Data Harmonisation between National Library Board, National Archives and National Heritage Board of Singapore

Shan Shan Chan  
National Library Board of Singapore  
Chan_Shan_Shan@nlb.gov.sg

Haliza Jailani  
National Library Board of Singapore  
Haliza_JAILANI@nlb.gov.sg

Keywords: data harmonisation; schema; mapping; crosswalking; single search interface; search engine optimization; controlled vocabularies; name headings; Taxonomy & Thesaurus Editor; OneSearch; cross-linkages

1. Introduction

The NLB Data Harmonisation Project aims to enhance user experience and the discovery of nuggets of resources from the rich collections of the National Library Board (NLB), the National Archives (Archives) and the National Heritage Board (Museums) of Singapore. Archives' and Museums' metadata records are ingested into NLB’s repository through the process of mapping, crosswalking and harvesting. All records can be searched through NLB's OneSearch, which is an integrated discovery service developed by NLB for the searching of physical and digital resources.

2. Data Preparation

Libraries, Archives and Museums metadata records share some common fields but largely use fields which are unique to their collections. NLB uses MARC21 for its physical resources and Dublin Core Libraries Application Profile (DC-Lib) for its digital collections. While Archives uses ISAD-G schema for archival description, Singapore's museums uses its own localised schema. As such, there is a need to harmonise these schemas so that seamless search can occur. Archives and libraries organise materials differently. The multi-level description of archives relates objects in a hierarchy and links the parts to a larger ensemble from the collection level perspective. Libraries organise at the item-level and groups these into collections for discovery. The museums' granularity of description for descriptive areas such as materials & techniques, styles and period, etc. are mostly not found in the descriptions for libraries and archives. Nonetheless, to achieve OneSearch, NLB takes the approach of crosswalking the various schemas to Dublin Core. The crosswalked records from MARC21, ISAD-G and the local schema are ingested into NLB's repository.

3. Controlled Vocabularies and Name Headings Integration and Standardisation

Like other NLs, NLB uses names authorised by Library of Congress Name Authority Cooperative Program (NACO) which observe strict rules for capturing every part of a name. Where names cannot be established in NACO, NLB uses a separate list from a local file. Archives has 8 databases comprising Posters; Oral History Interviews; Government Records; Audiovisual Recordings; Photographs; Maps & Building Plans; Straits Settlements, Overseas & Private Records; and Speeches & Press Releases. These databases are managed by different teams of archivists who do not necessarily share name headings or controlled vocabularies. Whereas Singapore's museums do not use any controlled lists at the time of the project. In the Archives & Museums portal prior to OneSearch, a single person may have more than one form of name. Search results for this person will not be unified and resources are retrieved according to the name a user enters. Merging of Archives collection as it is with the NLB collection will cause
The mapping of vocabularies between the three collections created a consolidated list of controlled vocabularies and name headings shared by NLB, Archives and Museums. The controlled lists of terms and name headings are managed using NLB’s vocabulary editor (Taxonomy & Thesaurus Editor). Upon receiving name headings provided by Archives and Museums, NLB’s team performed term-matching with NLB’s controlled terms and names. Unmatched names are created as new records. For both matched and unmatched names, the team needed to reconcile differences in the form of name used and come to an agreement on an authorized format with variant forms captured in the record. Material types between NLB and Archives were similarly mapped, merged and added to. For the museums, a study was made of the Art and Architecture Thesaurus (AAT) and recommendations were made on terms to be adopted for object categories and material types. These were then merged into NLB’s controlled list. The standardized terms reduce ambiguity and increase precision in searching. As a result, metadata of Archives and Museums content, regardless of format, can now be described consistently among NLB, Archives and Museums.

NLB uses ANSI/NISO Z39.19 for the construction, format and management of controlled vocabularies in the Taxonomy & Thesaurus Editor (TTE). To achieve standardization, NLB’s TTE will be integrated to the Archives’ indexing system. In addition, NLB has shared its policies and guidelines on name creation as well as an agreed upon, workflow between NLB and Archives. Training has been provided to Archives staff. Upon completion of the integration, further training will be provided to ensure processes are standardised so that the three agencies will be able to create and use a shared set of vocabularies and headings.

4. Technical Development

Five and one-half million NLB and 800,000 NAS metadata records were harmonised and OneSearch was launched in August 2014. On 3 June 2015, NLB completed the harmonisation of 80,000 museums metadata and enabled these for searching on OneSearch. The following were some of the technical developments:

1. OneSearch, the integrated search and online interface of NLB, Archives and Museums collections. This includes nine high-level groupings or containers of resources from the 3 agencies namely Books, Magazines & Articles, Audiovisuals, Images, Newspapers, Records & Papers; Websites, eJournals, and Physical Objects.

2. Implementation of search engine optimisation enhancements to improve the discovery of Archives and Museums content using the popular Internet search engines.

3. Deployment of embedded search service in NLB sites and cross-linkages within existing NLB and Archives content sites. Embedment of search service in Museums content sites is work in progress.

5. Benefits

The project was completed in August 2014. However, even before it was officially launched, the page view of Archives had seen substantial jump. During the period between April and November 2014, the page view of Archives was 2.1 million, nearly five times of FY14 target of 450,163. The increase of page views is largely attributed to the re-design of NAS website for search engine optimization and the launch of OneSearch in August 2014.

Users have also directly benefited as they can now access a wider range of materials from NLB, Archives and Museums through a single search without the need to go to three difference websites.

The Data Harmonisation has received extensive media coverage as OneSearch is the first search in Singapore which allows users access to a wide range of materials from Singapore’s...
Libraries, Archives & Museums, including digitised newspapers, literary works, museum artefacts, paintings, manuscripts and speeches, apart from the usual collection of books, magazines and audio-visual materials.