

**Julia Bullard** — The University of British Columbia  
**Nigel Town** — The University of British Columbia  
**Sarah Nocente** — The University of British Columbia  
**Aleha McCauley** — The University of British Columbia  
**Heather O'Brien** — The University of British Columbia

## **Thesaurus Construction for Community-Centered Metadata**

### **Abstract**

Community-engaged approaches to resource access require metadata practices that surface attributes relevant to local information needs and use terminology that reflects local language. This paper details the iterative and ongoing metadata work involved in facilitating access to aggregated items through the Downtown Eastside Research Access Portal. The challenges and strategies we describe here build upon and are relevant to knowledge organization projects seeking to repair issues of inaccurate and stigmatizing descriptive metadata for universal and local collections. After contextualizing the collection and the community, we describe our process in assessing areas of subject terminology in need of major repair, sources consulted for thesaurus terminology, and the approach we have taken to build a stand-alone thesaurus for this project, including our exploration and attempts at meaningful and respectful input into terms and term relationships.

### **Introduction**

Community-engaged approaches to resource access require metadata practices that reflect local information needs and language. This paper details iterative and ongoing metadata work in the context of a community-engaged project, the Making Research Accessible Initiative (MRAi). The MRAi began in the University of British Columbia's (UBC) Learning Exchange, an off-campus space located in the Downtown Eastside (DTES) of Vancouver which integrates community programming, student learning, and support for community-based research and knowledge exchange (Towle & Leahy, 2016). The MRAi grew into a formal partnership of UBC Library and UBC Learning Exchange, is focused on improving access to research for the DTES community and is attuned to the ways in which over researched, marginalized communities are often burdened by and negatively represented in university research (Tuck, 2009).

To mitigate issues of this research burden and extraction on the part of the universities, the MRAi developed a digital resource, the DTES Research Access Portal (RAP), that provides research articles, community reports, and research translations (e.g., summaries, infographics) and other information on topics relevant to the DTES (McCauley & Towle, 2022). This paper focuses on the description of these items for discovery and use. The challenges and strategies we describe here build upon and are relevant to knowledge organization projects seeking to repair issues of inaccurate and stigmatizing descriptive metadata for universal and local collections. After contextualizing the collection and the community, we describe our process of assessing areas of subject terminology in need of major repair, sources consulted for thesaurus terminology, and the approach we have taken to build a stand-alone thesaurus for this project, including our exploration and attempts at meaningful and respectful consultation and information-gathering.

The knowledge organization work we discuss in this paper represents the cumulative efforts of multiple cohorts of student library workers working within a specialized cataloguing infrastructure built through the MRAi as multi-institution, community-engaged project. This one aspect of the iterative design work—the building of a local thesaurus—was made possible by the previous and ongoing work of library professionals at UBC Library. We see this

terminology work as carrying out in a microcosm the larger project's values and priorities of: reciprocity in research-community relationships, repairing exploitative dynamics between universities and vulnerable communities, and broadening access to research-based and community-generated materials.

### **Background: A Community-Centered Digital Collection**

The MRAi Steering Committee includes representatives of UBC Library, UBC Learning Exchange, UBC School of Information, Simon Fraser University Library, and Vancouver Public Library. The MRAi was formed to respond to community concerns of extractive research practices in which residents of the DTES repeatedly participated in research studies yet did not have access to the publications resulting from these studies. Furthermore, the DTES community attracts a high volume of studies and a lack of coordination and visibility may have led to redundancy in research projects, with disparate research teams repeatedly drawing from the same pool of participants without understanding the cumulative burden of this practice (UBC Learning Exchange, 2014; Ubels et al., 2020). Beyond this exploitative history specific to university-community relations, the DTES is also the target of criminalization and stigmatization from the broader society and governing jurisdictions, especially about sex work, drug use, housing precarity, and mental health (DTES Literacy Roundtable, 2018). To mitigate these issues, the MRAi developed the Downtown Eastside Research Access Portal (DTES RAP) to aggregate research findings and other published information featuring or related to the DTES, and has endeavoured to center local stakeholders' needs in its design.

Prior to the official launch of the DTES RAP in March 2020, the MRAi worked in collaboration with the Supporting Transparent and Open Research Engagement and Exchange (STOREE) research project team on two community engagement events. The first was an evaluation session whereby we invited approximately a dozen community leaders and end users to a preview the DTES RAP prototype at UBC's Learning Exchange in Fall 2019. This session was comprised of several parts: an overview and demonstrations of the portal; activities concerning specific aspects and uses of the portal; and a focus-group-style discussion. Moderators took notes on suggestions, issues, and comments raised by participants throughout the session, and some participants gave written feedback during the focus groups. We received important input on topics and metadata we could consider as we completed the design of the portal.

The second event was a Metadata-a-thon in summer 2019 that brought together over 20 information professionals, community members, and university faculty and graduate students to locate materials for the digital portal and provoke discussion about description. We explored legacy library metadata, such as Library of Congress Subject Headings (LCSH), in terms of the contents in this collection and noted its limitations in the use of stigmatizing, harmful language. We also tested our initial list of DTES RAP "Topics," broad areas of classification that used community-preferred terms to facilitate browsing the collection. Additional metadata relevant to community needs included the access rights (such as open access) of a given resource, and related-work connections, allowing easier navigation between scholarly publications and their adaptations into clear language, infographic, and video formats. The collection has since grown to collect community materials, complementing the scholarly items with research, reports, and creative works produced in the neighbourhood. As of early 2023, the collection includes nearly 1800 items.

Given the volume of scholarly publications about the DTES, there is already extensive metadata in the RAP drawn from institutional repositories and publisher databases. Several metadata fields—here we focus on subject terms—are aggregated from multiple sources. In

assessing the appropriateness of the aggregate metadata to local needs, we found issues noted by the “Reimagine Descriptive Workflows” report common to aggregated digital collections: the imported terminology lacks the context of its original collection and is misaligned to local terminology (Frick & Proffitt, 2022). Furthermore, any intervention into this situation must account for layers of infrastructure: the digital collection is built on a university’s institutional repository (cIRcle), creating particular constraints and interdependencies affecting how metadata can be represented and altered for the collection’s needs. Our initial exploration of subject terms in the aggregate metadata revealed issues so serious and pervasive that we temporarily hid “subject” fields from item view, though they are still connected by indexing to keyword discovery.

### **Related Literature**

We situate our terminology work among numerous recent reparative metadata projects (Fox et al., 2020; Frick & Proffitt, 2022; Vaughan, 2018; Watson, 2021), including parallel thesaurus development projects aligned to historically marginalized communities (Berg et al., 2022; Cifor & Rawson, 2022). Reparative metadata work is a kind of “critical cataloguing,” practitioner- and community-driven movement away from dominant, legacy, and centralized cataloguing practices and toward local practices responsive to the needs of marginalized communities (Watson, 2021). New practices have often begun with critiques and assessments of existing systems, finding their limitations in describing communities made marginal including Indigenous Peoples (Bone & Loughheed, 2018) and LGBT2QIA+ communities (Adler, 2017).

In intervening in inaccurate and stigmatizing language relevant to this digital collection and community, we especially found alignment and inspiration from work assessing terminology for substance use (Broyles et al., 2014; Kelly et al., 2015; Kelly, 2004; Spiels and Conner, 2018; Wakeman, 2013; White et al., 2011; Xie et al., 2017), sex work (Benoit et al., 2018; Davies, 2015; Heverin & Zach, 2012; McMillan et al., 2018), HIV and AIDS (Dancy-Scott et al., 2018; Huber and Gillasp, 1998; Wood et al., 1996), and housing and homelessness (Berman, 2005; Hulchanski et al., 2009; Social Responsibilities Round Table of the American Library Association, 1996). Many of these works are grounded in and target the language used by specific communities of practice, such as medical researchers, and make specific recommendations as to more accurate and respectful terminology choices. These works are not limited to intervening through controlled vocabulary systems but make wider recommendations relevant to scientific and scholarly writing.

We find three of the five recurring recommendations Watson (2023) identifies across critical cataloguing scholarship characteristic of our approach here: “the use of “ethical outreach” when dealing with still-living identities, items, or groups,” “the replacement of dominant classification or cataloging on a local level”, and “consultation with described subjects.” While allowing for the tendency for reparative metadata practices to break rules and privilege points of view other than those in established systems (e.g., literary warrant), thesaurus features of the NISO standard (ANSI/NISO Z39.19-2005) and the core principles of domain analysis (Hjørland, 2002) are consistent with the aims of this project. Equivalence term relationships for addressing synonymity (USE/UF relationships), hierarchies and associations to assist in navigation among preferred terms (BT/NT and R/RT) and scope notes to clarify the coverage of terms (SN) are all relevant to resolving the terminology issues characteristic to aggregated metadata. The multiple stakeholders in the DTES RAP, including researchers, residents, and students, messily map onto Hjørland’s concept of a “domain,” though here we specifically prioritize the terminology of those with lived experience and who are themselves the referents for terms when seeking authoritative sources and resolving conflicts in language use. There are

few works detailing community-engaged thesaurus construction. However, the development of the Injured Workers Consultants Community Legal Clinic thesaurus was an inspiration to this project because of its approach to selecting term based on local usage and socially responsive thesaurus design (Soglasnova & Hanson, 2015).

During our research into work that intervenes in terminologies for subjects such as substance use, sex work, and HIV and AIDS, we also researched local and community-specific positions on these terminologies where possible as they directly affect the communities discussed in the work hosted through the RAP. We discovered that the online Addiction resource (Recovery Research Institute, 2017), the Vancouver Area Network of Drug Users (VANDU), the First Nations Health Authority (FNHA), the Western Aboriginal Harm Reduction Society (WAHRS), as well as the British Columbia Centre for Disease Control and Canadian Public Health Association all aligned with the three best practices that we drew from our literature review. They include 1) person-first language, 2) avoiding stigma, and 3) using consistent and accurate bio-medical language. VANDU, for example, is a DTES/Vancouver-based community organization run by and representing the interests of people who use drugs. They use non-stigmatizing, person-first language to refer to people who use drugs and argue for harm reduction and community-based intervention rather than prohibition and criminalization for illicit substance use. The FNHA, like VANDU, recommends using person-first language and avoiding words like “addict” or “user.” These terms may evoke implicit punitive biases compromising the quality of medical care and also may create unintended barriers to honest self-disclosure and treatment engagement for those suffering from alcohol or drug use conditions, and such language is inconsistent with other medical language and standards (Kelly et al., 2015). Dancy-Scott et al. (2018) likewise suggest that consistent, accurate language be used to describe HIV and AIDS: “HIV/AIDS”—a former subject heading that has since been changed—inaccurately equates the disease with its terminal stage and can carry stigma of people who are infectious or “dirty.” For these reasons, we found that community organizations that have adopted identified best practices for these vocabularies are appropriate resources that can inform the RAP’s access points, including subject and topic headings.

### **Thesaurus Construction**

In the following section, we detail the different modes of knowledge organization work that went into the current version of the thesaurus. Each type of work, and the subsections in which we describe them, are iterative and overlap in the chronology of the project, so that, for example, metadata extraction and processing has occurred multiple times throughout the project, and evaluation activities have taken place between and during the construction of different topic areas.

### **Exploration of Existing Vocabularies**

In focusing on subject terminology, we first reviewed existing, centrally-managed controlled vocabularies as possible alternatives to the imported metadata, including Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH), and other generalist vocabularies. After review, we concluded all such vocabularies contained problematic terminology for relevant topic areas, incompatible with the priorities of the project. Other vocabularies (cited in the related work section) specific to topics in the collection were inadequate to represent the collection, which is more diverse than the coverage of any one of these vocabularies. Rather than rely on the unedited, imported subject terminology or remove “subject” fields permanently, we chose to create a local thesaurus, starting a multi-year project to review imported subject terminology, select or create preferred terms among synonyms, and

map equivalent, hierarchical, and associative relationships among preferred terms (Aitchison et al., 2000; International Organization for Standardization, 2011).

### Metadata Extraction and Processing

We began with multiple passes of extracting the collection’s metadata and analyzing it for issues of consistency, completeness, and term use. We parsed and explored an up-to-date export of the collection’s metadata using OpenRefine, open access software built to navigate and batch edit messy data (<https://openrefine.org/>). This tool was useful in not only scoping aspects of the aggregate metadata for processing into thesaurus terms but also for generating overviews we could use to communicate terminology issues to non-specialists on the larger team.

Using OpenRefine, we split subject headings across multiple cells for ease of filtering by other attributes. Subject headings were then normalized to eliminate variants in title case and white space using the common transformation’s function. Next, the subject headings were extracted by topic tag and were split using the text facet function which produces a list of headings by frequency. This list of subject headings was copied into an Excel spreadsheet and the process was repeated for 18 topic tags. This working document was used to analyze common usage and to identify redundancies. It was also used to produce summary statistics such as the number of total subject headings in a topic, number of items without subject headings, and percentage of items in a topic without subject headings.

### Topic Areas

Focusing on the “Substance Use” and “Health and Wellbeing” topic tags, student workers reviewed the subject list manually to identify the relevancy of the subject headings for a given topic area. Early attempts at drafting a thesaurus were rooted in our literature review findings, identified best practices, and existing resources with suggestions for alternative terminologies such as *Addictionary*. We first mapped of existing terms, identified by their source vocabulary, to community-based vocabularies and resources. Figure 1 shows an early example of this mapping.

**Figure 1.** Excerpt of term mapping sheet between DTES RAP and controlled vocabularies

Subject Search Term	# of RAP Hits	Controlled Vocabulary/ Classification Scheme Source(s)	Suggested Alternative Term(s)	Suggested Alternative Term Source(s)
Abusers, Drug	3	LCSH; MESH	People who use drugs; People with lived / living experience	<i>Addictionary</i> ; Canadian Public Health Association; Government of Canada; Partnership to End Addiction
Addiction to alcohol	3	LCSH	People who use alcohol; People with alcohol substance use disorder	<i>Addictionary</i> ; Canadian Public Health Association

For a community-centered thesaurus, we weighed user and organizational warrant more heavily than literary and historical warrant (Bullard, 2017). We looked to community organization language guides such as the BC Centre for Disease Control, VANDU, *Addictionary*, Canadian Public Health Association, among others to establish preferred terms. Once preferred terms for concepts were established, the conceptual work of creating equivalency, hierarchical, and associative relationships could begin. Following NISO (2005)

standards, the thesaurus is currently displayed in alphabetical order with capitalized notation (BT, NT, RT, UF, USE) and indentation as visual cues to represent the term relationships.

Among the subject term clusters, the cluster within the “Substance Use” topic was a high priority given the terminology problems identified within the topic, as well as its high proportion of items—in 2021, there were 468 items within the topic (out of 964 total items), 829 unique subject headings distributed between them, and nearly 70% of these unique headings appearing on a single item. “People who use drugs,” for example, returned a different set of results than “People who use drugs (PWUD).” Such variation made the RAP more difficult to browse and language such as “Drug addicts” made discovery and access unnecessarily stigmatizing and painful. The high frequency with which these subject terms also appeared in items under other topic areas made this a clear starting point.

Following the rationale and groundwork laid out by work done in 2021, the development of the thesaurus continued on the “Substance Use” topic in 2022-2023. Due to the significant overlap among items in the “Substance Use” and “Health and Wellbeing” topic tags, thesaurus development was pursued simultaneously for these two topics. Together, these two topics comprised the second-highest proportion of undescribed items, having 317 items without subject headings.

A draft of the thesaurus for the “Arts, Culture, History and Heritage” (ACHH) topic was completed in the summer of 2022. This topic tag represented an area of consistent growth for the collection and contained the highest proportion of community materials which contain alternative (non-scholarly) genres. At the time of development, 59% (138 out of 234 items) of these items lacked subject headings, the highest proportion of undescribed items in the collection, making them challenging for users of the RAP to access.

As there are a wide range of subject headings in the RAP including LCSH and MESH as well as author supplied keywords, we included extensive lead-in term entries to expediate future indexing through find-and-replace functions. Scope notes were used extensively throughout the thesaurus to designate rationale and provide additional context for future indexers who may not have subject expertise in topic areas. The use of scope notes also supports future edits by providing an explanation for term usage, as shown in Figure 2.

**Figure 2.** Excerpt of thesaurus showing use of scope notes and lead-in terms

### **Needle Exchange Programs**

**SN Use when referring to services where people can dispose of used needles and be provided with clean needles, sterile water, alcohol wipes and health information (City of Vancouver).**

**UF Needle Exchange**

**UF Syringe Exchange programs**

**RT Needle Sharing**

**RT Supervised Injection Sites**

### **Testing and Evaluation**

Student workers tested their draft thesaurus sections by indexing a sample of 20-30 items in the RAP. This included a mixture of scholarly and community resources of a variety of genres including academic articles, community reports, podcasts, videos, and infographics. Student workers noted any problems that arose and used this experience to inform further editing of the thesaurus. Some of the issues encountered and addressed at the testing stage included eliminating redundancies, clarifying scope notes and term relationships, and redressing gaps in subject coverage.

During the development of the Arts, Culture, History, and Heritage topic portion of the thesaurus, we introduced further evaluation of thesaurus sections through a sample cataloguing activity (Figure 3) which was shared internally with student and other library workers. This activity asked participants to index five specific items in the collection under the ACHH topic and to provide notes on their experiences doing so. Reflections and learnings from this activity were used to inform further edits on the thesaurus draft with a particular focus on ensuring adequate coverage of collection resources, appropriate disambiguation, and confirming that preferred terms reflected community usage.

**Figure 3.** Instructions from a cataloguing activity for evaluating the Arts, Culture, History, and

This activity is comprised of a **cataloguing exercise** using the current draft thesaurus of Arts, Culture, History and Heritage terms and a short set of **reflection questions**. To complete the exercise, you will apply relevant terms from the thesaurus to a selected sample of 5 RAP items currently tagged in the "Arts, Culture, History and Heritage" topic. Following the cataloguing portion, there will be a short list of reflection questions that ask questions regarding language, usability, and thoroughness.

*\* If possible, these reflection questions will be used to engage in a short conversation via Zoom regarding the thesaurus. If we are unable to schedule a synchronous meeting to debrief your responses, your written responses will just be used!*

### **Heritage section of the thesaurus**

Substance use and health are complex topics with highly specific terminology. Further evaluation of the thesaurus sections arising from these topics was conducted in collaboration with domain experts. Researchers with domain expertise were asked to participate in a word association activity to assist in determining preferred terminology (Figures 4,5). The goal of this activity was to help determine user derived terms, user associations with terms, relevant synonyms and lead in terms, appropriate relationships between terms, and to identify terms that should be distinguished from each other (Spiteri, 2005). The design of the activity was derived by work from by (Spiteri, 2005) and (Soglasnova & Hanson, 2015).

**Figure 4.** Instructions for word association activity given to domain experts

**Word Association Activity**

For this activity we ask that you list words you associate with the supplied term. These words could be:

- Words that are commonly used as synonyms
- Similar terms or concepts that are important to distinguish from the given term
- Terms you would use to conduct a literature on the given term or concept
- Other associated terms you use or encounter in your work or research context

There are no wrong or right answers, and no minimum or maximum number of words that you must think of. For each of your chosen words, please explain briefly how you think the word is related to the original term. Feel free to make a note of terms you feel confident matches your expertise and on others that you may be less certain about. Your input is valuable and will contribute to ensuring the DTES RAP is accessible to a variety of users.

There is an example listed below for your reference, followed by a list of terms we would like your input on. Please fill in the provided text box with your answers and **bold** the confidence scale term you feel best reflects your confidence level about your answer.

**Figure 5.** Format example from word association exercise

**Example:**

<p>Original term:</p> <ul style="list-style-type: none"><li>• HIV [disease]</li></ul>	<p>My associated terms:</p> <ul style="list-style-type: none"><li>• Virus (HIV is a type of virus)</li><li>• AIDS (the terminal stage of the disease -- it is often confused with HIV but should be distinguished)</li><li>• Human immunodeficiency virus (synonym)</li><li>• Antiretroviral therapy [ART] (treatment for HIV)</li><li>• HIV Infected (a stigmatizing term that should be avoided)</li><li>• HIV response (umbrella term for strategies that seek to prevent the transmission of HIV)</li></ul>
<p>Confidence scale:</p> <p style="text-align: center;">                                            </p> <p>Not Confident at all   Slightly Confident   Somewhat Confident   <b>Fairly Confident</b>   Completely Confident</p>	

The activity was successful in generated user derived terms and in assisting to establish synonym control, hierarchical, and associative relationships. Additionally, researcher

commentary about terms that are important to distinguish from each other proved valuable for several terms. For example, researchers suggested that distinguishing peer run overdose prevention sites from state sanctioned supervised consumption sites was important and this resulted in two distinct, related terms with scope notes to clarify usage, as in Figures 6 and 7, below.

**Figure 6.** Thesaurus entry for Overdose Prevention Sites, as distinguished from Supervised Consumption Sites

Overdose Prevention Sites

SN Use when referring to safe spaces, typically run by peer or health workers who monitor people who use drugs while they use to help prevent and recover from overdoses/drug poisoning.

RT Overdose prevention

RT Supervised Consumption Sites

**Figure 7.** Thesaurus entry for Supervised Consumption Sites, as distinguished from Overdose Prevention Sites

Supervised Consumption Sites

SN Use when referring to formal social service sites that provide supports such as drug checking, clean drug equipment, and health information.

UF Drug Consumption Rooms

UF Observed Consumption Services

UF Safe Consumption Sites

UF Safe(r) Use Sites

UF Supervised Consumption Services

UF Supervised Inhalation Services

UF Supervised Smoking Facility

NT Supervised Injection Sites

RT Overdose Prevention Sites

### Findings

In our exploration of the imported subject terms and our development of a local thesaurus, we encountered some challenges familiar from other projects among the related literature. Early steps of working with the subject terminology in OpenRefine emphasized the difficulty with working with subdivision strings typical of LCSH, with combined headings such as “Women--Health and Hygiene--Research” needing to be broken down into their constituent parts for

analysis. In contrast, items catalogued with FAST (Faceted subject headings) or other post-coordinate vocabularies still generated issues of term choice but were more accessible for navigating and summarizing the aggregate subject terminology.

Since scholarly works go through subject indexing, the variation in the controlled vocabularies and perspectives of their respective source databases was reflected in the RAP's metadata, with a large extent of synonymy and variations in specificity (e.g., "Narcotics" or "Opioids") found throughout the aggregate subject terms. For example, the phrase "People who use drugs" or "PWUD," commonly used in community-centered writing, would produce different search results than the pejorative "Addicts" or even the MeSH term "Drug users." Given the multiple controlled vocabularies at play, navigation through specific terms was likely to fracture search results along disciplinary boundaries, so that the MeSH term would return only work sourced from medical journals, while the LCSH would return work sourced from general or social science collections, meaning that a user beginning with a term provided by a given item would unknowingly be limited to a subset of relevant materials biased toward a particular disciplinary view—a possible echo chamber. Intervention in the subject terminology was necessary to deliver on one promise of the RAP, to connect related works otherwise unknown to the disciplinary research groups working on similar topics.

Besides issues of synonymy and disciplinary boundaries, there were consistent issues of stigmatizing bias in the imported subject terminology. To some extent MeSH, and to a greater extent LCSH, use terminology that poorly fits this collection by medicalizing human behaviour (MeSH) and stigmatizing community members with outdated and biased language (LCSH). In addition to issues of respect for persons, we found instances where researchers' perspectives (typically those most privileged by literary warrant systems meant to reflect their publications) were poorly represented by these centrally managed thesauri. In considering the feedback received from domain experts through our word association activity, we generated the following recommendation for Decriminalization (Figure 8), with term relationships and scope notes directly contradicting existing LCSH relationships (Figure 9).

**Figure 8.** Thesaurus entry for Decriminalization, distinguish the term as an RT to Legalization

#### Decriminalization

SN Use to refer to a policy strategy in which non-criminal responses, such as fines and warnings, are available for designated activities, such as possession of small quantities of a controlled substance. It should be distinguished from *legalization*.

UF Prohibition

BT Drug Policy

RT Legalization

RT Harm Reduction

**Figure 9.** LCSH entry from id.loc.gov for Decriminalization, with Drug legalization as a Broader Term

## Decriminalization

Here are entered works on the process of making a formerly criminal act or omission noncriminal and without punitive sanctions.

### URI(s)

- <http://id.loc.gov/authorities/subjects/sh85036356>
- <http://id.loc.gov/authorities/sh85036356#concept>

### Broader Terms

- Criminal law
- Punishment

### Narrower Terms

- Drug legalization

More specific to this collection is a lack of subject access to non-scholarly works. In contrast to the extensive metadata imported with published scholarly works, non-traditionally published items such as podcasts, community-authored reports, and creative works such as poetry tended to arrive to the collection without subject indexing. This disproportionality in subject indexing created two challenges for the collection: 1) a lack of discoverability for non-scholarly materials through subject terms and 2) inherited biases from existed controlled vocabularies in our aggregated subject terms without a substantial counter from community-sourced items. The thesaurus therefore could not be constructed solely by choosing preferred terms from among the existing set but would also have to introduce new preferred terms sourced through community-centered materials and through consultation.

With respect to engagement and input from stakeholders, we found success in gathering perspectives through minimally burdensome requests, while more substantial consultation activities are still in development. The three minimally burdensome and stakeholder-specific requests we used were 1) the cataloguing exercise for library workers (Figure 3) 2) the word association exercise for domain experts (Figures 4,5), and 3) input through the community materials metadata form (Figure 10). Each of these engaged a different stakeholder group in ways that recognized their specific expertise and was complementary to their existing engagement with the DTES RAP.

*The cataloguing exercise* reached professionals whose work touched on the catalogue record but who were not typically engaged in cataloguing work nor the development of the thesaurus. These participants could contextualize the outcomes of subject indexing and, with a concise refresher on the syntactical structures of thesauri, navigate the draft thesaurus and assess whether relevant terms were available, intuitive, and findable. As professionals familiar with

the work aggregated in the RAP and who assisted in reference work, they had a good sense of the breadth of the collection as well as the terminology users expect to find in item descriptions.

*The word association exercise* introduced domain experts, such as health and policy researchers, to the function of the thesaurus but focused on their understanding of key terms, the use of those terms in their scholarly domains, and the subtle and sometimes critical distinctions among terms used in their own writing. We received feedback from 4/5 of the researchers we contacted in the first round of this exercise, with exhaustive and well-annotated notes on the history of terms in scholarly and public discourse, enabling us to create more lead-in terms as well as articulate term boundaries in scope notes. In addition to the direct input on term relationships and definitions, responses to this exercise also indicated to us the extent to which researchers are willing to engage in messy terminological work with knowledge organization projects and, once they are introduced to a given initiative, supportive of local knowledge organization labour. This finding is consistent with recent work on engagement with authors on subject indexing for their own work (Bullard et al., 2022; Koford, 2017) and points to a largely untapped resource of allies for local cataloguing work from researchers and scholars outside information science.

*The community materials metadata form* was created separately from this thesaurus project and facilitates library workers' processes of adding materials to the institutional repository outside typical faculty / staff / student internal processes. We found that engagement through this metadata form produced an opportunity to get informed input from creators on choices of subject terminology. Given that community materials are often produced by or with residents of the DTES, negotiating possible and appropriate subject terminology through the collaborative completion of the form notifies library workers to gaps in the thesauri and surfaces issues of fit or intuitiveness of terminology from a community perspective. Given the MRAi's goal to reduce burdensome extraction of community labour on academic projects, we find that this method, where community input is closely linked with existing collaborations and part of existing processes, strikes an important balance. This point of information sharing and exchange complements drawing from existing community resources that specify local terminology and word preference.

**Figure 10.** Community materials metadata form, including a space to negotiate subject terms

### Community Materials Metadata Form

Thank you for working with us to archive your materials in UBC's digital open access repository cIRcle!

To help people discover your materials in both cIRcle and our Downtown Eastside Research Access Portal (RAP), we include descriptions and metadata - data about data - to help people find materials on a topics or in a particular formats. To that end, we wanted to share what metadata we would like to use for your item(s) in case you have any input. If you would like to change any of the terms please feel free to make suggestions or recommendations. Please note we might change these tags over time in an effort to make your items more accessible and findable in these information systems but this is what we would like to start with.

**What we collect:**

**Citation:** How people can cite or reference your work **[FOR RAP]**

**Title:** **[Same on cIRcle and RAP]**

**Serial(series) Name:**

**Date:**

**Description:** Either the abstract of a paper or the description posted with the item on your website.

**Creators:** Anyone named as an author, interviewer, or otherwise involved in making your material

**Contributor:** anyone interviewed or consulted in your material

**Subject(s):** Any listed topics or themes about or listed on the material

**Topic(s):** This is what topic(s) your item will be listed under on the RAP.

**Affiliation (organization):** The name of the organization you work for, if there is one

**Category:** This tells people what type of material (community materials, creative work, tools and guides, etc.) your item(s) is (are)

**Genre:** This categorized your material with others made in the same style (thesis paper, documentary, poem etc).

**Type:** This tells people what format your material is in (video, text, podcast, etc)

### Future Work

Our work to assess sections of the draft thesaurus with various stakeholders has been effective for garnering feedback from both technical (e.g., librarians) and domain (scholars, clinicians) experts. We will continue to seek iterative input on the thesaurus with these groups, with the additional goal of involving people with lived experience. Our approach to consultation and feedback reflects the core project value that residents should not be overly burdened with requests for assistance but that their needs and perspectives must be centered in final decisions about resource description. Future work involving community input should also explore the effectiveness of the subject terms to enhance information retrieval and information discovery.

With three topic areas complete, covering the majority of subject terms and items in the collection, iterative work will complete the remaining topic areas in the thesaurus. Once the thesaurus is complete, implementation will require substantial back-cataloguing of the repository's collection and revising the metadata workflow for new items. These tasks are wide in scope and will require long-term planning and coordination with UBC Libraries to address

the interdependencies between the digital collection and cIRcle, the institutional repository that serves as its base architecture. However, we can employ creative technical solutions that allow us to take small steps toward implementation. For example, manual cataloguing of community materials and alternative genres, in collaboration with their creators, can make immediate use of the thesaurus for subject indexing and inform its ongoing maintenance.

Pragmatically, the DTES RAP serves the needs of a specific community with a digital portal that seeks to promote research access and community-preferred language. However, this project can have upstream effects. It serves a conceptual and technical space for experimenting with terminology and, as such, opens up conversations about democratizing research. Further avenues include supporting scholars to choose more thoughtful terminology for author-supplied keyword fields and to advocate for revisions to controlled vocabularies in their fields, and modelling effective collaboration with non-library experts around resource description. We also intend to publish the live thesaurus and documentation so that other projects may build on and learn from this work and apply it in their own unique contexts.

### Acknowledgements

We are grateful for the leadership from the Making Research Accessible Initiative Steering Committee, the support and involvement of the UBC cIRcle, and the generous input from each of the Downtown Eastside organizations and other stakeholders who have helped us advance our terminology work. Special thanks to Shannon Murray and Emily Conacher for their contributions to the thesaurus development over the years. This project was supported by the Social Science and Humanities Research Council of Canada Partnership Development Grant FAS F17-05624.

### References

- Adler, M. (2017). *Cruising the library: Perversities in the organization of knowledge*. Oxford University Press.
- Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: A practical manual*. Fitzroy Dearborn Publishers.
- Benoit et al. (2018). Prostitution Stigma and Its Effect on the Working Conditions, Personal Lives, and Health of Sex Workers. *The Journal of Sex Research*, 55(4–5), 457–471. <https://doi.org/10.1080/00224499.2017.1393652>
- Berg, M., Bains, S. K., & Suri, S. (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), 1–7. <https://doi.org/10.18357/kula.223>
- Berman, S. (2005). Classism in the stacks: Libraries and poor people. *Counterpoise*, 9(3), Article 3.
- Bone, C., & Lougheed, B. (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>
- Broyles et al. (2014). Confronting inadvertent stigma and pejorative language in addiction scholarship: A recognition and response. *Substance Abuse*, 35(3), 217–221. <https://doi.org/10.1080/08897077.2014.930372>
- Bullard, J. (2017). Warrant as a means to study classification system design. *Journal of Documentation*, 73(1), 75–90. <https://doi.org/10.1108/JD-06-2016-0074>
- Bullard, J., Watson, B., & Purdome, C. (2022). Misrepresentation in the Surrogate: Author Critiques of “Indians of North America” Subject Headings. *Cataloging & Classification Quarterly*, 60(6–7), 599–619. <https://doi.org/10.1080/01639374.2022.2090039>
- Cifor, M., & Rawson, K. J. (2022). Mediating Queer and Trans Pasts: The Homosaurus as Queer Information Activism. *Information, Communication & Society*, 0(0), 1–18. <https://doi.org/10.1080/1369118X.2022.2072753>

- Dancy-Scott et al. (2018). Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years. *JMIR Public Health Surveill*, 4(2), e50. <https://doi.org/10.2196/publichealth.8552>
- Davies. (2015). The criminalization of sexual commerce in Canada: Context and concepts for critical analysis. *The Canadian Journal of Human Sexuality*, 24(2), 78–91.
- Fox, V., Bennyhoff, N., George, K., Grant, E., Gross, T., Kellett, C., Kirkland, A., Jurgemeyer, K., Perera, T., & Pettitt, K. (2020). *Report of the SAC Working Group on Alternatives to LCSH "Illegal aliens"*. ALA/ALCTS/CaMMS Subject Analysis Committee.
- Frick, R. L., & Proffitt, M. (2022). *Reimagine Descriptive Workflows: A Community-informed Agenda for Reparative and Inclusive Descriptive Practice*. OCLC. <https://doi.org/10.25333/wd4b-bs51>
- Heverin, T., & Zach, L. (2012). Use of Microblogging for Collective Sense-Making During Violent Crises: A Study of Three Campus Shootings. *Journal of the American Society for Information Science*, 63(1), 34–47. <https://doi.org/10.1002/asi>
- Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4), 422–462. <https://doi.org/10.1108/00220410210431136>
- Huber and Gillaspay. (1998). Social constructs and disease: Implications for a controlled vocabulary for HIV/AIDS. *Library Trends*, 47(2), 190–208.
- Hulchanski et al. (2009). Homelessness: What's in a Word?" (Introduction). In *Finding Home: Policy Options for Addressing Homelessness in Canada*. University of Toronto. [https://www.homelesshub.ca/sites/default/files/attachments/FindingHome\\_Full.pdf](https://www.homelesshub.ca/sites/default/files/attachments/FindingHome_Full.pdf)
- International Organization for Standardization. (2011). *Information and documentation—Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval* (Standard ISO 25964-1:2011).
- Kelly et al. (2015). Stop talking 'dirty': Clinicians, language, and quality of care for the leading cause of preventable death in the United States. *American Journal of Medicine*, 128(1), 8–9. <https://doi.org/10.1016/j.amjmed.2014.07.043>
- Kelly, J. F. (2004). Towards an addictionary: A proposal for more precise terminology. *Alcoholism Treatment Quarterly*, 22(2), 79–87. [https://doi.org/10.1300/J020v22n02\\_07](https://doi.org/10.1300/J020v22n02_07)
- Koford, A. B. (2017). Engaging an Author in a Critical Reading of Subject Headings. *Journal of Critical Library and Information Studies*, 1(1), Article 1. <https://doi.org/10.24242/jclis.v1i1.20>
- McCauley, A., & Towle, A. (2022). The Making Research Accessible Initiative: A Case Study in Community Engagement and Collaboration. *Partnership: Canadian Journal of Library and Information Practice and Research / Partnership: Revue Canadienne de La Pratique et de La Recherche En Bibliothéconomie et Sciences de l'information*, 17(1), 1–31. <https://doi.org/10.21083/partnership.v17i1.6454>
- McMillan et al. (2018). Usage of the Terms Prostitution, Sex Work, Transactional Sex, and Survival Sex: Their Utility in HIV Prevention Research. *Archives of Sexual Behaviour*, 47, 1517–1527. <https://doi.org/10.1007/s10508-017-1140-0>
- National Information Standards Organization. (2005). *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the construction, format, and management of monolingual controlled vocabularies (Z39.19-2005; Revised 20)*. NISO Press.
- Recovery Research Institute. (2017, January 26). *Addictionary*®. Recovery Research Institute. <https://www.recoveryanswers.org/addiction-ary/>
- Social Responsibilities Round Table of the American Library Association. (1996). *Resolution on Subject Headings Relating to Class and Poverty* [Text]. Round Tables. <https://www.ala.org/rt/srrt-resolutions-1996-resolution-subject-headings-relating-class-and-poverty>
- Soglasnova, L., & Hanson, M. (2015). Socially Responsive Design and Evaluation of a Workers' Compensation Thesaurus for a Community Organization with Selective Application of Cognitive Work Analysis: A Case Study. *Cataloging & Classification Quarterly*, 53(8), 905–926. <https://doi.org/10.1080/01639374.2015.1044632>

Julia Bullard, Nigel Town, Sarah Nocente, Aleha McCauley & Heather O'Brien. 2023. Thesaurus construction for community-centered metadata. *NASKO*, Vol. 9. pp. 51-66.

- Spiehs and Conner. (2018). Considerations for substance-use disorder language: Cultivating a shift from 'addicts in recovery' to 'people who thrive.' *Journal of Public Health Policy*, 39, 372–378. <https://doi.org/10.1057/s41271-018-0127-y>
- Towle, A., & Leahy, K. (2016). The Learning Exchange: A Shared Space for the University of British Columbia and Vancouver's Downtown Eastside Communities. *Metropolitan Universities*, 27(3), 67–83.
- Tuck, E. (2009). Suspending Damage: A Letter to Communities. *Harvard Educational Review*, 79(3), 409–428. <https://doi.org/10.17763/haer.79.3.n0016675661t3n15>
- UBC Learning Exchange. (2014). *Community based research dialogue summary report*.
- Ubels, N., McCauley, A., Towle, A., De Forest, H., & Baron, D. (2020). Working together to mitigate information privilege with the Downtown Eastside Research Access Portal. *BCLA Perspectives*, 12(1).
- Vaughan, C. (2018). The Language of Cataloguing: Deconstructing and Decolonizing Systems of Organization in Libraries. *Dalhousie Journal of Interdisciplinary Management*, 14(0), Article 0. <https://ojs.library.dal.ca/djim/article/view/7853>
- Wakeman. (2013). Language and addiction: Choosing words wisely. *American Journal of Public Health*, 103(4), E1-2. <https://doi.org/10.2105/ajph.2012.301191>
- Watson, B. M. (2021). Advancing Equitable Cataloging. *NASKO*, 1–28. <https://doi.org/10.7152/nasko.v8i1.15887>
- White et al. (2011). Alcohol/drug/substance 'abuse': The history and (hopeful) demise of a pernicious label. *Alcoholism Treatment Quarterly*, 29, 317–321. <https://doi.org/10.1080/07347324.2011.587731>
- Wood et al. (1996). *HIV/AIDS and HIV/AIDS-Related Terminology: A Means of Organizing the Body of Knowledge*. Routledge.
- Xie et al. (2017). Updating our language around substance use disorders. *CMAJ*, 189, E1566. <https://doi.org/10.1503/cmaj.733490>