

## SCROL application profile

Steven Wu  
National Heritage Board,  
Singapore  
steven\_wu@nhb.gov.sg

Barbara Reed  
Recordkeeping Innovation  
Pty Ltd, Australia  
b.reed@records.com.au

Paul Loke  
National Heritage Board,  
Singapore  
paul\_loke@nhb.gov.sg

### Abstract

The National Heritage Board's virtual repository project resulted in SCROL ("Singapore Cultural Resources Online") which was launched in August 2006. This web-based repository pulls data and images from multiple information resources without the need to modify the existing databases. For advanced resource discovery, SCROL is equipped with a metadata search facility that retrieves relevant museum and archive records which have been ingested from several independent databases. Arguably, the SCROL application profile development has been an evolutionary process that restructured and refined existing schemas and standards in use. The final phase saw the consolidation of metadata elements through a methodological approach that also integrated a comprehensive subject taxonomy. This paper presents a case study in designing a Dublin Core-compliant application profile for cultural heritage. The rationale used in selecting and refining the SCROL element set so as to accommodate existing schemas in order to enable data re-use is explained. Future maintenance and review issues are also discussed.

**Keywords:** SCROL; cultural heritage; virtual repository; application profile; resource discovery; metadata search; archive; subject taxonomy; metadata element set; element refinement; National Heritage Board; A2O.

### 1. Introduction

The National Heritage Board of Singapore ("NHB") has under its wings the National Archives of Singapore ("NAS") and four museums, namely, National Museum of Singapore ("NMS"), Asian Civilisations Museum ("ACM"), Singapore Arts Museum ("SAM"), Singapore Philatelic Museum ("SPM"), Peranakan Museum (PM) as well as the Heritage Conservation Centre (HCC).

The virtual repository project was conceived in mid 2004 to integrate NHB's disparate repositories in a cost-effective manner, and to construct a comprehensive web search portal that would provide seamless and intelligent access to all digitized information via the Internet. The project's main deliverables were the SCROL portal and the SCROL application profile. SCROL implementation was completed in August 2006.

Prior to SCROL, users have been able to access archive records over the web using A2O ("Access to Archives Online") set up by NAS. The A2O provides integrated access to the national archive records. These records are stored in separate databases according to type including photographs, oral history recordings, audiovisuals and sound recordings, speeches, maps and building plans, posters, private records and government records. Each record format requires a slightly different form of metadata description. A2O application profile evolved from these metadata descriptions implemented as database schemas. Whilst A2O is still in use, SCROL aims to provide enhanced search capabilities by inclusion of a subject taxonomy.

For the museums, a separate collection management system, IMCMS ("Integrated Museum Collection Management System") has recently been set up using a DC-compliant solution. The IMCMS metadata standard subsumes that of an older system, ARTS 2, in use by NHB's museums. In setting up SCROL, it was intended that its application profile straddled both the A2O as well as the IMCMS standards. Presently, IMCMS stores over 100,000 artefact records and associated digital images.

A full list of NHB databases and repositories is found in Appendix 1 (Loke et al., 2005). Collectively, these databases store over 3 million records and images.

## **2. Project Goals**

In developing the application profile and taxonomy, the general guidelines were that the application profile and taxonomy should be:

- Simple to adopt, allow ease of interpretation, adherence, verification and maintenance;
- Independent of and flexible enough to allow seamless implementation and maintenance on the proposed solution to achieve desirable performance targets as well as enable continuous improvements in search performance;
- Extensible to allow future additions or refinements of metadata elements and taxonomy fields; and
- Compliant with the Dublin Core and other international standards where appropriate.

## **3. Project Activities**

In developing SCROL application profile, the following activities were examined and undertaken by the project team:

- an overview of the methodology used to design the metadata standard;
- an analysis of existing and relevant information resource discovery metadata standards referenced during the project;
- issues considered during the design of the metadata standard;
- an analysis of existing and relevant taxonomies referenced during the project;
- implementation issues associated with the metadata standard and taxonomy;
- recommendations on mechanisms and processes to update the metadata standard and taxonomy; and
- recommendations on conduct of reviews of the standard and taxonomy.

## **4. Application Profile Development**

### **4.1. Overview**

The objectives of the SCROL project are to implement a unified virtual media repository to provide a single interface to the public and organisations maintaining these repositories. The standard supports the aims of the project by defining a core set of elements that are needed to retrieve information about cultural heritage resources. These core elements are then mapped to each institution's databases, thus inheriting the data values.

Thus, the SCROL application profile thus serves two roles: firstly as an information resource discovery metadata standard, and secondly as a tool to enable mapping to existing cataloguing systems in each independent institution involved in SCROL, thus enabling reuse of existing data (Loke et al., 2005).

As an information resource discovery tool, SCROL aims to tag resources so they:

- can be readily retrieved;
- can be assessed for relevance to a particular purpose;
- can be reused; and
- are able to be exchanged between multiple repositories of information resources.

## 4.2. Assumptions

The following assumptions were made as the basis for development of the standard:

- That for longevity and sustainability, the standard adopted must be widely applicable to the cultural heritage resources sector and in line with international approaches to cultural heritage portal developments.
- The metadata standard must be appropriate to multiple uses that can cross domains of discipline specific requirements and therefore operate at a higher level than any one discipline specific set of requirements.
- The focus of the initiative is the general public, rather than curatorial or archival staff.

## 4.3. Methodology

To develop the application profile, we undertook the following steps:

1. analysis of existing metadata standards to determine appropriateness as the SCROL application profile;
2. development of a draft set of elements for discussion with stakeholders;
3. refinement of initial draft set following discussion and recommendations on the elements for the SCROL application profile; and
4. development and documentation of the standard.

### 4.3.1. Development of a Draft Set of Elements

An initial working draft of metadata elements was compiled from the analysis work. This set of recommendations proposed for discussion 17 elements and 11 element refinements. At the initial discussion stage, 10 of the elements were proposed as mandatory.

Discussion among NHB institutions involved consideration of the relevance of the proposed elements and element refinements. Further issues targeted for discussion, clarification or verification with stakeholders were:

- individual institutional usage;
- the degree of complexity;
- whether any existing standard vocabularies could be used to populate the elements;
- mandatory and optional elements.

### 4.3.2. Refinement and Finalisation of Draft Element Set

Following the discussions, a recommendation for the SCROL metadata element set was approved. The discussions did not change the recommended approach to the metadata element set, and the final SCROL application profile contains:

- 17 elements, 10 of which are mandatory, and
- 11 element refinements.

Finalisation of the elements, usage requirements and encoding schemes continued after the formal discussion stage. A summary of the final standard is presented in Appendix 2 (Reed et al., 2005a, 2005b, 2006a & 2006b).

## 4.4. Issues Encountered in the Design of the Application Profile

In compiling the standard, a number of issues were considered, raised or decided which may be important for future maintenance.

#### 4.4.1. Different Models of Underlying Database Design

Most of the underlying databases in use by the NHB institutions (indeed, by any organisation) are designed to work within the parameters of their own functionality. Often they do not have information extraction and re-use as a primary requirement, but are designed to meet discipline specific functionality. This meant that in undertaking the mappings between individual databases and the SCROL, in a number of places multiple elements in the underlying database were mapped to a single SCROL element. It becomes a technical problem to ensure that the extracted data then brings with it its underlying field tag, so that the information can be understood in context. For example, from the GRID database, two elements Date\_Opened and Date\_Closed were mapped to the SCROL element Date. (Reed et al., 2006b)

#### 4.4.2. Dependence on Quality of Data in the Underlying Databases

The extent to which the SCROL can operate as a portal is totally dependent on the data quality and completeness of the underlying databases it extracts data from.

The A2O databases were developed over a period of time based on various archival standards. On the other hand, the IMCMS is a new system recently implemented for the museum sector of NHB. Much effort and time is being applied to translating previously disparate and non-uniform older systems into its structures. However, the data in both IMCMS and A2O databases are dynamic as new knowledge and interpretations of the objects will continuously be incorporated into the resource descriptions.

#### 4.4.3. Issues Arising With Specific Elements

As a part of the project, we constructed a metadata mapping between the SCROL standard and the various systems which exist within the NHB institutions. In the mappings, we identified source data in systems with which to populate mandatory elements, or proposed default values. Some issues were encountered when dealing with specific elements/element refinements. One element and one element refinement are highlighted here to illustrate the challenge of mapping elements from multiple sources (Reed et al., 2006b).

**Subject element:** This element will have the SCROL taxonomy associated with it. Some of the underlying databases have their own taxonomies to control the attribution of terms to be mapped to this element (e.g., PICAS, Posters). In other databases, no controlled language is used, but a descriptive free text field is used. The approach taken in this project is to inherit the existing subject terms enabling retrieval on these terms, while also attributing the relevant SCROL taxonomy terms.

**Format Medium element refinement:** This element refinement relates to the physical or digital manifestation of the resource. Within the Dublin Core community it is common to adopt the encoding scheme IMT (Internet Media Types). A medium encoding scheme has been developed for the SCROL application profile. The element's documentation is shown in Table 1.

### 4.5. Encoding Schemes

Encoding schemes are tables that prescribe the values which will be used to populate a metadata element. In the SCROL application profile, we use three encoding schemes, in addition to the significant work undertaken on the subject taxonomy. The other encoding schemes are:

- SCROL.Spatial scheme (uses SGMS\* location)
- SCROL Medium scheme
- SCROL Type scheme

\*SGMS stands for Singapore Government Metadata Standard

For the subject taxonomy, it was decided to adopt two widely available thesauri for cultural heritage resources – the Getty Art and Architecture Thesaurus (AAT) and the UNESCO thesaurus.

TABLE 1. Format Medium element refinement documentation.

<b>Definition</b>	The physical or digital manifestation of the resource			
<b>Purpose</b>	This element allows users to: <ul style="list-style-type: none"> <li>• search for items in a specific format</li> <li>• alert users to technology or equipment that may be needed to access records in a particular format</li> <li>• allow users to decide whether they wish to retrieve/download records based on the size or dimension of the resource</li> </ul>			
<b>Obligation</b>	Mandatory			
<b>Repeatability</b>	Not Repeatable			
<b>Scheme/Values</b>	SCROL Medium scheme			
<b>Element Refinements</b>	<i>Element Qualifier Name</i>	<i>Scheme</i>	<i>Obligation</i>	<i>Repeatability</i>
	Format Extent	-	Optional	Not Repeatable
	Format Medium	[SCROL Medium]	Mandatory	Repeatable
<b>Default Value</b>	The relevant encoding scheme, where used, must preface the values provided in the element. This can be set up as a default.			
<b>Example</b>	<i>Element Name</i>	<i>Scheme</i>	<i>Example</i>	
	Format Medium	[SCROL Scheme] Medium	Audio; Digital (for a digital audio resource)	
	Format Medium	[SCROL Scheme] Medium	Wood; Ivory; Gold (for a chest of drawers made of wood, inlaid with ivory and gold)	
	Format Medium	[SCROL Scheme] Medium	Microfilm	
	Format Medium	[SCROL Scheme] Medium	Paper	
	Format Extent	-	1.86 Mb (for an electronic image)	
	Format Extent	-	00:50:00 (for an audio file)	
	Format Extent	-	196 x 209 cm (for embroidery)	
<b>Source</b>	<i>System</i>	<i>Database Table Name &amp; Element Name(s)</i>		
	IMCMS	<i>For Extent:</i> Dimensions Page No and No of volumes (for literature) Image width (for multimedia) Image height (for multimedia) File data size (for multimedia) <i>For Medium</i> Material		
	A2O MICAS	<i>For extent:</i> [AV_Catalogue] Record Duration <i>For medium</i> [AV_Catalogue] Medium. Values 'm' translates to 'moving image' and 's' translates to 'Audio'		
	A2O PICAS	<i>For medium</i> [PC_Catalogue] Medium. However, Default to 'Image' except where value in medium is 'n' which translates to 'negative'; and 'p' translates to 'photograph'		
	A2O GRID	<i>For medium:</i> Default to 'Microfilm'		
	A2O PRISM	<i>For medium:</i> Default to 'Microfilm'		
	A2O CARDS	<i>For medium:</i> Default to 'Image'		
	A2O CORD	<i>For extent:</i> [OHC_Acquisition] Total Running Time [OHC_Catalogue] Pages <i>For medium:</i> Default to 'Audio'		
	A2O POSTERS	<i>For medium:</i> Default to 'Image'		
	A2O STARS	<i>For medium:</i> Default to 'Audio'		
<b>Date Issued</b>	October 2005			

## **5. Mechanisms for Updating and Maintaining the Application Profile**

Maintenance of the SCROL application profile is critical in ensuring that the standard continues to meet the requirements of its intended users. The application profile has been designed for flexibility and extensibility, but such extension needs to be carefully managed to ensure the inheritance of data values from the underlying databases remains valid through any extensions proposed in the future. Encoding schemes can be altered quite easily using the design methodology adopted without compromising the integrity of the standard.

### **5.1. Review Scope and Mechanism**

A review of the SCROL application profile and encoding schemes (including subject taxonomy) was made following the implementation of the pilot scheme. The first review mainly covered the implementation of SCROL up to the database ingestion and SCROL performance in live use.

Subsequent reviews will extend into user experience in using the application profile and Encoding Schemes (including subject taxonomy). We anticipate that the Encoding Schemes are the most likely to require amendment at each review point. The element level of the SCROL application profile, on the other hand, is expected to remain relatively stable.

A number of separate, but related tasks in relation to approving, processing and implementing changes to the application profile and associated documentation are needed to be in place for a review process to be successfully executed. In particular, these include:

- Establishing responsibility for approving changes to the standard and encoding schemes, and, if necessary, assessing and resolving conflicting requests for change;
- Establishing responsibility for documenting changes to the standard and associated documentation;
- Maintaining the mappings between the SCROL application profile and the underlying databases.

One organizational hurdle in the maintenance process is that no sector of NHB is logically the owner of the standard and taxonomy. While the technology implementation portion of the project logically belongs to the STO (Strategic Technology Office), the skills in managing the application profile and taxonomy belong more to the curatorial or archival sectors of NHB (Loke et al., 2006).

## **6. Conclusion**

The SCROL project has highlighted both technical as well as organizational hurdles in implementing an application profile. Nevertheless, the development of SCROL application profile has demonstrated the feasibility of the overall approach. When fully operational, SCROL is expected to be a valuable tool for resource discovery by all user groups. With continuing maintenance, SCROL application profile will adapt to changing user requirements particularly in the areas of data encoding and taxonomy. The success of the project depends on appropriate methodology, user ownership, and auditable processes.

## **Acknowledgements**

We are grateful to the National Heritage Board for supporting the dissemination of knowledge on which this paper is based. In particular, we would like to thank Philip Chua, CITO (NHB) for his encouragement.

## References

- Loke, Paul, et al. (2005, Feb.). *VMR specifications*.
- Loke, Paul, et al. (2006, Aug.). *SCROL post-implementation review*.
- Reed, Barbara, et al. (2005a, Oct.). *DCMES-SCROL-IMCMS* (Metadata implementation at NHB institutions).
- Reed, Barbara, et al. (2005b, Oct.). *SCROL Metadata Standard version 1.0*.
- Reed, Barbara, et al. (2006a, Aug.). *Final report on standard and taxonomy development version 1.0*.
- Reed, Barbara, et al. (2006b, Aug.). *Final metadata mapping report and appendixes version 1.0*.

## Relevant URLs

- NHB : <http://www.nhb.gov.sg>
- SCROL : <http://www.scrol.sg>
- A2O : <http://www.a2o.com.sg>

## Appendix 1. NHB Databases and Repositories

Mappings for SCROL elements were extracted from the following databases:

- Integrated Museum Collections Management System (IMCMS) – A system that supports the planning of exhibitions and integrates cataloguing, imaging, conservation, inventory and storage data across NHB's various institutions.
- Picture Archives Singapore (PICAS) – A text and image retrieval system on the web, which helps to disseminate a huge collection of historical photographs built up over the past few centuries.
- Moving Image and Sound Archives Singapore (MISAS) – A computerized cataloguing, retrieval and information management system providing fast and easy access to the audio-visual collection in the National Archives of Singapore (NAS).
- Speech-Text Archival and Retrieval System (STARS) – A system that provides public access to a growing collection of ministerial speeches, which were transferred to the National Archives of Singapore (NAS) mainly by Ministry of Culture and the Ministry of Information, Communication and the Arts (MICA).
- Cartographic & Architectural Records Database Singapore (CARDS) – A text and image retrieval system on the web, which helps to disseminate maps and building plans.
- Private Records Information System (PRISM) – A search and retrieval system for private collections acquired from individuals, clubs and societies, professional bodies, religious organizations and clan associations. These records provide valuable insights into Singapore's history.
- Government Records Information Database (GRID) – A system which offers insights into the forces and factors that shaped the policies of post-war Singapore.
- POSTERS – A text and image retrieval database for posters on the web. Posters constitute movie posters, concert posters or campaign posters such as family planning and tourism promotion.
- Collection of Oral History Recording Database (CORD) – A text and image retrieval system, which helps to disseminate social memories of individuals who lived through the era or witnessed historical events in the making. It has interviews with personalities from all walks of life, and their stories are for the benefit of future generations.

## Appendix 2. SCROL Element and Element Refinement Summary

Element	Element Refinement
Contributor	
<i>Creator</i>	
Publisher	
<i>Rights Holder</i>	
<i>Identifier</i>	
<i>Title</i>	
<i>Subject</i>	
<i>Description</i>	
<i>Date</i>	
Language	
Format	Extent
	Medium
<i>Type</i>	
Coverage	Spatial
	Temporal
Audience	
<i>Rights</i>	Access Rights
	Copyright
	Cost
Relation	Requires
	Is part of
	Has format
	Has part
Provenance	

*Italics* indicates mandatory elements.