

# A metadata application profile for collection-level description of digital folklore resources

Irene Lourdi<sup>1,2</sup>

<sup>1</sup>Department of Archive and Library Sciences,  
Ionian University, Corfu, Greece

<sup>2</sup>Libraries Computer Centre, National &  
Kapodistrian University of Athens, Athens,  
Greece

elourdi@lib.uoa.gr

Christos Papatheodorou<sup>1</sup>

<sup>1</sup>Department of Archive and Library Sciences,  
Ionian University, Corfu, Greece

papatheodor@ionio.gr

## Abstract

*The preservation and representation of folklore collections is a basic priority for every country because they are valuable for studying the customs and the tradition of specific groups of people and places. However, for the heterogeneity and diversity of folklore resources (text, images, photographs, 3D objects, sound recordings, maps or even digital material), it is difficult to create a unified and semantically rich description concerning both the collections and their objects. In this paper we introduce an integrated metadata model by mixing elements of different metadata standards in order to make the navigation to the digital collection efficient and to provide the users with rich meaningful information retrieval to collection objects. The model fits with the requirements and the unique characteristics of the folklore collections and is focused on facilitating the retrieval of information to all the structural levels.*

## 1. Introduction

The folklore collections are a valuable source for study and research about the cultural heritage of a society. They refer to various aspects of their life such as: customs, folk tales, music, architecture, clothing, handicraft, texts and the oral tradition. In one sense, the folklore collections contain objects that reflect the common way of thinking, living and believing of a society, while there is a constituent affection from external elements that cannot be prevented. The wide scientific range of scope that a folklore collection covers makes the preservation and representation of the material in the web an urgent priority for every

organization owner, in order to satisfy the retrieval of information by various users and the navigation to old cultural material.

In this paper we represent an effort to improve the retrieval of information of large compound digital folklore collections with different kinds of compound objects. Our goal is to preserve and make available to every user all the precious information by providing an “information landscape” adapted to the needs of users or groups of users. For this purpose we develop a metadata model that enables efficient navigation to the collection structures and improves the retrieval of cultural heritage information. In the next section we represent the main characteristics of a folklore collection and the requirements for developing a “good” digital folklore collection, while in section 3 we evaluate the existing metadata standards for collection-level description and we refer to metadata schemas that have been developed for specific projects. In section 4 we analyze our case and we represent an integrated metadata model for the description of folklore collections and finally in section 5 we discuss the conclusions and our further work.

## 2. Folklore collections

### 2.1. Main characteristics

The main difficulty for managing and making retrievable such kind of collections is the fact that they contain material with unique characteristics. One of the most important is the *variety* of cultural heritage material. According to the guidelines of Unesco [1], an ethnographic collection may consist of either text written on different kinds of materials, still images, photographs, 3D objects, sound recordings, maps or

even digital material. Consequently, it is quite justifiable to have collections and sub-collections with *complex structure* and *rich semantics*. Furthermore, folklore collections concern a *varied audience* of different educational level that uses and searches information about cultural heritage and traditions and at last the content of the cultural heritage resources is characterized by an “*active character*”.

All these characteristics make quite clear the reasons why it is difficult to retrieve the right kind of information from a folklore collection that is composite and is consisted of many sub-collections and different kinds of material. The discovery of data is quite complex because the collection is multilevel and there are many structural levels that need to be accessed and searched.

## **2.2. Requirements for creating a “good” digital folklore collection**

Our work is motivated by the digitization project of the collection of the Folklore Department of the National & Kapodistrian University of Athens, which consists of sub-collections of different kinds of material, such as the sub-collection of notebooks written by students, the sub-collection of sound recordings or non sound recordings which some of them are accompanying material of the notebooks, the sub-collection of physical objects that are exposed in the library, the sub-collection of photographs and the sub-collection of sound and non sound recordings.

In order to understand better the following requirements, it is important to analyze further the structure of the sub-collection of notebooks. Each notebook has been written by a student after making local research by visiting specific places and taking interviews from the habitants of these places. The narrative text is separated in chapters and subsections and there are also indexes and tables of contents; most of the notebooks contain as accompanying material to the basic text maps, photographs of habitants and places and small objects stuck on a specific page related to the narrative text and the most recent notebooks are carrying sound recordings with songs and folk music. It is a folklore collection big in size, rich in relationships between the resources and with complex structure. The volume and the variety of the material require the representation of the collection with a multi-hierarchy structure of sub-collections according to the type of the objects, the corresponding chronological period and geographic region.

A “good” digital folklore collection should provide besides the structural decomposition, also the semantic decomposition with the appropriate markup language,

in order to expand the discovery of information [2]. Thus, the basic requirements of a digital folklore collection are:

a) *The retrieval of information*: a digital folklore collection needs to allow the retrieval of information by many categories of scientists and simple users such as: historians, philologists, librarians, psychologists, lawyers dealing with public law, ethnologists, reporters, musicians [3]. Therefore, the implementation of the digital collection shall allow the dissemination of the related knowledge to the users through the web and shall provide a familiar environment for searching and discovery of information.

b) *Organization of the collection*: according to the above requirement and in order to accomplish a full retrieval of information from the whole collection, it is necessary to organize the collection and its sub-collections into groups with specific criteria and to map the relations that exist between them and between the objects. The variety of material of a folklore collection justifies the need for separating the material into collections and sub-collections with complex structures and rich semantics. The main criteria that must be taken into consideration in separating the resources, according to Johnston and Robinson [4] are: i) the common subject and topic coverage of the resources, ii) the specific usage or purpose that each resource has in the context of digital collection, iii) the common provenance, iv) the same category of material, v) the specific spatial coverage or specific temporal coverage and vi) the same category of object such as: photographs, clothes, fairy tales, etc.

c) *Collection - level description and definition*: it is strongly observed by the emerging distributed approaches the importance of high level collection description in order to help the navigation, discovery and selection of the cultural content [5]. Also, it is strongly requested in our case to have a detailed collection-level and sub-collection-level metadata description, that will provide information about the contents and the size of the collection, the metadata schemas that are used in describing separately the various objects of the collection (as it is said above the folklore collections contain a variety of material that will be described with the appropriate metadata schemas) and also about the purpose and the historical context that the collection or sub-collection has been created and other metadata elements for the administration and the technical requirements.

## **3. A data model for the description of folklore collections**

### **3.1 Basic requirements**

Based on the above analysis the main requirements for the design of the desired metadata model are:

- a) The model should express both the subject coverage of the objects and the details of their creation. The digital folklore collections are valuable source for study and learning the cultural heritage of a country, so the data model besides structure must also express the semantic definition of folklore material.
- b) We need to depict the collection and sub-collections hierarchy in order to make all the structural levels searchable and retrievable.
- c) The proposed data model needs to be interoperable with the schemas that will be used separately for the item-level description. So collection-level description should provide the basic features of the metadata schemas that are used for item level description and mappings between them.

### **3.2 Existing metadata standards for collection-level description**

It is observed that the last years there has been an intense interest in complementing metadata schemas for collection-level description either for helping the users to decide easier whether the collection is of their interest or for better administration of large complex collections. Because of the difficulties of having searchable and fully retrievable digital folklore collections, we have studied many existing metadata standards and specific domain schemas for describing such kinds of collections and how these can be fitted in our case.

For library collections the most known standards are Dublin Core Collection Description Application Profile [6], that has been developed recently, the RSLP [7] model and the Marc format [8]; for archival collection there is ISAD [9] for archival description and EAD [10] for the encoding of the finding aid and there is also the “Z39.50 Profile for Access to Digital Collections” [11], which is proposed by CIMI for museums. Also there have been developed many metadata application profiles for specific projects by extending a metadata standard and enriching it with domain elements for specific purposes. Some examples of these application profiles for collection-level description are the Alexandria Digital Library (adl) [12] metadata schema, Renardus collection description [13], Riding and Agora experience [14], National Library of Australia [15] and many other that have been implemented to serve specific user needs and project requirements.

The above metadata schemas are not expressive enough in providing a collection-level description of

folklore resources. Given the fact that every schema has been designed for general collection descriptions or for the needs of particular projects, we observe that there exist important missing elements from each other and that every schema offer specific attributes for retrieval of information, which are not enough for covering the basic information that is suggested to have for our digital collection. For example according to our case, only the adl schema recognize the existence of information about meta-metadata description, which denotes the metadata schemas that are used for describing the digital objects inside the digital collection, but it does not contain other elements such as: information for the contents of collection. The Dublin Core Collection Description Application Profile provides rich description and information for the audience or the contents of the collection, but it does not provide meta-metadata information or information for the legal status of the collection; the same also with the Marc standard, which is very difficult and cost in time for implementing it and does not either contain information for meta metadata. The RSLP schema is very good for creating machine readable descriptions of collections but it refers mostly to collections of physical objects and so much digital collections because they are missing elements that are appropriate for describing and cataloguing digital resources. On the other hand, the archival descriptions are not adequate for our case because even though many of the elements are very useful for retrieval of information, they do not provide full retrieval in object level, which is required for the folklore collection. And at last the “Z39.50 Profile for Access to Digital Collections” (CIMI description) is very poor for the collection and sub-collections level description.

Our intention is not to reject these metadata standards but to exploit them in order to construct an application profile that fits exactly with the requirements of folklore collections and that will improve and help users to discover exactly the information that they are looking for or to guide them to the digital collection that it is best for their interest. What we need is to make the retrieval of information easy by providing to the users as many access points as they need and especially at collection-level description.

### **4. The proposed integrated data model**

Based on the above requirements, we developed a data model, which integrates elements from different metadata schemas and is enriched by new local domain elements when it is necessary. Our expectation is to use the model to: a) help the discovery of information collection level in such a way that the user can decide

whether the collection is of his interest or not, before continuing further with the navigation and the searching of each object inside the collection and b) to satisfy the demands of specialized users providing them the best way of acquiring the information that they need from all the structural levels.

The proposed model for collection-level description combines elements from a variety of metadata schemes to describe many thematically interlinked sub-collections with compound objects. In particular we integrate the following collection-level metadata standards into an application profile, suitable for describing and improving the retrieval of the folklore collections: the basic Dublin Core schema (dc) and the refined terms (dcterms), the Dublin Core Collection Description Application Profile (refined here as dc-cdap for reasons of separating it from the basic Dublin core schema), ISAD (isad), rslp, Alexandria Digital Library (adl) and IEEE-Learning Object Metadata (lom) [16]. Table 1 presents the main metadata elements that we propose. After each element is the metadata standard that the element belongs to.

Table1. The proposed metadata model for folklore collection description

<b>PROPOSED METADATA ELEMENTS</b>	
<b>Title</b> (dc-cdap)	<b>Owner</b> (dc-cdap)
<b>Alternative title</b> (dc-cdap)	<b>Accrual Status</b> (dc-cdap)
<b>Identifier</b> (dc-cdap)	<b>Relation</b> (dc& dcterms)
<b>Spatial coverage</b> (dc-cdap)	<b>Location: physical &amp; electronic repository</b> (rslp)
<b>Temporal coverage</b> (dc-cdap)	<b>Metadata schema</b> (adl)
<b>Accumulation date range</b> (dc-cdap)	<b>Metadata mapping</b> (adl)
<b>Custodial History</b> (dc-cdap)	<b>Medium</b> (dcterms) (i)
<b>Size</b> (dc-cdap) (i)	<b>Contributor</b> (dc)
<b>Audience</b> (dc-cdap)	<b>Structure</b> (lom)
<b>Type</b> (dc-cdap)	<b>Legal Status</b> (isad)
<b>Subject</b> (dc-cdap) (i)	<b>Note</b> (rslp/ isad)
<b>Language</b> (dc-cdap)	<b>Rights</b> (dc)
<b>Creator/collector</b> (dc-cdap)	<b>Date: issued, available, created original or digital</b> (dcterms)
<b>Sub-Collection</b> (dc-cdap)	<b>Source</b> (dc)
<b>Super-Collection</b> (dc-cdap)	<b>Abstract</b> (dcterms)
<b>Associated collection</b> (dc-cdap)	<b>Table of contents</b> (dcterms)
<b>Access rights</b> (dc-cdap)	<b>Scope/purpose</b> (isad)

First of all, it is important to mention that the Dublin Core Collection Description Application Profile has been used as the basic metadata standard. Specifically the elements that come from it, as it seems in table1, are: *title* and the refined term *alternative title*, *accumulation date range*, the refined terms of coverage *spatial* and *temporal coverage*, *identifier*, *subject*, *language*, *type*, *audience*, *custodial history*, *accrual status*, the refined term of format *size*, *collector*, *owner*, *sub-collection*, *super-collection*, *associated collection* and *access rights*. Furthermore, we have added elements from the basic Dublin Core schema so as to cover some points of the collection that need to be explained more for providing a rich picture of the folklore collection and for retrieving all the information required by the users. These elements are: the refined terms of description *table of contents* and *abstract* which can give in thorough the contents of the collection or sub – collection. Additionally to the elements of relations in proposed model we have included the element *relation* with all the refined terms that it can be specified, in case that we want to denote relations that are not covered by the elements of sub-collection, super-collection and associated collection; further we have the additional elements of *contributor* and *source* of the collection. At last we have added the element *rights* because the more information we keep for the folklore collection the better is for the protection of the content and *date* with all the refined terms and especially by separating the refined term *created* for physical and for digital collection.

For the same reasons and for specific purposes, as explained above, we have added entities from ISAD such as: the *legal status* of the collection, the *note* for giving other kind of information for the collection that don't fit in other entity and the *scope/purpose* because we consider important to specify for which purpose the collection has been created or how it will be used. Also we have used the element *structure* from Lom denoting the structure of the collection (i.e hierarchical, linear or networked) and the element *location* from Rslp in order to specify both where is placed the physical and digital collection. But the most important elements are the ones about metadata information (*metadata schema & mapping*) from the schema of Alexandria Digital collection, which describe the metadata schemas that will be used for item-level description of the resources of the composite folklore collection and the mappings between them. This kind of information is valuable for the description of the folklore collection because the collection contains heterogeneous resources that require further to be described separately with different metadata standards according to the format of each resource.

The model has been implemented in RDF using its XML syntax because it allows combining different metadata namespaces and helps in expressing the structure of a collection and the relationships that exist inside and outside the collection. Also we define in our model the “inherit metadata”, using the notation “(i)”, which denote those metadata elements that will be implemented automatically by the system for saving time besides from the metadata elements that will be filled manually (contextual metadata) [17].

Up to now we have tested our model with the sub-collection of notebooks written by students of the Folklore Department of the University of Athens of Greece, which as it has been described, is a very good example of a large compound collection with varied resources and complex relationships between them.

## 5. Conclusions

The current scientific research intends to create and implement a general metadata model that facilitates the retrieval of information of digital folklore collections consisting of heterogeneous resources by giving a description for the collection with the appropriate data model. The model focuses more on collection level and less on item level because the nature of folklore collections requires to provide information for every structural level and the first level of information is the collection and sub-collection level. Our expectation is to establish a model for affectively describing and administering many folklore collections with each other and their metadata elements in the environment of a digital library, so as the users can easily retrieve information and comprehend the content and context of the collections for their own satisfaction. It is the basic step for a user to understand the contents of a folklore collection and to decide if the collection is what he is looking for. Resources, collections and other services need to be defined and structured so that human and machine users can discover and make effective use of them.

The future work will be directed to the semantic decomposition of the collection by using markup languages. With the right metadata elements and markup language we can create formal records that express the semantic internal structural organization of the collection and we will have the opportunity to search not only for information about the resources but also for information inside them. Succeeding in this we will also have the opportunity to create innovative services in web-based interactive environments according to the users’ needs.

## 6. References

- [1] UNESCO, Memory of the world programme: general guidelines to safeguard documentary heritage [http://www.unesco.org/webworld/mdm/administ/pdf/MOW\\_FIN.PDF](http://www.unesco.org/webworld/mdm/administ/pdf/MOW_FIN.PDF) [viewed 15/07/2003]
- [2] SHREVE, G. M. and ZENG, M. L. Integrating resource metadata and domain markup in an NSDL collection, In *DC-2003: Proceedings of the International DCMI Metadata Conference and Workshop*, Seattle, WA, pp. 223-229
- [3] SOERGEL, D., *et al*, The many uses of Digitized Oral History Collections: Implications for Design. <http://guir.berkeley.edu/projects/oralhistory/SoergelECDL2002.pdf> [viewed 01/10/2003]
- [4] JOHNSTON, P., Robinson, B. Collections and Collection Description. Collection Description Focus, Paper1, 2002 <http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf> [viewed 08/09/2003]
- [5] DEMPSEY, L. Scientific, Industrial, and Cultural Heritage: a shared approach. *Ariadne* Issue 22, 1999 <http://www.ariadne.ac.uk/issue22/> [viewed 18/02/2004]
- [6] Dublin Core Collection Description Application Profile <http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2003-08-25/> [viewed 01/02/2004]
- [7] RSLP Collection Description Schema <http://www.ukoln.ac.uk/metadata/rslp/> [viewed 01/03/2004]
- [8] Marc Standards <http://www.loc.gov/marc/> [viewed 01/03/2004]
- [9] International Standard Archival Description ISAD(G) [http://www.ica.org/biblio/cds/isad\\_g\\_2e.pdf](http://www.ica.org/biblio/cds/isad_g_2e.pdf)
- [10] Encoded Archival Description (EAD), 2002 <http://www.loc.gov/ead/> [viewed 01/02/2004]
- [11] Z39.50 Profile for Access to Digital Collections Library of Congress <http://www.loc.gov/z3950/agency/profiles/collections.html>
- [12] HILL, L. *et al*. Collection Metadata Solutions for Digital Library Applications, 1998 (ADL description)
- [13] Renardus Collection Level Description (RCLD) <http://renardus.sub.uni-goettingen.de/renap/rclld.html> [viewed 05/02/2004]
- [14] BRACK, E.V. *et al*. Collection Level Description – The Riding and Agora Experience, 2000 *Dlib* Vol 6, No 9 <http://www.dlib.org/dlib/september00/brack/09brack.html> [viewed 02/02/2004]
- [15] NLA: National Library of Australia <http://www.nla.gov.au/preserve/pmeta.html> [viewed 05/02/2004]
- [16] IEEE LOM 3.1 [http://grouper.ieee.org/LTSC/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://grouper.ieee.org/LTSC/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)
- [17] CORNELL Univ. 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> [viewed 02/02/2004]