So a digital project needs to design and clarify the preservation metadata set that prefers to be kept for digital objects. The model needs to contain information about the digital object *provenance* (Who has had custody/ownership of the digital object?), *authenticity*, *any preservation activity* (What has been done to preserve the digital object?), *technical environment* and *rights management*. There are some quite remarkable efforts that deal with this matter like OAI reference model, PREMIS, CEDARS guide and many others.

5. Conclusions

Crucial factors in designing a effective and flexible digital cultural collection are: interoperability, verification, documentation, reusability, persistence etc. Except from the selection of digital repository system, focus should be given to three issues: composite collection/objects description, metadata definition and project purposes, but there are not strict rules to follow. In order to manage and expose the wealth of composite cultural heritage collections, a functional metadata policy covering collection-level and item-level descriptions as well as facilitating effective access to digital content should be defined. This policy, in most cases, results in the integration of more the one metadata schemes. Furthermore, metadata interoperability issues since a variety of metadata standards are used to describe heterogeneous material collections. The basic purpose is to create a cultural digital collection that will serve not only local needs but will be reusable in new and innovative contexts.

The most important indicators for an effective planning of a cultural digital collection are whether: a) users can easily find the information they need (information retrieval), b) the digital collection administrator is able to preserve and locate digital data and manage big collections without much effort (functionality), c) digital planning respect the institution's management policy and priorities, d) there is respect for the intellectual property rights and e) digital information resources fit into larger context of significant related national and international digital library initiatives (e.g., ability to contribute metadata to more inclusive search engines) and be exchangeable across platforms.

6. References

Categories for the Description of Works of Art (CDWA)[online]. Available from: <http://www.getty.edu/research/institute/standards/cdwa/>.

CORNELL Univ.2001, Mixing and mapping metadata to provide integrated access to digital Library collections, In DC-2001: Proceedings of the International Conference on Dublin Core and metadata applications, Japan,<http://www.nii.ac.jp/dc2001/proceedings/product/paper-23.pdf>.

Dempsey, L., Scientific, Industrial, and Cultural Heritage: a shared approach. Ariadne Issue 22, 1999, Available from: <http://www.ariadne.ac.uk/issue22/>.

Dublin Core Metadata Initiative (DCMI) [online]. Available from: <http://dublincore.org/>.

Dublin Core Collection Description Application Profile, Available from: <http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2003-08-25/>

Encoded Archival Description (EAD) [online]. Available from: <http://www.loc.gov/ead/>

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/>.

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>.

Joint Information Systems Committee (JISC), Cedars Project, Available from: <http://www.leeds.ac.uk/cedars/indexold.htm>.

Library of Congress. 2004. Metadata Encoding and Transmission Standard (METS) [online]. Available from: <http://www.loc.gov/standards/mets/>.

Library of Congress. 2005. Machine Readable Cataloging 21 (MARC 21) [online]. Available from: <http://www.loc.gov/marc/>.

Library of Congress. 2006. Metadata Object Description Schema (MODS) [online]. Available from: <http://www.loc.gov/standards/mods/>.

Lourdi, E., Nikolaidou, M., and Papatheodorou, C., Implementing Digital Folklore Collections , Proceedings of ISCA Third International Conference on Computer Science, Software Engineering, Information Technology, e-Business, and Applications (CSITeA-04) December 27-29, Cairo, Egypt, 2004.

MDA. SPECTRUM Units of Information [online]. Available from: <http://www.mda.org.uk/spectrum.htm/>.

NSO, Understanding Metadata (NISO press, 2004).

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>.

Research Libraries Group. 2005. Descriptive metadata guidelines for RLG cultural materials [online] RLG. Available from: http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf.

Research Support Libraries Program (RSLP). Collection Description Schema [online]. Available from: <http://www.ukoln.ac.uk/metadata/rslp/>.

RLG and the library consortium OCLC, Preservation Metadata maintenance Activity. Available from: <http://www.loc.gov/standards/premis/>.

The Open Archives Initiative Protocol for Metadata Harvesting [online]. Available from: <http://www.openarchives.org/>.

VRA Core Categories [online] (Updated 20 February 2002) Available from: <http://www.vraweb.org/vracore3.htm>.

W3C. Resource Description Framework (RDF) [online] Available from: <http://www.w3.org/RDF/>.