

# Multilingual Metadata: Aligning Digital Heritage Systems with Cultural Values

Robin Dresel<sup>1</sup>   and Pamela Low<sup>1</sup>  

<sup>1</sup> National Library Board Singapore

## Abstract

Metadata systems traditionally prioritise technical efficiency over cultural authenticity. This paper examines how Singapore's National Library Board redesigned metadata practices for two major cultural heritage digital projects—the Encyclopedia of Singapore Tamils (EST) and Prominent Malays of Singapore (PMoS)—to align digital systems with community values.

Breaking from locally established conventions, we implemented largely monolingual metadata records, creating separate collection identifiers, language-specific navigation paths, and culturally-aware controlled vocabularies. This approach required overcoming technical constraints in content management systems designed for English-language dominant workflows.

Our methodology involved close collaboration with over 500 community contributors. Key innovations include collection name separation, multilingual controlled vocabulary integration, and community-driven category translation that reflects cultural mental models rather than literal translations.

Implementation demonstrates that separate monolingual metadata records can preserve cultural authenticity while maintaining system functionality. These approaches create valuable training data for AI systems working with multilingual cultural heritage, offering a replicable model for institutions seeking to align digital systems with diverse community needs rather than technical convenience.

## 1. Introduction

Metadata systems serve as gateways to information resources, enabling users to discover and access textual, visual, and auditive materials. As intermediaries between users and content, these systems ideally empower informed resource selection for diverse informational needs.

However, metadata systems are inherently biased, reflecting both creator experiences and embedded value systems [1]. The Dewey Decimal Classification's Christian focus and English cataloguing dominance exemplify such bias.

Singapore exemplifies this challenge through its practice of cataloguing key access points of resources in their source languages. While preserving linguistic authenticity—Tamil resource titles, Creators, Publishers and other fields are being catalogued in Tamil, the respective fields for Malay resources in Malay, Chinese and English, the other two official languages, respectively - this approach creates access barriers for users who might benefit from resources without full language proficiency.

Resource constraints make creating metadata in all four official languages impractical, leading to the current practice of maintaining English-only website interfaces—reflecting the role of English as Singapore's business language. This challenge reflects broader patterns where digital libraries are rarely multilingual, despite users expecting multilingual access to digital libraries [2].

## Short Paper

Available Aug 01, 2026

Correspondence to  
Robin Dresel  
robin\_dresel@nlb.gov.sg



Copyright © 2026 Dresel & Low. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Against this backdrop, NLB co-created two projects: the Encyclopedia of Singapore Tamils (EST) and Prominent Malays of Singapore (PMoS). These portals document their respective communities' cultural heritage while, for the first time, offering fully bilingual access—EST in English and Tamil, PMoS in English and Malay.

Providing separate monolingual metadata accesses for these resources became essential to support this vision. However, what appeared straightforward encountered technical challenges from systems designed for monolingual cataloguing. This paper examines how we overcame these limitations and the insights gained through reimagining metadata practices for cultural equity.

## 2. Value-Driven Requirements Analysis

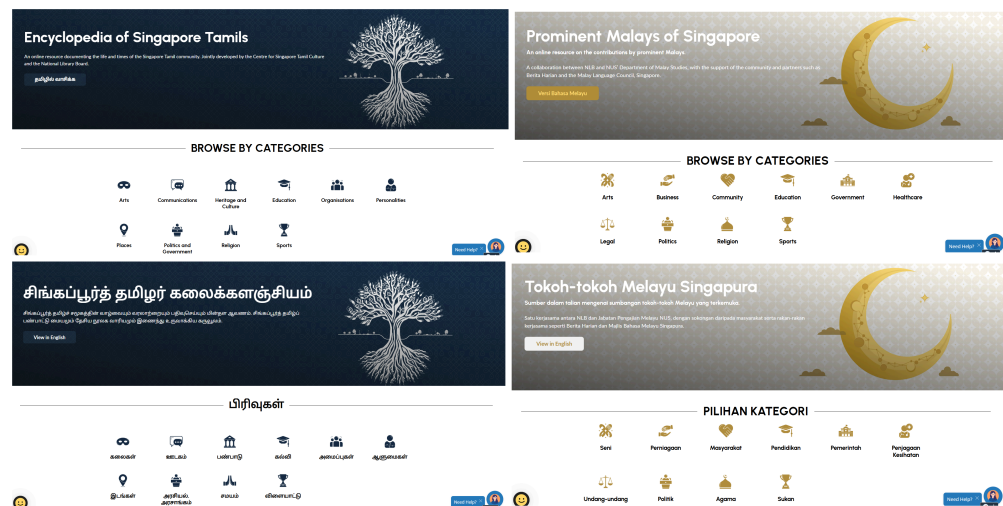
The Encyclopedia of Singapore Tamils (EST) and Prominent Malays of Singapore (PMoS) represent community-driven cultural documentation projects. EST, was launched in 2025 through collaboration between the Centre for Singapore Tamil Culture and NLB. The project documents “the life and times of the Singapore Tamil community” while serving broader audiences interested in Tamil cultural contributions to Singapore. This dual purpose - preserving community heritage while fostering intercultural understanding - necessitated separate monolingual presentations in both Tamil and English. PMoS, launched in short succession, documents “contributions by prominent Malays” through collaboration between NLB and the National University of Singapore's Department of Malay Studies, supported by community partners including Singapore's Malay-language newspaper *Berita Harian* and the Malay Language Council.

Both platforms exemplify intergenerational knowledge transfer while serving as cultural gifts to Singapore's multicultural society.

The crucial innovation was creating truly separate monolingual representations rather than English-dominant versions with translations. To that end, fully monolingual landing pages were created for EST and PMoS, allowing the reader to enter the collection in a completely English or entirely Tamil/Malay website environment. The landing pages display ten topical browsing categories, whose Tamil and Malay versions were carefully chosen and translated through community consultation.

Involving the community in the project is a critical step to counter biases, as the importance of cultural knowledge in metadata creation cannot be overstated [3]. Examples of cultural differences are manifested in metadata through various elements, from fields for names that vary between languages like English and Japanese, to decisions about transliterations and the use or lack of diacritics [1], [3].

NLB's content management system generates webpages from metadata using predefined templates before the content is added. Enabling vernacular subject access necessitated creating both articles and metadata in each respective language. However, recognising limited Tamil content availability, the team made a pragmatic decision: Tamil articles would include English equivalent subjects, acknowledging the community's



**Figure 1.** Landing pages for EST (left column) and PMoS (right column) in English (top) and Tamil and Malay respectively (bottom). The English and vernacular landing pages mirror each other and are fully monolingual.

effective bilingualism. The same was done for the Malay articles. This careful balance expanded access while respecting linguistic preferences and content realities.

Having established the community-driven requirements for authentic cultural representation, the following section examines how these values translated into specific technical implementation decisions and the challenges encountered in adapting existing library systems.

### 3. Technical Implementation of Value-Aligned Metadata

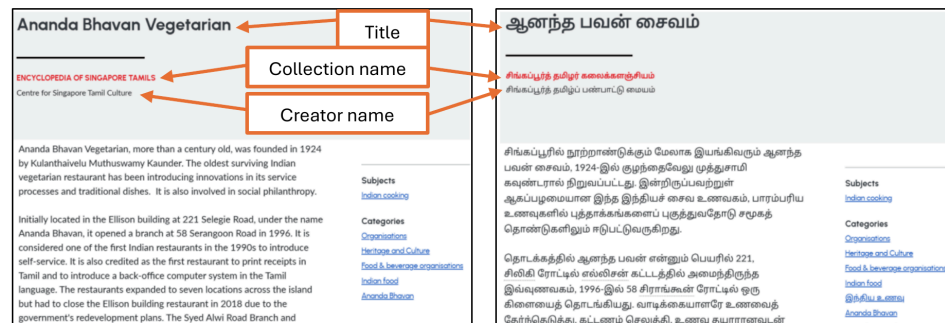
NLB's content management system generates webpages dynamically from metadata records using predefined templates. This architecture, designed for English-dominant collections, required fundamental modifications to support the dual-monolingual approach. The following technical analysis focuses primarily on the EST for clarity, with both platforms sharing identical practices and architectural constraints. Differences with PMoS are noted where relevant.

#### 3.1. Breaking Convention for Cultural Equity

Traditional double-barrelled approaches aim to broaden resource discovery, making materials accessible through both vernacular and English searches, typically with language faceting. However, common practices for multilingual metadata have not yet emerged [4], and different institutions adopt varying approaches—for instance, Library and Archives Canada's Bilingual Cataloguing Policy proposes to describe resources in one or two languages [4].

To enable immersive language experiences, NLB implemented truly monolingual articles encompassing collection names, creator names, and embedded bilingual resource captions.

With articles existing in both languages, double-barrelled entry points, where titles are recorded in both the source language and English, became redundant. While double-



**Figure 2.** Webpage display for English EST article *Ananda Bhavan Vegetarian* (left) and the Tamil version *ஆனந்த பவன் சைவம்* (right), showing parallel website structure.

barrelled approaches enhance vernacular language discoverability, they create noise in monolingual searches by including unwanted multilingual results. The monolingual approach improved search precision: users searching *ஆனந்த பவன் சைவம்* would find Tamil versions, while English searches for *Ananda Bhavan Vegetarian* would yield English versions exclusively. However, acknowledging the projects' dual monolingual nature, each article includes links to its counterpart language version, maintaining connections between monolingual representations. Implementing this approach required creating parallel website structures for both languages. The alternative of using `xml:lang` denotations would have required substantive changes to the frontend and backend solutions and was therefore left for future enhancements.

This separation extended to collection names identifying all project-related articles. The dual monolingual approach required cataloguing of articles separately in their respective languages, necessitating language-specific collection identifiers: “Encyclopedia of Singapore Tamils” and “சிங்கப்பூர்த் தமிழர் கலைக்களஞ்சியம்” for EST; “Prominent Malays of Singapore” and “*Tokoh-tokoh Melayu Singapura*” for PMoS.

### 3.2. Human-Centered Navigation Architecture

The multilingual approach directly reflects users' cultural mental models. The category ‘Arts’—translated as *கலைகள்* in Tamil and *Seni* in Malay—exemplifies how identical concepts carry different cultural connotations across languages. These differences reflect deeper challenges where concepts that cannot be translated between languages, such as “ant nests” in Finnish versus “ant hills” in English, or where concepts don't exist in another language, or are more specific in another language [5]. While English ‘arts’ typically emphasises individual creativity and achievement, Tamil and Malay versions highlight artists within cultural environments, often incorporating spiritual contexts absent from Western secular frameworks, such as Indian Dance. As Matusiak et al. note, ‘the language of discipline and access point does matter, as it is a key to resource discovery.’ [2]. However, it is a considered choice, since Language appropriate terms require maintaining equivalencies [4], adding to existing work.

### 3.3. Multilingual Controlled Vocabularies

Extending the principle of two separate monolingual versions, controlled vocabularies required translation, prioritising higher-frequency terms to enable category-based

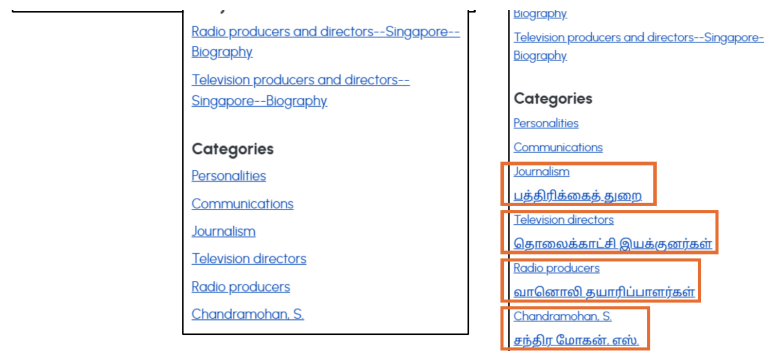


Figure 3. Local vocabulary terms tagged to the English and Tamil versions of an EST article. English-

**Figure 3.** Local vocabulary terms tagged to the English and Tamil versions of an EST article. English-language articles are tagged with English terms only, while Tamil-language articles are tagged with both English terms and, for key terms, their Tamil variants as well.

browsing. We expanded controlled vocabularies to include Malay and Tamil versions alongside English variants within the local browsing taxonomy. Cataloguers then assigned both English and the vernacular terms for key terms to respective language article versions while retaining only the English variants for English articles.

This approach aligns with vocabulary mapping as an approach for controlled vocabularies [2], though it requires careful curation. As Riva observes, ‘when subject heading mappings have been carefully curated by bilingual cataloguers and the subject heading languages are compatible in structure, the results can be very good.’

However, implementation required retaining English subject terms within Tamil-language pages, as Tamil terms point exclusively to Tamil content, which is in limited quantity and would disadvantage users by omitting access to related English-language materials, which dominate NLB’s collections.

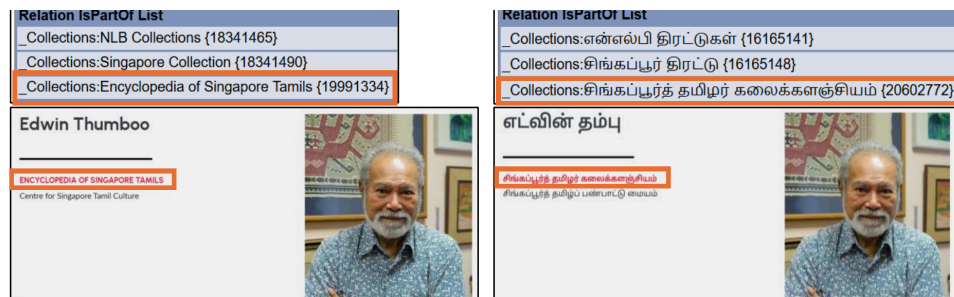
These vocabulary translations have implications beyond immediate project needs. Culturally sensitive translations from these projects provide valuable training data for AI systems working with multilingual cultural heritage. The translations represent high-quality resources for developing culturally aware AI applications.

#### 4. Challenges and Future Directions

Achieving fully monolingual articles required more than bilingual content creation. Since NLB’s content management system relies on metadata to generate website structure, the metadata itself required augmentation to overcome system constraints designed for monolingual workflows.

While titles were catalogued in article languages, site sub-headers defaulted to displaying content type - “Articles” - for these projects. Displaying the project names required developing logic to retrieve the collection name in the correct language. This necessitated moving away from the conventional use of double-barrelled collection names such as “Encyclopedia of Singapore Tamils சிங்கப்பூர்த் தமிழர் கலைக்களஞ்சியம்” to establishing English and vernacular collection names separately. When English articles are tagged with the English collection name and Tamil articles with the Tamil version,





**Figure 4.** English and Tamil monolingual collection names tagged in the *isPartOf* field (top) of the English and Tamil article records, retrieved by the frontend system and displayed on the respective article webpages (bottom) for clear branding.

in `dcterms:isPartOf` field, the site can retrieve and display the name matching the article language.

The second challenge involved achieving monolingual image captions from “Description” fields. The preferred method of using language qualifiers would have required extensive database schema modifications affecting system-wide functionality, which was not achievable within the scope of these projects.

The third constraint involved displaying National Archives of Singapore (NAS) images within articles. The predefined webpage template allows images from the same database as the articles to be displayed in the articles, by adding the image UIDs to the articles’ `dcterms:hasPart` fields. However, the database contained only images previously prepared for online use through public requests, whose display is restricted to those requests. Unprepared images remained in NAS’ database (a different database), whose UIDs are not recognised by the retrieval logic of the `dcterms:hasPart` field of the articles’ database. The eventual decision, which upheld NAS’ image usage restrictions, was to ingest and catalogue the preview-resolution, watermarked images. To prevent these images from being individually discoverable within NLB’s website ecosystem, their records included metadata feeding a logic to suppress the generation of webpages.

Given the resource intensity of article cataloguing, AI-assisted metadata generation presented potential value. However, current English-only explorations demonstrate that AI outputs require librarian validation, yielding no immediate efficiency gains. This aligns with broader findings that results of automatic translations are rather crude, machine translation accuracy is limited, and all natural language fields require human verification [2]. Future improvements in AI reliability may warrant revisiting this approach for multilingual applications.

## 5. Conclusions: Towards Culturally Responsive Metadata Systems

The Encyclopedia of Singapore Tamils and Prominent Malays of Singapore demonstrate how libraries can transcend traditional cataloguing practices to better serve multi-cultural communities. While showcasing community-driven content creation, these projects primarily exemplify institutional commitment to cultural equity through truly

monolingual article versions that provide immersive language experiences while maintaining interlingual connections.

The implementation revealed substantial challenges in adapting monolingual library systems for multilingual metadata. Technical constraints necessitated pragmatic compromises such as bilingual image captions and hybrid English-vernacular category structures.

The technical solution required metadata innovation, abandoning locally practiced double-barrelled cataloguing for separate monolingual representations. This approach enhanced search precision while preserving cultural authenticity and linguistic preferences. Developing multilingual controlled vocabularies and culturally informed category translations further supported authentic user experiences across language communities.

Future research priorities include: (1) conducting user studies comparing monolingual versus double-barrelled approaches for cultural heritage discovery, (2) developing cost-benefit frameworks for institutions considering similar implementations, (3) implementing technical standards for multilingual content management systems, such as SKOS-XL, and (4) investigating AI-assisted translation validation for culturally sensitive terminology.

## References

- [1] D. Donathan, M. Nason, M. Tullney, J. Shi, and J. P. Alperin, "Evaluating Multilingual Metadata Quality in Crossref," arXiv, 2025. doi: [10.48550/arXiv.2503.11853](https://doi.org/10.48550/arXiv.2503.11853).
- [2] K. K. Matusiak, M. Ling, E. Barczyk, and C.-J. Shih, "Multilingual Metadata for Cultural Heritage Materials: The Case of the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings," *The Electronic Library*, vol. 33, no. 1, pp. 136–151, 2015, doi: [10.1108/EL-08-2013-0141](https://doi.org/10.1108/EL-08-2013-0141).
- [3] C. Jilovsky, L. Sukkar, and E. Varga, "Multi-lingual Cataloguing: culture, practice and systems." 2008.
- [4] P. Riva, "The Multilingual Challenge in Bibliographic Description and Access," *JLIS*, no. 1, 2022, doi: [10.4403/jlis.it-12737](https://doi.org/10.4403/jlis.it-12737).
- [5] S. Niininen, S. Nykyri, and O. Suominen, "The Future of Metadata: Open, Linked, and Multilingual – the YSO Case," *Journal of Documentation*, vol. 73, no. 3, pp. 451–465, 2017, doi: [10.1108/JD-06-2016-0084](https://doi.org/10.1108/JD-06-2016-0084).