

PROJECT REPORT

The South Asian Canadian Digital Archive Thesaurus: Thesaurus Construction for the South Asian Canadian Diaspora

Magnus Berg

South Asian Studies Institute, University of the Fraser Valley

Satwinder Kaur Bains

South Asian Studies Institute, University of the Fraser Valley

Sadhvi Suri

South Asian Studies Institute, University of the Fraser Valley

The South Asian Canadian Digital Archive (SACDA) is a soon-to-be-released digital repository developed by the South Asian Studies Institute at the University of the Fraser Valley, located in Abbotsford, British Columbia, Canada. SACDA partners with memory institutions, individuals, families, and organizations to digitize, describe, and provide online public access to heritage materials created by, or relevant to, the South Asian Canadian diaspora. This project report will detail how SACDA is building a customized thesaurus to classify its digitized archival holdings, augment existing subject headings and thesauri, and fill in taxonomical gaps. Building on prior work done by alternative thesauri like the Homosaurus, Association for Manitoba Archives Indigenous Subject Headings, Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) Thesauri, and the International Thesaurus of Refugee Terminology, among others, the SACDA thesaurus intends to fill in a vital gap in South Asian Studies subject control, particularly from a Canadian perspective.

Keywords: South Asian studies; thesaurus construction; digital repositories

Introduction

The South Asian Canadian Digital Archive (SACDA) is a forthcoming digital repository developed by the South Asian Studies Institute (SASI) at the University of the Fraser Valley in Abbotsford, British Columbia, Canada. SACDA partners with memory institutions, individuals, families, and organizations to digitize, describe, and provide online public access to heritage materials created by, or relevant to, the South Asian Canadian diaspora. SACDA uses the Library of Congress (hereafter LoC) Subject Headings (LCSH) and Thesaurus for Graphic Materials (TGM) to assign subject headings to records; however, the terms available to describe South Asian Canadian diasporic archival material are often lacking or outright non-existent. In many instances, broad headings for describing materials in SACDA are not available in LCSH and TGM. For example, while LCSH does have headings for Muslim or Christian gay men, there is no heading available for Sikh or Hindu gay men. In other cases, we require headings specific to our records or local context, such as Jor Malla, a gathering festival specific to Paldi, a small logging town on Vancouver Island. While it is evident why this festival is not included in LCSH, its presence and significance in our records means that it warrants a subject heading in our repository.

To address these gaps, SACDA created a custom thesaurus relevant to the South Asian Canadian diaspora, to be used in conjunction with LCSH and TGM. The thesaurus construction is modeled after similar community-centered projects like the Homosaurus, which functions as a companion to existing thesauri rather than an outright replacement (Zwaaf 2020, 207–8). Alternative thesauri, such as the Homosaurus, provide

cataloguers and information professionals with alternatives when dominant vocabularies like LCSH are insufficient to describe subjects that are culturally or locally specific. The SACDA thesaurus aims to fill in a similar gap. The thesaurus has been under construction by a small team of temporary project staff since October 2020. While it was initially conceived as a small controlled vocabulary limited to a few hundred concepts, it has since expanded in response to the limited concepts available in alternative vocabularies and to better classify the records contained within SACDA. This project report details why and how the SACDA thesaurus is being developed and where we hope to take the thesaurus in the coming years.

The Importance of the SACDA Thesaurus

Ever since the publication of Berman's (1971) *Prejudices and Antipathies*, the presence of oppressive structures, misdescription, offensive and racist terminology, and erasure by dominant and colonial taxonomies and classification systems are subjects that have been widely covered in library and information studies (LIS) scholarship (Berman 1971; Knowlton 2005; Johnson 2010; Littletree and Metoyer 2015; Biswas 2018; Bone and Lougheed 2018; Howard and Knowlton 2018; Lo 2019; Watson 2020). The description of information resources for equity-seeking groups is a growing topic of research. Though the critical reappraisal of cataloguing practices is not a new phenomenon, critical librarianship and critical cataloguing have emerged in both LIS online communities and scholarship as a means of encouraging active application of critical theory in librarianship practice (Martin 2021; Schroeder and Hollister 2014). Despite this flurry of intellectual and professional activity, there are few alternatives to LCSH in North America, ultimately requiring institutions to choose between continuing to use subject headings that may do a disservice to their stakeholders or to design their own controlled vocabulary where one does not exist.

The lack of relevant subject headings for South Asian studies is attributable to its underrepresentation in LIS and in mainstream scholarship. As observed by Biswas (2017, 3), far fewer information professionals have a subject focus on South Asian studies than on queer studies, women's studies, or Black studies, which results in a lower number of subject access points and standards for South Asian studies. Since the LoC makes additions to the LCSH based on literary warrant—the philosophy that headings are only warranted when sufficient material on the topic exists—the amount of available South Asian studies scholarship also has a direct correlation to the number of relevant subject headings produced by LoC (Lo 2019).

Moreover, the terminology that LCSH uses to classify people from South Asia/the South Asian diaspora is inconsistent. While most subject headings in LCSH use some variation of *South Asian* or *South Asian American*, there are still around fifty headings that use *East Indian*, including *Foreign workers*, *East Indian*, *East Indian Americans*, and *East Indian diaspora*. In some cases, there are even synonymous headings for the same concept, such as *South Asian American teenagers* and *East Indian American teenagers*. The subject headings are also in some cases colonial. Biswas argues that the use of "East Indian" in LCSH headings "represents a problematic vestige of colonialism because the term's origins are inextricably tied to a European imagination that was shaped by the writings and experiences of maritime travelers, Christian missionaries, and colonial expeditions" (2017, 1). TGM suffers from a wholly separate issue, in that it is a far broader taxonomy and avoids explicit reference to social identity categories altogether. As contended by Adler, "any mention of race [in TGM] is veiled—it is only to be implied . . . [images] cannot be retrieved or found by way of subject headings that describe race or ethnicity" (2017, 123). Due to the inconsistency of South Asian/South Asian diasporic subject headings in LCSH and the relative dearth of headings available to describe a wide variety of subjects common in archival records documenting the South Asian Canadian diaspora, the SACDA team found it necessary to construct our own thesaurus to fill in these gaps, and in some cases replace existing LCSH or TGM headings.

First Phase Thesaurus Construction

Approach

SACDA is built by a small project team with limited funding, which limits the scope of work we could complete on the thesaurus during this first phase. The team is composed of a temporary archivist, who functions as project lead, in addition to student researchers and co-op students. Due to the nature of student labour, there was high turnover on the project team from term to term, and subject expertise was typically divided between computer science and information science. In an average term, half of the project team was composed of South Asian Canadians or recent immigrants or international students from the South Asian subcontinent, with the remaining half composed of European Canadian settlers. Team members relied heavily on the community knowledge of our colleagues on the project team proper, on the advisory committee set up for this project, and within the larger research staff at SASI. While we initially set out to create a very small list of controlled terms to augment LCSH and TGM, limited to a couple hundred

use within the diaspora, such as replacing the LCSH term *Sikh temples* with *Gurdwaras*, *Hindu temples* with *Mandirs*, and *Mosques* with *Masjids*. While the thesaurus, and the forthcoming SACDA portal more broadly, will primarily be in English, we sometimes created headings in Indic languages when it was deemed that there was not an appropriate equivalent in English, that using an English translation would dilute its meaning, or that our users would be more likely to search for a particular subject using an Indic language term rather than an English language term. Examples include *Sarna sthals*,⁴ which has no appropriate English equivalent, and wedding ceremonies or rituals like *Kanya Daan* (the giving of the girl to the groom) and *Vatna* (pre-wedding cleansing ceremony for the bride and groom), which would have been subsumed under overly broad headings like *Marriage service* or *Rites and ceremonies*.

Maintenance

The SACDA project team is made up of staff and student researchers in different disciplines and with different skill sets. As such, it was important for us to ensure that there was an approval process in place for the thesaurus to avoid issues like duplicate headings, synonymous headings, or inconsistent syntax. Further, since the thesaurus was being actively constructed in parallel to the rest of the repository's development, there was a lot of potential for records to be inconsistently classified. To ensure that our headings were created and applied in a consistent way, we adhered to the workflow depicted in Figure 1. This workflow allowed us to be nimble enough to create new subject headings on the fly, while also ensuring that our headings were created and applied in a consistent manner. This workflow also allowed us to incorporate multiple viewpoints and skill levels without compromising the organizational integrity of the thesaurus. The archivist's role as gatekeeper was primarily to avoid synonymous terms, duplicate terms, and inconsistent syntax rather than to question the validity or necessity of the proposed heading. Instead, the archivist assisted in revising or reshaping terms suggested by those without an LIS background so that they conformed to a common syntactical standard.

Classification of Identity

Classifying groups according to identity or geography has both positive and negative connotations. On the one hand, the classification of groups according to race, religion, sexual orientation, gender, class, caste, and ability levels has historically been used to perpetuate oppressive structures and systems against a particular identity group, up to and including genocide. For some, using any kind of identity category to classify records is a form of ghettoization. On the other hand, classification can also result in a sense of shared identity and belonging (Coleman-Fountain 2014; Porta et al. 2019; Rinderle and Montoya 2008; Kodama and Ebreo 2009). Not including certain identity categories in metadata records could be seen as a form of erasure and could ultimately make it difficult for users to retrieve records that deal explicitly with particular identity categories or that use disagreeable colonial nomenclature.⁵ Both of these viewpoints were taken into account when developing the SACDA thesaurus.

Ultimately, the SACDA team chose to create granular, identity-based thesaurus terms for two reasons: one, to facilitate information retrieval of specific, identity-based research topics that may be of interest to our users; and two, to create subject headings that could be used by other organizations whose collections are less focused on the South Asian Canadian diaspora. To avoid issues that can sometimes arise with identity-based subject headings, we attempted to avoid "neutral" identity headings, which often reinforce stereotypes about certain subjects and/or ghettoize marginalized groups. For example, when creating a new heading for women (e.g., South Asian Canadian women journalists), we generally made a parallel term for men (e.g., South Asian Canadian men journalists). This allowed our cataloguers to use a term specifically for male South Asian Canadian journalists rather than using a broader "South Asian Canadian journalists" heading as a "neutral" heading when describing records with male subjects. We also created terms for racialized groups outside of South Asian Canadians, such as East Asian Canadians and White Canadians. This avoided similar issues around kyriarchical ideas of "default," "neutral," or "dominant" headings, and will allow our users to create more targeted queries around intercultural relations.

Future Work

At present, the SACDA thesaurus is composed of over sixteen hundred subject headings and continues to be actively developed by members of the SACDA project team. The thesaurus is currently stored using our

⁴ Sarna sthals are sacred groves to practitioners of the Sarna religion.

⁵ While the SACDA Thesaurus makes a point to avoid colonial and outdated terminology, there are instances in our descriptions where we retain outdated/colonial/offensive/othering terminology when it is contextually relevant. This is done to facilitate the study of such language by scholars and to avoid erasing the history of white supremacy and colonization evident in archival records.

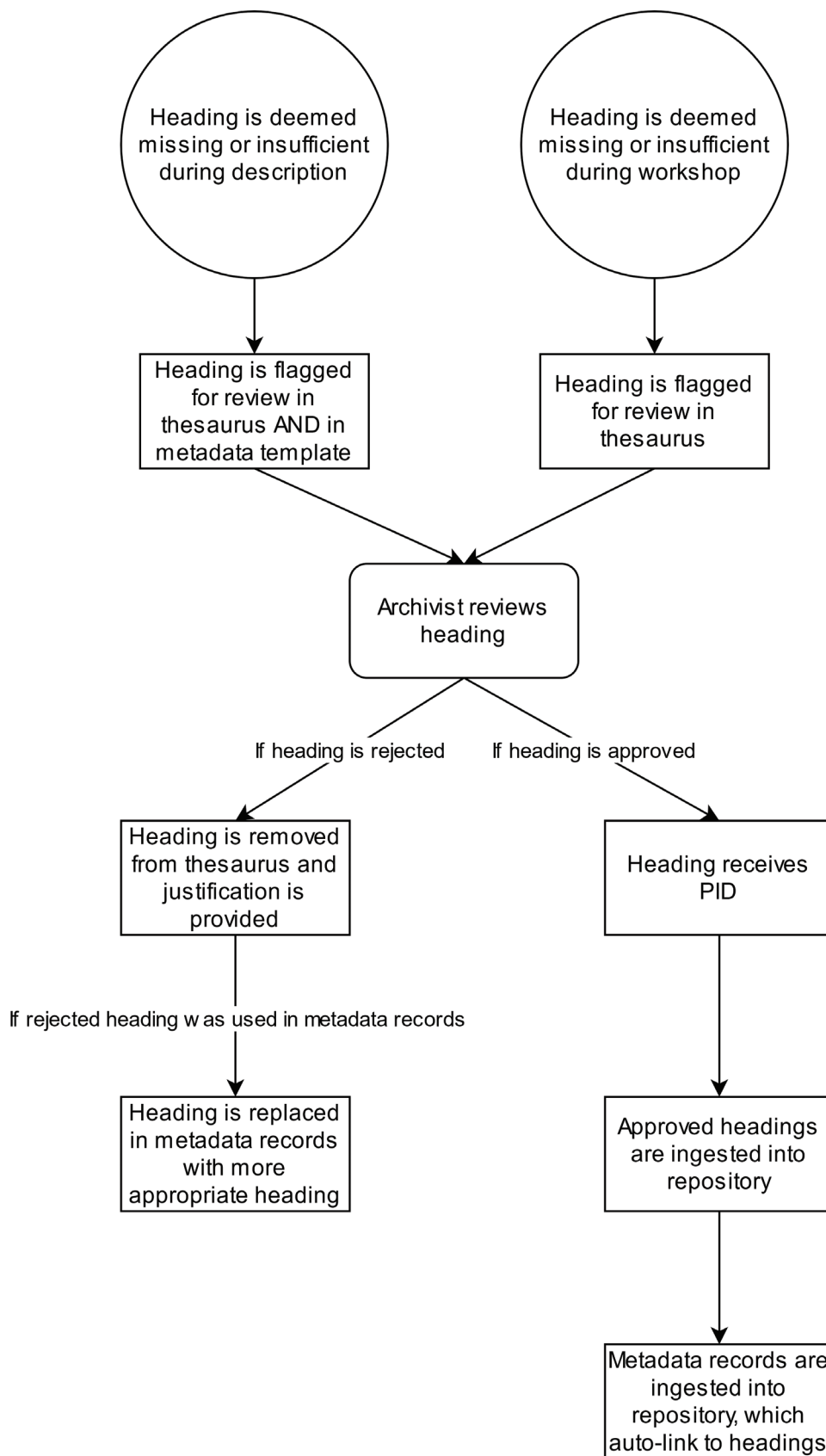


Figure 1: Flowchart illustrating the workflow used by the South Asian Canadian Digital Archive (SACDA) team to create new headings for the SACDA Thesaurus. New headings are flagged for review by the archivist, who either approves or rejects the heading. Rejected headings are replaced with more appropriate headings from the SACDA Thesaurus, LCSH, or TGM and justification is provided. Approved headings receive a persistent identifier (PID) and are subsequently ingested into the repository.

digital repository system, Collective Access. Because of the limited time and resources the SACDA team can devote to the development of the thesaurus, it has a simple, flat structure. Aside from the name of the heading itself, the only other information recorded with each heading is a persistent identifier. Although this simple structure works while the thesaurus and project team are relatively small, additional resources and structure would be necessary to scale the thesaurus. We hope to eventually provide more structure for the thesaurus and create linkages between subject headings, such as related headings and parent headings. We also hope to eventually convert the thesaurus to a format like RDF and make it available for use by other GLAM organizations.

Acknowledgements

The following people contributed to the development of the SACDA thesaurus during its first phase: Sadhvi Suri, Alisa Sohi, Dr. Satwinder Kaur Bains, Tim Ubels, Benjamin Arends, Tania Teixeira, Magnus Berg, and Thamilini Jothilingam.

Competing Interests

The authors declare that they have no competing interests.

References

- Adler, Melissa. 2017. *Cruising the Library: Perversities in the Organization of Knowledge*. New York: Fordham University Press.
- Berman, Sanford. 1971. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Metuchen: Scarecrow Press.
- Biswas, Paromita. 2017. "Rooted in the Past: Use of 'East Indians' in Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 56 (1): 1–18. <https://doi.org/10.1080/01639374.2017.1386253>.
- Bone, Christine, and Brett Loughheed. 2018. "Library of Congress Subject Headings Related to Indigenous Peoples: Changing LCSH for Use in a Canadian Archival Context." *Cataloging & Classification Quarterly* 56 (1): 83–95. <https://doi.org/10.1080/01639374.2017.1382641>.
- Canadian Museum of Immigration at Pier 21. n.d. "Chinese Immigration Act, 1923." Accessed November 16, 2021. <https://pier21.ca/research/immigration-history/chinese-immigration-act-1923>. Archived at: <https://perma.cc/6ZM6-DNNK>.
- Chan, Arlene. 2016. "Chinese Head Tax in Canada." *The Canadian Encyclopedia*. Last modified June 3, 2020. <https://www.thecanadianencyclopedia.ca/en/article/chinese-head-tax-in-canada>. Archived at: <https://perma.cc/5BB4-9ZQD>.
- Coleman-Fountain, Edmund. 2014. "Lesbian and Gay Youth and the Question of Labels." *Sexualities* 17 (7): 802–17. <https://doi.org/10.1177/1363460714531432>.
- Howard, Sara A., and Steven A. Knowlton. 2018. "Browsing through Bias: The Library of Congress Classification and Subject Headings for African American Studies and LGBTQIA Studies." *Library Trends* 67 (1): 74–88. <https://doi.org/10.1353/lib.2018.0026>.
- Johnson, Matt. 2010. "Transgender Subject Access: History and Current Practice." *Cataloging & Classification Quarterly* 48 (8): 661–83. <https://doi.org/10.1080/01639370903534398>.
- Knowlton, Steven A. 2005. "Three Decades Since *Prejudices and Antipathies*: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 40 (2): 123–45. https://doi.org/10.1300/J104v40n02_08.
- Kodama, Corinne M., and Angela Ebreo. 2009. "Do Labels Matter? Attitudinal and Behavioral Correlates of Ethnic and Racial Identity Choices among Asian American Undergraduates." *College Student Affairs Journal* 27 (2): 155–75. <https://eric.ed.gov/?id=EJ882661>.
- Littletree, Sandra, and Cheryl A. Metoyer. 2015. "Knowledge Organization from an Indigenous Perspective: The Mashantucket Pequot Thesaurus of American Indian Terminology Project." *Cataloging & Classification Quarterly* 53 (5-6): 640–57. <https://doi.org/10.1080/01639374.2015.1010113>.
- Lo, Grace. 2019. "'Aliens' vs. Catalogers: Bias in the Library of Congress Subject Heading." *Legal Reference Services Quarterly* 38 (4): 170–96. <https://doi.org/10.1080/0270319X.2019.1696069>.
- Martin, Jennifer M. 2021. "Records, Responsibility, and Power: An Overview of Cataloging Ethics." *Cataloging & Classification Quarterly* 59 (2-3): 281–304. <https://doi.org/10.1080/01639374.2020.1871458>.
- McRae, Matthew. n.d. "The Chinese Head Tax and the Chinese Exclusion Act." Canadian Museum for Human Rights. Accessed November 16, 2021. <https://humanrights.ca/story/the-chinese-head-tax-and-the-chinese-exclusion-act>. Archived at: <https://perma.cc/J7H8-VGYD>.

- Porta, Carolyn M., Amy L. Gower, Camille Brown, Brittany Wood, and Marla E. Eisenberg. 2019. "Perceptions of Sexual Orientation and Gender Identity Minority Adolescents About Labels." *Western Journal of Nursing Research* 42 (2): 81–9. <https://doi.org/10.1177/0193945919838618>.
- Rinderle, Susana, and Danielle Montoya. 2008. "Hispanic/Latino Identity Labels: An Examination of Cultural Values and Personal Experiences." *Howard Journal of Communications* 19 (2): 144–64. <https://doi.org/10.1080/10646170801990953>.
- Schroeder, Robert, and Christopher V. Hollister. 2014. "Librarians' Views on Critical Theories and Critical Practices." *Behavioral & Social Sciences Librarian* 33 (2): 91–119. <https://doi.org/10.1080/01639269.2014.912104>.
- Watson, Brian M. 2020. "'There was Sex but no Sexuality*': Critical Cataloging and the Classification of Asexuality in LCSH." *Cataloging & Classification Quarterly* 58 (6): 547–65. <https://doi.org/10.1080/01639374.2020.1796876>.
- Zwaaf, Katrina. 2020. "The Homosaurus <http://homosaurus.org>." *Technical Services Quarterly* 37 (2): 207–8. <https://doi.org/10.1080/07317131.2020.1728140>.

How to cite this article: Berg, Magnus, Satwinder Kaur Bains, and Sadhvi Suri. 2022. The South Asian Canadian Digital Archive Thesaurus: Thesaurus Construction for the South Asian Canadian Diaspora. *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 6(3). <https://doi.org/10.18357/kula.223>

Submitted: 24 June 2021 **Accepted:** 24 January 2022 **Published:** 27 July 2022

Copyright: © 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

KULA: Knowledge Creation, Dissemination, and Preservation Studies is a peer-reviewed open access journal published by University of Victoria Libraries.

OPEN ACCESS 