

No image in this paper is to be used without the express permission of the National Heritage Board, Singapore.

Author: HO, SWEE ANN

Institution: Heritage Conservation Centre / National Heritage Board, Singapore

Contact: ho_swee_ann@nhb.gov.sg

Keywords: heritage cataloguing, controlled vocabulary, taxonomy,
Heritage Conservation Centre

Title: ***Use of Controlled Vocabulary in the Documentation of the Singapore National Collection***

1. Introduction

The National Heritage Board (NHB), Singapore, launched an NHB-wide digital Collections Management System (CMS) in an endeavour to digitalise the National Collection (NC) in 2005. Managed by the Heritage Conservation Centre (HCC), the CMS (presently known as Singapore Collections Management System (SCMS)), contains the inventory data and research information of the NC. In 2014, to better engage the public digitally, NHB proposed a Digital Engagement Framework (DEF), resulting in the launch of *Roots* in 2016 which functions as the NHB resource portal, hosting NHB content related pages on a single platform. To enable greater search-ability and accessibility of the NC records on *Roots*, HCC was tasked to enhance the records in SCMS.

Former HCC colleagues had recognised the need to enhance the data in the then-CMS for better retrieval for many years. These were expressed in two papers which were presented the 2005 and 2010 CiDOC conferences respectively. Teh (2005, p. 2) and Low and Doerr (2010, p. 5) stated that record retrievals would be a challenge due to the inconsistent documentation standards and lack of terminology control used to document the NC. Low and Doerr (2010, p. 6) discussed the need for controlled vocabularies and a need for a thesaurus, however, due to the specialised skill, effort and time required then, it could not be done. In short, the two papers emphasized the need for controlled vocabulary, local terms, and enhancement of research content in the CMS to aid more efficient record/information retrieval. Low and Doerr (2010, p. 11) suggested a team specifically dedicated to do these tasks due to the significant amount of time and effort required. In December 2014, coincidentally with the proposal of the DEF, this team, the Heritage Cataloguing Team was formed at HCC to enhance the records in the SCMS.

2. Challenges with existing records

The team met with a few challenges when they started to (re-) organise and structure the data in the system. Firstly, the data was not standardised due to different standards in information capturing as the birth of the NC is dated way back to the late 19th century, (during the days of the Raffles Library and Museum – presently the National Museum of Singapore (NMS)). When the records were transferred from hard copy into the then CMS, data was just transcribed according to how the handwritten notes were read and deciphered. This resulted in variations in the naming and spelling of terms and object descriptions. In addition to that were spelling errors. An example would be, the country name of Singapore. The different forms of spelling found were “Singapore”, “S’pore”, “Singapura”, “Sg” and spelling error, “Singapore”. Hence, one would not be able to retrieve all the object records associated with Singapore in the collection with the search keyword “Singapore”. There were different terms used for the same object type, for example, “bangles” and “bracelets” were used interchangeably to tag the same object type.

The second challenge identified was the way the objects were categorised. The objects were categorised according to how they were stored. This categorisation, which was to aid with storage and retrieval of physical objects then, continued when we moved from hardcopy documentation into a CMS. This resulted in objects of the same form but of different material type to be placed in different categories. For example – sculptures; there are gold, silver, brass, ceramic and bronze sculptures in the collection. These sculptures were placed (tagged) into different categories according to their material and not held together by a “Sculpture” category. The free text search engine

No image in this paper is to be used without the express permission of the National Heritage Board, Singapore.

for both the former SCMS and Roots does keyword searches and would retrieve any record with the word “sculpture” – be it a postcard, photograph, contemporary art sculpture, or a traditional sculpture. Hence, to retrieve all traditional sculptures in the collection, a user would then need to perform five different searches using the five material categories, each one at a time, as that was the ability of the system back then. (See Image 1: *Category filter on former Roots (2016 – 2020)*) This was not efficient or user-friendly for search and retrieval.

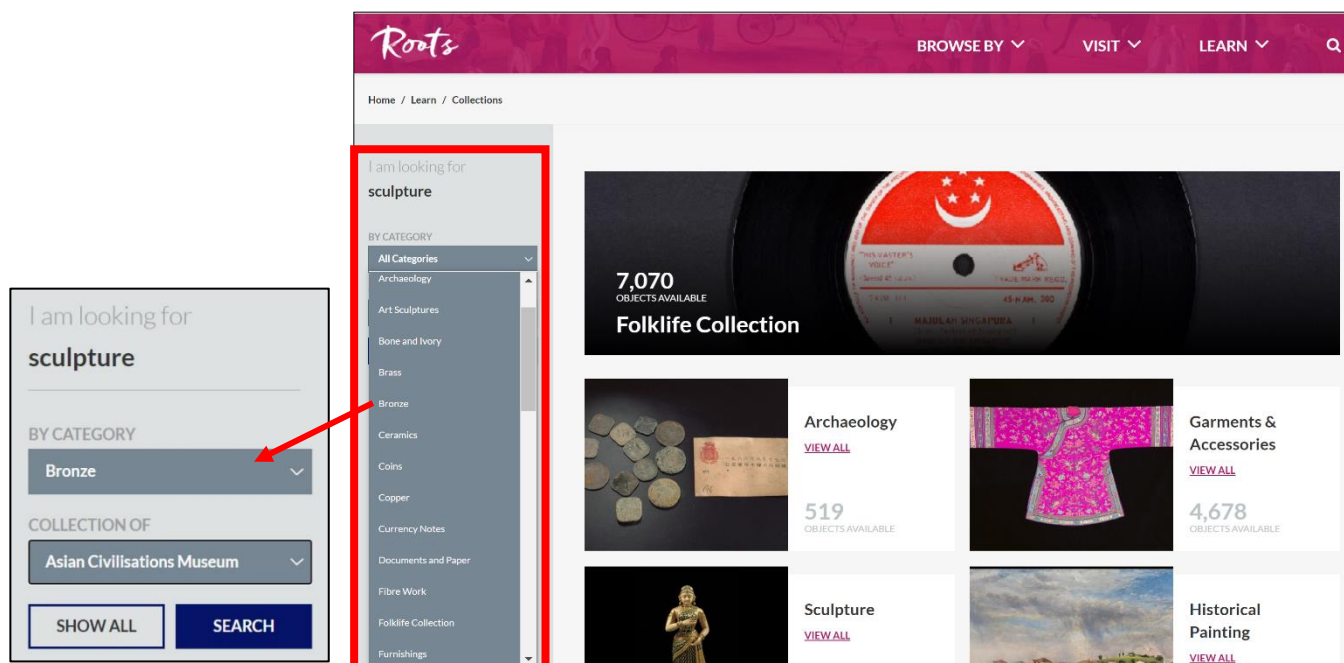


Image 1: Category filter on former Roots (2016 – Nov, 2020)

Artefact image credit: Courtesy of Asian Civilisations Museum and the National Museum of Singapore, NHB

The third challenge, which is related to the second challenge mentioned above, was that the simple NC search page on Roots then had only two filters, which are, “Category” and “Collection” (museum names) as shown in Image 1 above. This made narrowing of searches a challenge. However, these were the only two fields with the most consistent data that could be used as filters back then.

3. Use of controlled vocabulary from international and local thesauri

With the understanding that controlled vocabulary helps in standardisation and organisation of data, the team explored several established controlled vocabularies and thesauri available online and considered the suitability of the terms in the thesauri.

As Art & Architecture Thesaurus (AAT) is used for describing items relating to fine art, architecture, decorative arts, archival materials and material culture (The Getty Research Institute, no date), hence it is suitable for the NC. The Library of Congress Name Authority (LCNA) was selected for its proper nouns, such as names of places and people, especially Southeast Asian creators' names. The Library of Congress Subject Headings (LCSH) was added to describe the context and subject matter of the artefacts better. Thesaurus of Geographic Names (TGN) which has context for locations was later included. In 2016, National Library Board (NLB), Singapore's Knowledge Organisation Systems (KOS) terms which are predominantly Singapore-centric was adopted. Table 1 shows examples of metadata fields and the corresponding controlled vocabularies/ thesauri used when assigning metadata (terms) to them (Ho, 2019, p. 255).

Elements (Fields)	Controlled Vocabulary/ Thesaurus (source of the metadata used)
Object/Work Type	AAT, NHB Controlled Vocabulary
Creator	LCNA, KOS
Creation Date	ISO 8601 format
Creation Place/ Original Location	TGN
Styles/Periods Indexing Terms	AAT - Styles and Periods Facet
Materials Name	AAT- Materials Facet
Subject	AAT, LCSH, KOS, NHB Controlled Vocabulary

Table 1: Examples of fields and corresponding controlled vocabularies/thesauri

4. Challenges faced with implementation of controlled vocabularies

With that, the proper heritage cataloguing work started with the anticipation that this will enhance the metadata of the NC records. However, the records were catalogued on MS Excel sheets at the initial years while awaiting the cataloguing functionalities to be built into the SCMS. It was a challenge then to visualise the standardisation of terms and retrieval of records on the many individual Excel sheets.

Between 2017-2018, when a sizeable number of records had been catalogued, there was a realisation that while the usage of controlled vocabulary from thesauri standardised a few terms, it did not help to standardise the application of most of the terms. This was because each Cataloguer would attribute terms based on his/her own domain expertise, experience, perspective and interpret the given scope of use differently. The differently selected controlled vocabulary by each Cataloguer were not inaccurate, however, they did not help in standardising and organising the objects. For example, a sculpture of Buddha – using the AAT, some were tagged as “sculpture (visual work)”, others as “Buddha (visual work)” and “reliefs (sculptures)” by different Cataloguers. This would affect the retrieval of Buddha sculptures as a whole. We tried a few ways to circumvent this by having meetings to discuss the terms we had used for each object type and by listing them in groups. However, this was not manageable due to vast variety and number of object types we had in the collection.

Besides that, while cataloguing, the Cataloguing Team was faced with the challenge of describing our Southeast Asian artefacts more accurately as the terms in the established international controlled vocabularies and thesauri were too general for these. These artefacts are known by their common indigenous names, such as “congkak”, a local traditional game, which otherwise would just be generally tagged as “games” following AAT. Hence, it was essential to develop our own local terms (NHB Controlled Vocabulary) in order to accurately describe the artefacts. This realises Teh’s (2005, p. 5) statement that HCC then had plans to start a working group to develop a thesaurus but was not able to do so.

5. Development of NHB Controlled Vocabulary and NHB Object Type Taxonomy

In 2018, inspired by some taxonomy modules we saw during a work trip to Amsterdam and London, we started thinking about “controlling” the controlled vocabularies used by all of our Cataloguers using a taxonomy. Starting small, we decided to work on the most popularly used field for search in the collection, that is the Object Type field. A taxonomy supports search via the relationship of synonyms, variant terms, faceted search, and browsing of a collection using hierarchies which aids with better organisation of data (Hedden, 2016, pp. 3-15). The Cataloguing Team engaged the services of a knowledge management consultant to impart knowledge on taxonomy development and to guide us for a few sessions when we started developing the NC object type taxonomy in 2019. It was new and a challenge to us due to the huge number and the vast variety of the collection, but we found it

No image in this paper is to be used without the express permission of the National Heritage Board, Singapore.

beneficial for data organisation. The Cataloguing Team generated lists of artefacts from the SCMS, studied them and tried to group the artefacts after the taxonomy workshop. We learnt about and worked on --hierarchy levels required; categorising, sorting and grouping the artefacts by open card sorting method; labelling categories; testing the suitability of the categories by close and open card sorting; linking preferred terms with synonyms and related terms; and creating of scope notes for local terms created for the NHB Controlled Vocabulary which are fitted into the taxonomy (Ho, 2019, p. 258).

In 2017, there was an initial attempt to create the NHB Controlled Vocabulary but was short lived due to the lack manpower and experience issues in this area. However, after gaining knowledge through developing the object type taxonomy, development of the NHB Controlled Vocabulary commenced again in 2020. We referenced the format based on AAT's. Presently, we have created terms for object type and subject fields. These terms are created in the Vocabulary Module in the SCMS and attributed to the artefacts as well. A guide on the criteria of creating local terms was developed. Each term has scope note which will give the definition of the term and the scope of use. There are plans to share the NHB Controlled Vocabulary with other institutions when it is well established.

In response to Teh's (2005, pp. 2-3) concern that native descriptors might limit the retrieval of the Southeast Asian artefacts, we learnt that this would not happen as native descriptors can be linked to the displayed (preferred) term, for example – "kamcheng" which is a Peranakan (Straits Chinese) pot and can be linked to "containers". A "kamcheng" can be tagged with its indigenous name and still be retrieved by non-native users using the term "containers" as the two terms are linked (Ho, 2019, p. 257). Besides indigenous names, variant spellings and synonyms of a term can also be linked to the term itself to aid search at the backend and will cast a wider yet accurate net during retrieval.

Presently, all HCC Cataloguers must select object type terms from the NHB Object Type Taxonomy to tag to the artefacts. The terms consist of AAT and NHB Controlled Vocabulary terms. The taxonomy structure is developed in the SCMS and we were able to visualise better the organised data and retrievals when we started cataloguing into the SCMS in 2019. Meanwhile, for the other fields, the Cataloguers who vet the records discuss amongst each other to ensure the consistency of the terms tagged to the artefacts and not clutter the data in the system again. As the Cataloguers got used to the guidelines and house rules for tagging, the challenges began to ease.

6. Outcome of the Use of Controlled Vocabulary in SCMS

The use of controlled vocabulary and taxonomy enhanced the search and retrieval in our internal database and online portal Roots, as well as the browsing feature.

Inspired by Bloomsbury Fashion's and online shopping browse features, the Cataloguers worked with the Strategic, Communications and Digital (SCD) department colleagues and vendors to create the Collections Classification page on Roots which features the NHB Object Type Taxonomy for browsing. This is enabled by having controlled vocabulary in the system to standardise terms and a taxonomy to guide the attribution of object type terms. The Collections Classification page is a work-in-progress which presently consists of approximately 3,000 artefact records but updates are being made to add more records as they get tagged with the terms from the taxonomy. (See *Image 2: Collections Classification page on Roots*)

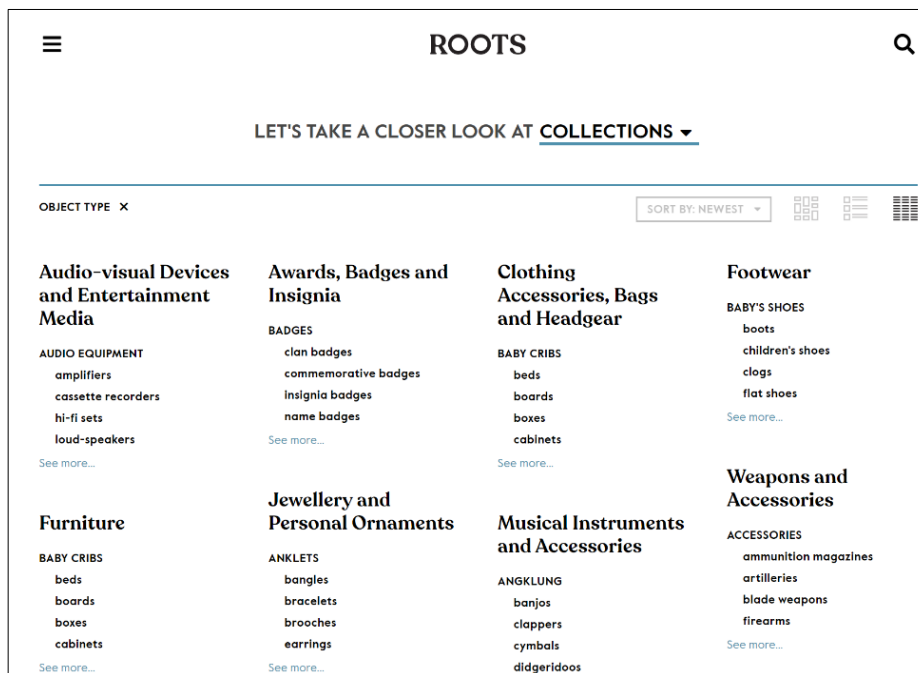


Image 2: Collections Classification page on Roots which is based on the NHB Object Type Taxonomy (URL: <https://www.roots.gov.sg/filter/collectionresearch>)

The new search feature on the Collections page is able to include new filters to narrow down search results with the use of controlled vocabulary using the new framework. Image 3 below shows the new filters for the new Collections search page on the right, vis-à-vis the former one on the left which had only two filters.

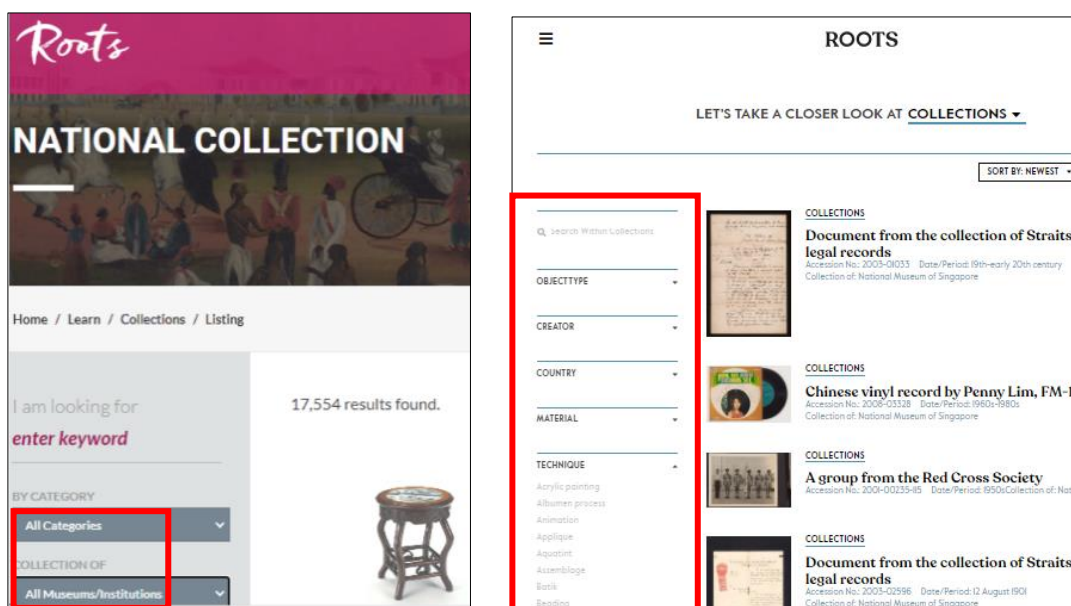


Image 3: Comparison of filters between the former and the new Collections page on Roots: 2016 (left) and 2020 (right)
Artefact image credit: Courtesy of Asian Civilisations Museum and the National Museum of Singapore, NHB

No image in this paper is to be used without the express permission of the National Heritage Board, Singapore.

With the data more organised and structured, it enhanced the retrieval in the SCMS. (See Image 4: *Retrieval in SCMS*) The retrieval is based on the hierarchies in the NHB Object Type Taxonomy. For example, game cards, can be retrieved by the label “card games” or by the “Games, Toys and Sports Equipment” category as a whole. Prior to the creation of the taxonomy and controlled vocabulary for object type, the game cards would have been “drowned” in the Folklife Collection (a category filter) that consists of 40,960 artefacts. However, with the use of the taxonomy and controlled vocabulary, the retrieval is narrowed down to 73 card games and they are differentiated from other types of cards in the collection as well. This results in more efficient retrievals.

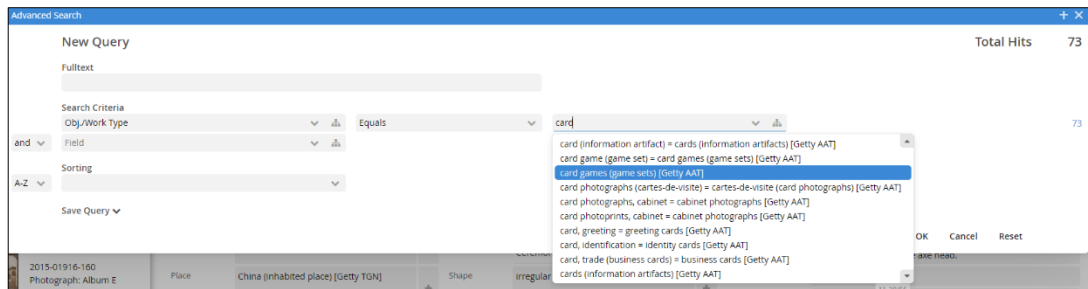


Image 4: Retrieval in SCMS

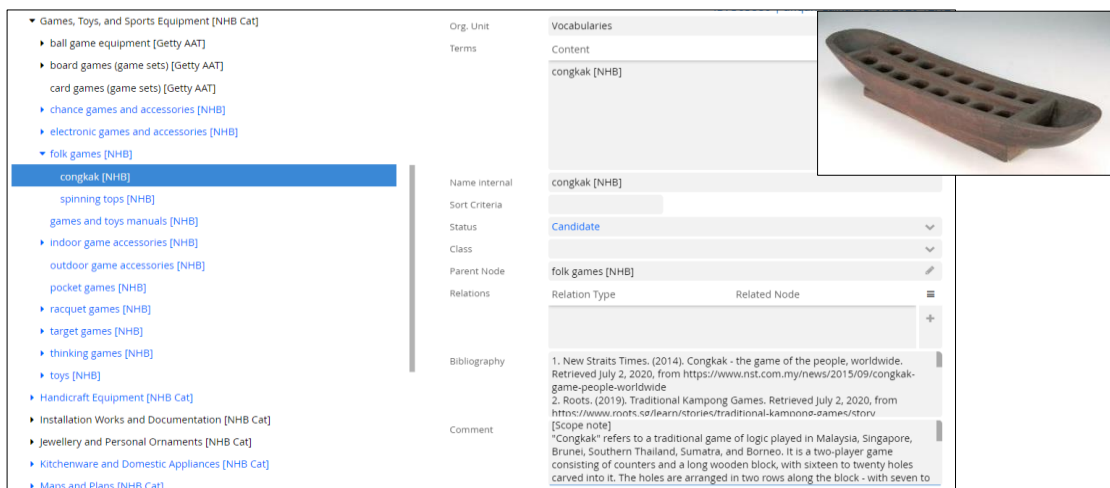


Image 5: NHB Controlled Vocabulary and NHB Object Type Taxonomy in SCMS

We are also able to create as well as link Malay and Indonesian local terms to English terms when we created the NHB Controlled Vocabulary which aids in a more accurate and wider retrieval. (See Image 5: *NHB Controlled Vocabulary and NHB Object Type Taxonomy in SCMS*) “Congkak” which is a Malay name for a popular local traditional game does not have an English equivalent name and is not found in any of the thesauri we are using. Users who know the native name for this game, can now retrieve the artefacts directly using the word “Congkak” or “folk games” in the object type field, instead of filtering through the many records in the Folklife Collection.

7. Conclusion

The use of controlled vocabulary and taxonomy enabled new features to be created on Roots to enhance the browsing, search and retrieval of the NC. We are still at a learning phase and the work is still in-progress due to the huge number of objects in the NC and is every-growing. The processes of developing controlled vocabulary and taxonomy are time consuming but they are a worthwhile investment of effort as they increase the search-ability and accessibility of the NC records; providing users with a better search experience. Moving forward, AI is a tool we are looking to explore to aid in this work for the future.

No image in this paper is to be used without the express permission of the National Heritage Board, Singapore.

References:

Hedden, H. (2010) *The accidental taxonomist*. Information Today, Inc., Medford, New Jersey.

Ho, S. A. (2019) 'Heritage cataloguing: the HCC experience' in *Collections care: staying relevant in changing times, Asian and beyond*. National Heritage Board, Singapore, pp. 249-261. Available at: <https://www.roots.sg/learn/resources/publications/Heritage-Conservation-Centre/collections-care> (Accessed 1 September 2020)

Low, J. T. & Doerr, M. (2010) 'A postcard is not a building: why we need museum information curators', paper presented at the ICOM International Committee for Documentation (CIDOC) Conference, Shanghai, China, 7th - 12th November. Available at: <http://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2018/12/low.pdf> (Accessed 31 August 2020)

Teh, E. E. F. (2005) 'A custodian's challenge: a museum documentation standard for all? Experience from Heritage Conservation Centre (HCC), Singapore', paper presented at the ICOM International Committee for Documentation (CIDOC) Conference, Zagreb, Croatia, 24th – 27th May. Available at: <http://cidoc.mini.icom.museum/archive/past-conferences/2005-zagreb/> (Accessed 31 August 2020)

The Getty Research Institute (no date) *Getty Vocabularies* [Online]. Available at: <http://www.getty.edu/research/tools/vocabularies/> (Accessed 16 September 2020)

-End of document-