5. Impact and Benefits

• Though the system was not implemented, the KM team finally realises the power and impact (offering an additional dimension to locate asset) of a tool like TaxoFolk.
• Common tags used by users turn out to be a valuable source of information to consider during taxonomy revamp especially on the setting of terms and phrases; further discussion with users on how certain tags/labels are used reveals the mental models of many users.
• At the time when I showed TaxoFolk to the organization, they were not piloting other projects hence I would say much of the change and understanding are attributed to TaxoFolk. The KM team appreciates that during taxonomy revamp, more consideration needs to be paid to a user’s perspective on what they perceive and their use of common terms (which may well be different from terms that the organization has adopted).
• Delivering explanations at various levels of abstraction, patience, and provide a prototype or demo system for users to try.

6. Next Steps

Tag clouds can also be generated from user contributed tags and are provided to users as an additional mean to navigate over knowledge assets. These tags have improved the findability of explicit assets in the organization. From the research perspective, we intend to apply for funds to support personalization of the TaxoFolk hybrid navigation system for individual users; at present it is one system for all users based on the contributed tags.

HALIZA JAILANI

Using knowledgebases using named entity recognition, query expansion and Linked Data, to support enhanced discovery

1. About the Case Organization

The National Library Board Singapore manages the National Library, 26 Public Libraries and the National Archives, with a mission to make knowledge come alive, spark imagination and create possibilities. NLB supports knowledge seeking activities by providing a trusted, accessible and globally-connected library and information service.

2. About the Challenge

With the merger of National Archives Singapore (NAS) and the National Library Board in November 2012, it became critical to harmonise library and archives metadata, name headings & controlled vocabularies so the public can discover resources from both agencies without having to go through two portals. Metadata and terminologies need to be aligned. Knowledge organisation processes need to be centralised for efficiency and cost savings, and common tools shared. Not many libraries and archives in the world have merged their collections, difficult to do as 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.

Names are largely not standardised in NAS, where 8 different databases are managed individually by different teams. As a result a single person may have more than one form of name. Search results for this person are not unified and resources are retrieved according to the name a user enters. Merging of NAS collection as it is with the NLB collection will cause search results to be more fragmented. A search for an entity whether a person, an organisation or a place will not pull content resources about the entity into a single list. 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.

In addition, the library world is undergoing a shift in thinking on how libraries catalogue resources and enable them to be discovered. The long-time library cataloguing standard AACR2 (Anglo-American Cataloguing Rules 2nd Edition) was replaced with a less rigid standard that is more attuned to the needs of a digital world, called RDA or Resource, Description and Access. Increasing efforts are also being made to break the coded information out of the catalogue record so that they can be used as Linked Data and library materials can become searchable on the Internet.

3. What We Did

NLB used a three-pronged approach to organise the merged collections and improve resource discovery: (i) Data harmonisation & query expansion (ii) Linked Data implementation, (iii) Knowledge-base building.
NLB & NAS data harmonisation

Mapping was done between the archival schema, ISAD (G) and Dublin Core a generic schema for interoperability, where NLB had previously crosswalked 1.5m MARC records to DC. We managed the hierarchical ISAD (G) with item-level DC relationships and collection level description. About 6,000 people and organization names were extracted from the archives indexing system, de-duplicated and mapped to NACO and NLB local files. The finalised set was centralised in NLB’s Knowledge Organisation System (KOS) and maintained in the Thesaurus & Taxonomy Editor (TTE), a vocabulary management tool. System enhancements and data integration were done for the archives, a new website launched and OneSearch, an integrated search for NLB (4.5m titles) and NAS (0.8m titles) was delivered in Aug last year. Due to the different forms of names, requests from owners, aliases and use in different languages, we are now working to provide a query expansion service that will enable the non-preferred names and associated forms of name to be searchable as desired.

NLB Linked Data implementation

We implemented a Linked Data management system and created an extensible data model and an extended service vocabulary to support a web feature for NLB & NAS websites. More than 250,000 records and named entities were transformed into RDF. We incorporated a selection of the National Heritage Board’s galleries and museums metadata so as to create an RDF triple store of library, archive and museum data. Going beyond publishing the RDF data, NLB takes this a step further by conceptualizing a widget for NLB and NAS websites that will allow users to perform contextualised searches based on entities found on the website. Term extraction was done for unstructured data to enable nuggets of information to be extracted and added to NLB’s triple store. Based on the raw data extracted, named entities were identified and added to the knowledge-base. The data was also analysed so that explicit relationships can be built into the extended service vocabulary.

Knowledge-base Building

The above two approaches require a strong knowledge-base to meet discovery needs effectively. NLB named entities are described by a set of properties such as latitude and longitude and dates of birth and death. These set of properties need to be enriched if we are to deliver a useful service. We have started enriching this knowledge-base and are working to enhance it by tapping on existing data on the Internet using Linked Data. By leveraging on other knowledge-bases, we hope to be part of a global network of contextualised data.

4. Challenges and Lessons Learned

Working beyond the library makes us realise that data exist at different levels and are organised differently because of this. Example, orientation or arrangement of work is important for museums and archives to understand the meaning and context of the artefacts and archival materials. NLB do not manage its content this way. Do we capture and control vocabularies for such purposes? Names may be authorised in different forms for Getty’s Union List of Artist Names vs NACO. Which should be used or should a decision even be forced? Materials & technique and object/work type classifications for museums are at a very high granularity compared to the type categories found in libraries and archives. Compromises need to be made and interoperability can only occur at a high level. Insufficient metadata exist and vocabularies are not controlled or are unlikely to be controlled due to various reasons. NLB is looking at cleaning and reconciling data for the three agencies but this effort has its limitations.

5. Impact and Benefits

NAS content access and use has jumped substantially and we are now working to deliver OneSearch for the museum’s artefacts on the SGCool website (80,000 items) by mid this year. Linked Data exposes poor quality data. A good outcome from this is a commitment to improve the metadata descriptions and the use of applicable standards that would aid in collection organisation and stakeholders’ buy-in.

6. Next Steps

NLB is now in a position to share building blocks and learning points from its semantics-based resource discovery journey. NLB is also intending to share local terms, vocabularies and name headings to organisations with local cultural heritage collections and welcome their participation and use of the knowledge-base for their own resource description and discovery. NLB has set targets and resources for the coming years to increase the knowledge-base and publish them on international registries. Internally, we will also be transforming a further 700,000 authority and metadata records into RDF to enrich the web feature. We will continue to publish our RDF data on international registries and begin work with other national libraries to link the collection into a network of Linked Data.
Resource Discovery@NLB: Knowledge Organisation Systems

Knowledge-base Building
Linked Data
Data Harmonisation
Query Expansion

NLB SearchPlus
- Type: Photograph (10,048)
- Electronic Book (1,145)
- Article (230)
- Video Recording (140)
- Book (44)
- Show 15 more

- Collection:
  - SNAP (10,377)
  - Singapore Heritage Collection (10,231)
  - Lee Kip Lin Collection (10,129)
  - Singapore Infopedia (230)
  - Reference Closed Access (73)
  - Show 15 more

- Subject:
  - Architecture and Landscape – Streets and Places (10,205)
  - Architecture and Landscape – Building Types – Commercial Buildings (5,314)
  - Streets – Singapore – Photographs (4,824)
  - Streets (18339165) (3,423)
  - Commercial buildings (18339159) (2,856)
  - Show 15 more

Recording Details
- Title: Pasir Ris sunrise
- From: Pop
- Preview Music
- Play Full Recording

Information about this recording
- Performer(s): Crowd (Musical group)
- Conductors: -
- Composers: -
- Lyricists: -
- Arrangers: Matthew, Kevin M., Nelson, Tim
- Instruments: -
- Rights statement: All rights reserved by the respective intellectual property rights owners

Keywords: -
- LCSH Songs, English–Singapore
- nlbh5artfcs#>Performing Arts#>Music

Description: This photograph shows the stilt walkers with hand stilts performing a body-theatre piece called “Time Out,” by Antagon Theatre AKTion from Germany at the closing celebrations of the Singapore Arts Festival 2007 at Jurong East. The Enterprise building (formerly known as Unity House or McDermott House) can be seen in the background. Title devised by Library staff.

Coverage: Jurong East
Singapore
Knowledge-base Building of Singapore Entities

What is needed?

Specific taxonomies for key user groups (Arts, Cultural Heritage, Singapore Memory)

Singapore-centric vocabulary

Standardised local names (People, Organisations, Places, Events)
Data Harmonisation: Integrated Search for Library, Archives, Museums

80,000 museum records
Museums

4.5m library records
National & Public Libraries

Subscribed resources
Infopedia
BookSG
MusicSG
NewspaperSG
NORA
PictureSG
Singapore Memories
WAS

eJournals
Databases

880,000 archives records
National Archives

Books
Periodicals
AV
Ephemera
Rare materials

Posters
Oral history
Govt records
AV records
Photographs
Maps
Building plans

SGCool
Art Sculptures
Bone & Ivory
Documents & Paper
Furnishings
Historical Paintings
Installations, etc.
Users have to go to THREE places to look for resources.
Query Expansion

- Grouping relevance under one search
  - Different forms of names & in various languages
  - Related names
  - Former names (name changes)

E.g. Display Name: Lee Kuan Yew
    Preferred Term: Lee, Kuan Yew, 1923-2015

Variants
UF Harry Lee (_People) (NPT)
UF Lee, Harry Kuan Yew (_People) (NPT)
UF Li, Guangyao (_People) (NPT)

Related Names e.g.
- Lim Chong Pang (son) is Related To Lim Nee Soon (father)
- Lim Boon Keng (business partner) is Associated With Lim Nee Soon

Singapore. Ministry of Communications and Information (2012- )

Associations

Predecessor Singapore. Ministry of Information, Communications and the Arts (_Organisations)
ENGtoTAM கூட்டவியல் தொடர்பு அமைப்பு (_Organisations_TM)
Linked Data @ NLB

- Archives Online
- NLB Infopedia
- Singapore Memory
- NLB HistorySG
- Museums Portal

Linked Data Web Feature

- Linked Data Management System: entities extraction, data transformation, triple store & URI registry
- Ontology Building
- RDF data sets

External Data (e.g. DBPedia, LOC)

Mobile App

Linked Data browser
data.gov.sg

Knowledge-base Building
Data Harmonisation
Query Expansion

All Rights Reserved. National Library Board Singapore.
Building and linking entities to resources for improved discovery

NLB's Knowledge Base (KOS)

People Data

Corporate Data

Geographic Data

Enhancing Entities

Knowledge Discovery

Linked Data

Query Expansion

Named Entity Recognition & Text Extraction

From 2,000 entities (FY15) to 4,000 entities (FY16)
On 9 August 1965, Singapore separated from Malaysia to become an independent and sovereign state. The separation was the result of deep political and economic differences between the ruling parties of Singapore and Malaysia, which created communal tensions that resulted in racial riots in July and September 1964. At a press conference announcing the separation, then Singapore Prime Minister Lee Kuan Yew was overcome by emotions and broke down. Singapore's union with Malaysia had lasted for less than 23 months.

Singapore in Malaysia

Prime Minister Lee Kuan Yew signed the Malaysia Agreement in London on 9 July 1963. The agreement spelt out the terms for the formation of the Federation of Malaysia, comprising Singapore, Malaya, Sarawak and North Borneo (Sabah), which was to take place on 31 August 1963. The terms for Singapore's entry into Malaysia, which were agreed upon by both the Singapore and federal governments, were published in a White Paper in November 1961. This White Paper documented the outcome of talks between Lee and then Malayan Prime Minister Tunku Abdul Rahman on Singapore's inclusion into Malaysia. The terms included the margins of Singapore's autonomy, Singapore's political representation in the federal government, the status of Singapore citizens and Singapore's revenue contribution to the federal government. Prior to the signing of the Malaysia Agreement in London, there was a week of "arduous and gruelling negotiations" over the more thorny issues of a common market between Singapore and Malaya, and the portion of Singapore's revenue and taxes that would go to the federal government. With these issues settled, Singapore began its journey as part of Malaysia.

A Difficult Union

Even before the proclamation of the formation of the Federation of Malaysia on 16 September 1963, Singapore and Malayan leaders were mindful that the differences in the political approach and economic conditions between the two countries "cannot be wiped out overnight."
Ong, Teng Cheong

**Synopsis**

President of Singapore (1993-1999); Member of Parliament for Kim Keat Constituency (1972-1993); Chairman of the People’s Action Party (PAP) (1981-1993); Secretary-General of NTUC (1983-1993); Deputy Prime Minister (1985-1993).

**Achievement:**

First elected President of Singapore (1993-99). Pushed for the Mass Rapid Transit (MRT) system, which also changed land-use patterns in Singapore.

---

**Archives Online**

**KOS**

**SMP**

**marc record**

**DC record**

---

**Singapore Infopedia**

**Related Subject**

- Subject List
- LSC: Social ethics -- Singapore
- LSC: Special weeks -- Singapore
- LSC: Special events -- Singapore
- LSC: Law and government -- Politics

**Work by**

- Official opening of Yuhua Community Centre: guests at balcony
- Official opening of Yuhua Community Centre: guests seated
- Official opening of the Yuhua Community Centre: guests enjoying dinner

**Memesries**

- Digital article
- Thumbbook
- Arts events
- Music
- Singapore
- Thumbbook
- Photographs

---

**Subject List**

- nlsbtx: Events >> National Campaigns
- nlsbtx: Politics and Government
- LSC: Social ethics -- Singapore
- LSC: Special weeks -- Singapore
- LSC: Special events -- Singapore
- LSC: Law and government -- Politics
Challenges and Learning Lessons

**Open Data**
- Regulatory role for open data standards
- Public domain rights to data
- Strategy for data sharing

**Data Quality**
- Data integrity, cleanliness & currency
- Levels of data specificity
- Unstructured Data – concept extraction

**Service Value-add**
- Stakeholder buy-in
- Staff competency & resource crunch
- Integration & interoperability
Thank you for your attention..

haliza_jailani@nlb.gov.sg