



Shifting Paradigms The Impact of Streaming on Diversity in Academic Library Film Collections

Trevor Deck , Amelia Clarkson  and James Mason 

Volume 10, 2024

URI: <https://id.erudit.org/iderudit/1115613ar>
DOI: <https://doi.org/10.33137/cjal-rcbu.v10.41603>

[See table of contents](#)

Publisher(s)

Canadian Association of Professional Academic Librarians / Association
Canadienne des Bibliothécaires en Enseignement Supérieur

ISSN

2369-937X (digital)

[Explore this journal](#)

Cite this article

Deck, T., Clarkson, A. & Mason, J. (2024). Shifting Paradigms: The Impact of Streaming on Diversity in Academic Library Film Collections. *Canadian Journal of Academic Librarianship / Revue canadienne de bibliothéconomie universitaire*, 10, 1–32. <https://doi.org/10.33137/cjal-rcbu.v10.41603>

Article abstract

Investigations into format shifts from physical to digital access in libraries often centre print materials. Similarly, recent calls to action for an increasing focus on acquisition of materials that support equity, diversity, and inclusion (EDI) efforts within postsecondary institutions often centre print resources. For academic libraries, media like film have unique access and acquisition models that do not correspond to print and pose unique challenges extending back to the Hollywood studios that create and distribute films. This paper explores the dual shifts in academic libraries toward collecting fewer physical films and collecting more content to support EDI mandates, and asks: first, whether the shift away from collecting physical media may also be a shift away from including diverse perspectives in film collections; and second, if we have the data to draw a measurable and demonstrable conclusion. A comprehensive literature review traces efforts to assess markers of diversity in large library collections and/or film collections over the past two decades and helps establish a methodology that combines analyzing data from the library catalogue and Wikidata. Findings revealed that the completeness and consistency of the data over time makes drawing strong conclusions difficult and demonstrated the challenges of this approach in addressing EDI analysis, even when augmenting catalogue metadata with Wikidata. Curation and choice are perhaps more important in building a diverse film collection than questions of format alone, despite the challenges in assessing and collecting film which is and has always been a format in rapid and continual flux.

© Trevor Deck, Amelia Clarkson and James Mason, 2024



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>



This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>



Shifting Paradigms: The Impact of Streaming on Diversity in Academic Library Film Collections

Trevor Deck

University of Toronto

Amelia Clarkson

University of Toronto Mississauga

James Mason

University of Toronto

ABSTRACT

Investigations into format shifts from physical to digital access in libraries often centre print materials. Similarly, recent calls to action for an increasing focus on acquisition of materials that support equity, diversity, and inclusion (EDI) efforts within postsecondary institutions often centre print resources. For academic libraries, media like film have unique access and acquisition models that do not correspond to print and pose unique challenges extending back to the Hollywood studios that create and distribute films. This paper explores the dual shifts in academic libraries toward collecting fewer physical films and collecting more content to support EDI mandates, and asks: first, whether the shift away from collecting physical media may also be a shift away from including diverse perspectives in film collections; and second, if we have the data to draw a measurable and demonstrable conclusion. A comprehensive literature review traces efforts to assess markers of diversity in large library collections and/or film collections over the past two decades and helps establish a methodology that combines analyzing data from the library catalogue and Wikidata. Findings revealed that the completeness and consistency of the data over time makes drawing strong conclusions difficult and demonstrated the challenges of this approach in addressing EDI analysis, even when augmenting catalogue metadata with Wikidata. Curation and choice are perhaps more important in building a diverse film collection than questions of format alone, despite the challenges in assessing and collecting film which is and has always been a format in rapid and continual flux.

Keywords: *academic libraries · equity, diversity, and inclusion · film · library collections · media*

RÉSUMÉ

Les enquêtes sur le changement de format de l'accès physique à l'accès numérique dans les bibliothèques se concentrent souvent sur les documents imprimés. De même, les récents appels à l'action en faveur d'une concentration accrue sur l'acquisition de matériel soutenant les efforts d'équité, de diversité et d'inclusion (EDI) au sein des établissements postsecondaires se concentrent souvent sur les ressources imprimées. Pour les bibliothèques universitaires, les médias comme le cinéma ont des modèles d'accès et d'acquisition uniques distincts de l'imprimé et posent des défis uniques qui remontent aux studios hollywoodiens qui créent et distribuent des films. Cet article explore la double évolution des bibliothèques universitaires vers la collecte de moins de films physiques et la collecte de plus de contenu pour soutenir les mandats EDI, et se demande : premièrement, si l'abandon de la collecte de supports physiques peut également être un abandon de l'inclusion de perspectives diverses dans les collections de films ; et deuxièmement, si nous disposons des données nécessaires pour tirer une conclusion mesurable et démontrable. Une revue complète de la littérature retrace les efforts visant à évaluer les marqueurs de diversité dans les grandes collections de bibliothèques et/ou de films au cours des deux dernières décennies et aide à établir une méthodologie combinant l'analyse des données de catalogues de bibliothèque et de Wikidata. Les résultats ont révélé que l'exhaustivité et la cohérence des données au fil du temps rendent difficile la tâche de tirer des conclusions solides et ont démontré les défis de cette approche pour aborder l'analyse EDI, même en agrégeant les métadonnées des catalogues avec Wikidata. La conservation et le choix sont peut-être plus importants dans la constitution d'une collection de films diversifiée que les seules questions de format, malgré les défis liés à l'évaluation et à la collecte de films qui sont et ont toujours été un format en évolution rapide et continue.

Mots-clés : *bibliothèques universitaires · collections de bibliothèques · équité, diversité et inclusion · film · médias*

WITHIN Media Librarian organizations and working groups, there is a common anecdotal assumption that a move away from collecting physical media will lead to a loss of diversity in the collection, in part due to the perceived lack of choice offered by large streaming platforms and packages compared to the title-by-title selection of physical formats. This leads to the central questions guiding this research paper: what are the potential implications for representation in media collections if libraries fully embrace the format shift toward streaming and abandon collecting physical materials? Does collecting physical materials ensure a more diverse media collection than streaming alone? Do we have adequate sources of collection data available to answer these questions?

Finding an answer to these questions requires an exploration of the history of film acquisitions models in libraries in relation to distribution models heavily influenced by and biased towards the Hollywood studio system. In comparison to

other library resources, film is a format always in flux and transition. In defining what EDI looks like in film, whether it is the creators, the content, or both, it becomes clear that just as Hollywood influences format and distribution models for film, it creates and perpetuates problems with a lack of diversity in the medium. This comes to light in analyses of commercial streaming platforms such as The Criterion Channel (Buchanan and Ugwu 2020) and reaches library platforms such as Kanopy by the summer of 2020 (Louie 2020), evaluating representation in the films available. While there is some literature investigating diversity in media collections in academic libraries, the question of format and access rarely enters the conversation. This literature provides useful models for defining aspects of EDI and identifying them within collections, however a new methodology is required to address format alongside diversity in a media collection.

Local Context

In the summer of 2020, the Black Lives Matter and Anti-Black Racism movement reached new levels of urgency following the public murder of George Floyd, and academic libraries across North America responded by pushing for change across all areas of service, including collection practices (American Library Association 2020). Like many, the University of Toronto Libraries (abbreviated to UTL) released an anti-racism statement and action items that included a Collections Diversity Plan (UTL Collections Committee 2022). This document uses the language of book and serials acquisitions, referring to a knowledge divide created by spending on “Big Deal contracts” and “large-scale approval-plans” limiting discretionary funds that could ultimately broaden and diversify content (UTL Collections Committee 2022). However, a similar model has arguably emerged in media acquisitions, particularly film, in which large streaming platforms with preselected content consume the majority of budgets and physical media acquisitions get deprioritized, while at the same time its vendors become increasingly scarce.

UTL’s collection practices surrounding *format* remain without guiding documentation, but anecdotal evidence and recent action provide the shape of a deprioritizing of physical media collections in favour of streaming access. The closing of UTL’s Media Commons service desk in Summer 2022 indicates a perceived lack of need for a dedicated physical film collection, and the requisite designated space and subject experts. The shuttering of physical media vendors in recent years, such as long-standing CD vendor Arkivmusic and the DVD vendor Action! Media Library, gives way to conversations about whether this is a new challenge or instead an opportunity to abandon collecting these formats. The cost of maintaining both physical and streaming collections becomes increasingly untenable. While 88% of libraries procure

streaming media content from streaming platforms (MacDougall and Ruediger, n.d.), their annual fees can cost tens of thousands of dollars (King 2014). The average spent on all streaming licensing for libraries totaled \$22,187 USD as of 2015 (Tanasse 2021), and as of 2020 the highest ranked “biggest challenge” for maintaining streaming media is cost, as identified by 89.80% of respondents in a 2020 IFLA report (Tanasse 2021). It is no wonder that as both demand for streaming access and its acquisition cost rises, the perceived value of investing in physical media alongside dwindles.

Taking UTL as an example, the opportunities presented in the Diverse Collections Plan to address Equity, Diversity, and Inclusion (abbreviated EDI) make explicit reference to books and serials but leave out media as a unique format with its own challenges. The ACRL Guidelines for Media Resources in Academic Libraries presents the assumption that while media requires the same principles of collection management and preservation that apply to print and other library collections, “media formats, delivery, and access mechanisms will continue to evolve and shift at a pace that exceeds print resources” (Guidelines for Media Resources for Academic Libraries in Higher Education Task Force of the Association of College and Research Libraries 2006). As such, the challenges outlined in UTL’s plan simultaneously provide answers to and raise questions about building a diverse media collection.

In describing the opportunity to redirect funds from “the purchasing and licensing of mainstream commercial academic content” to “endangered content and marginalized voices” (UTL Collections Committee 2022), the language “mainstream” and “commercial” is also an accurate description of the content available via many large subscription streaming platforms for university libraries. The opportunity sets mainstream content in opposition to, or perhaps simply incompatible with, marginalized voices, and as unlikely to become endangered. Building on this idea, another challenge addressed in UTL’s Diverse Collections plan is a reliance on vendor selection for acquiring materials potentially undermining the discoverability and value of small publisher collections, thus negatively affecting the diversity of collections. We wonder whether streaming platforms that offer a predetermined selection of films also obscure or omit small, independent films that are more likely to add diversity to the media collection.

Finally, UTL’s plan comes close to exploring format-based collection policies, in addressing how the removal of material from public spaces in tandem with an increased reliance on digital resources might erase the *visible* diversity of a collection. The opportunities presented to address this challenge do not question the reliance on digital resources and the move away from physical materials as erasing diversity itself in the collection; ultimately this is our guiding question for the following research. The impetus for UTL’s Collection Diversity Plan and the proliferation of

so many others like it springs from a call to action for anti-racism in the summer of 2020. Library staff were asked to think critically about institutional policies and practices that may have contributed to harm and develop actionable plans to reduce harm and improve representation, including in collections. However, language used in UTL's Plan document also highlights the ways in which practices in a digital age embrace passive rather than active collection strategies. Is the call to embrace digital formats and their attendant content delivery models, such as Big Deals or large streaming platforms, in opposition to the call to interrogate and diversify library collections?

Hollywood Context

It is impossible for a library evaluating film formats to limit the critical scope to the confines of academia; questions of availability, accessibility, diversity, and representation extend towards Hollywood itself far more than print materials within academia are governed by the press. As King notes, compared with books, movies “have existed for a century rather than a millennia, and every generation or so there are dramatic technological changes that lead to new formats and methods of delivery” (King 2014, 295). Indeed, Fossatti frames the many moments of technological change for film as “a catalyst for a process that has never found rest: the continuous transformation of audiovisual media or, maybe even more aptly...their inherently transitional nature” with “existing logics of production, distribution, and exhibition” challenged or championed within the industry but spreading to audiences, including those in academia (Fossatti 2021, 13-14).

In a 2012 article on the move in Hollywood from shooting with 35mm film to digital technology, Alimurung interviews directors such as Christopher Nolan, who point out that digital technology is cheaper and faster for both creators and distributors, allowing studios to forgo physically printing and shipping movies with digital releases (Fossatti 2021). The article further explores Hollywood's fraught history with format preservation, pointing out that:

Not too long ago, studios simply threw films away. Paramount planned to burn its old nitrate. MGM was set to dump its original negatives — including those for *Gone With the Wind* and *The Wizard of Oz* — into the ocean. What did they need those for, they figured? They'd made copies. Luckily for the studios, archivists at UCLA and Eastman House took the prints instead. Because, years later, MGM wanted to digitize its old movies and needed the originals back. The copies they'd made, on Kodak stock, had faded (Fossatti 2021).

Hollywood history such as this, in which archivists at an academic institution figure prominently in preserving content that the industry itself does not understand has future utility, reveals the locus of anxiety within academia over format shifts. As the

current director of the UCLA Film & Television Archive points out, the most recent shift to digital is also at risk of forgoing long-term storage due to rapid obsolescence, stating “in the last 10 years of digitality, we’ve gone through 20 formats!” (Fossati 2021). If studios embrace change but are not committed to preserving film history across formats, uncertainty cascades to the institutions that attempt to collect and preserve access to this history.

Large Hollywood studios at least have the resources to consider preservation and sustainability efforts, and the 2007 report *The Digital Dilemma*, coauthored by the Academy and the Library of Congress National Digital Information Infrastructure and Preservation Program (The Science & Technology Council, The Academy of Motion Picture Arts and Sciences 2007), highlighted the urgency of doing so for major studios. In a follow up report from 2012, the same organizations turn to how this issue affects Independent Filmmakers and Documentarians (The Science & Technology Council, The Academy of Motion Picture Arts and Sciences 2012). Paradoxically, digital technology lowers the barrier for these creators to make and distribute movies but increases the uncertainty of long-term preservation. This is also of particular interest to the question of diversity and representation in creator identities as one might assume that, given Hollywood’s known whiteness, independent filmmakers and their works would contribute to collection diversity. A 2022 article by Michael Tran, however, explores the racial barriers to independent filmmaking in the US (Tran 2022). Tran finds that despite discourse surrounding independent film as inclusive and non-commercial, many of the mechanisms of racial exclusion that underlie Hollywood’s lack of diversity exist in independent film: economic uncertainty of creative labour, racialized market logistics, economic stigmatization of artists of colour, and reliance on closed social networks for career advancement all keep creators from advancing in the independent film scene (Tran 2022). Furthermore, creators also face barriers to making art reflective of their experiences, meaning that independent film may not bring diversity in either creator identity or content as often assumed. Ultimately, in addition to facing challenges to preserving their works, independent filmmakers face “the combination of racial and market logics that excludes people of color from equal participation in the industry operat[ing] at full force in the more economically uncertain arena of independent film” (Tran 2022).

For libraries, the forces at work within Hollywood filmmaking are inescapably tied to the quality and content of the films they can collect. An industry with massive, deep-rooted barriers to participation for marginalized creators, and with a history of rapid format shifts and haphazard preservation practices, inherently passes these problems onto institutions that seek to study it and preserve access to its creations.

An analysis of diversity and representation across film formats in a library collection must engage with the power structures that first decide who gets to make works and how these works are distributed.

Literature Review

Studies related to assessing diversity and representation in film collections over the past two decades can be roughly categorized into two approaches. The first involves assessing diversity across many categories and formats that comprise a collection. The second involves assessing specific categories of diversity, such as gender, within specific formats. Formats are often print or DVD, and for media collections, no study has attempted to compare diversity and representation in a physical versus streaming film collection. This is particularly interesting given that most libraries today have a mix of both, but increasingly face decisions about resource allocation and user preference that tend towards streaming over physical formats. This is particularly timely as during the pandemic streaming became the ascendant format and libraries now face questions about what is gained and lost in this shift, without a method of assessment. The question of assessing equity, diversity, and inclusion as a concept across many categories in a media collection, while also comparing these markers across formats, requires a combined approach that has not been tested in the literature.

In 2000, Chu proposed developing criteria for evaluating whether “multimedia,” including film and video, is “multicultural,” using the working definition where “multiculturalism is broadly defined as the inclusion of all cultures, where culture includes racial, religious, or social groups” but pointing out that “libraries lack criteria for evaluating multicultural multimedia materials” (Chu 2000, 257). They further outline analysis on an item-by-item basis to understand context, bias, and treatment of subjects within the film, which while comprehensive would be challenging for a large collection. They also crucially note that the mere presence of subject headings, such as “Jews,” provides no insight into how a work treats the group in question, or whether other groups receive discriminatory treatment within the same work. Two years later, Gray and Abaid wrote about strategies for building a “multicultural” media collection via collection development and assessment policies, but similarly discuss evaluation strategies that involve assessing individual items for accurate portrayals of different cultures (Gray and Abaid 2002, 9). Both articles address the shortcomings and potential bias of one librarian doing the work, the need to potentially consult external perspectives, and ultimately conclude that available methods “are limited in their application to specific conditions and fail to provide one set of criteria to evaluate any multimedia material on any multicultural topic and language” (Chu 2000, 257).

Twenty years later such criteria still do not exist for media collections, though other authors have attempted to further develop applications for specific conditions. Ciszek and Young outline methods for diversity-related collections assessment for large academic libraries including the use of bibliographies for print collections, searching and subsequently collecting titles based on reading lists (Ciszek and Young 2010). Comparable filmographies have largely emerged since the summer of 2020, and as film scholar Raquel Gates points out, the instigating “Anti-Black Racism” lists of that time “reduce Black art to a hastily constructed manual to understanding oppression, always with white people as the implied audience” (Gates 2020). Given the pervasive whiteness of both librarianship and Film Studies in academia, the value of filmographies as a key to ensuring an inclusive collection may be reductive and perpetuate a monoculture.

Proctor turns to the library catalogue as a potential record of diverse representation, searching across the collection for Library of Congress subject headings related to LGBTQ issues and identities, but notes that inconsistent cataloguing or lack of appropriate subject headings in the classification system limit this method (Proctor 2020, 229). In addition, subject terms and collection organization in libraries are not designed to engage with multidisciplinary or intersectional ideas (Proctor 2020, 229), still failing to address Chu’s original notion of establishing criteria that can evaluate any multicultural topic. Using the catalogue record but focusing on media specifically, Tillay and Chapman assess the diversity of creator gender identities in the Howard-Tilton Memorial Library’s DVD collection, examining what percentage of films in the collection are directed by women (Tillay and Chapman 2019). They search for director names in the 245 field from a master list of female directors and note that since library metadata and catalogue records have only incorporated demographic data about creators since the early 2000s, an augmented methodology might pull gender information from Wikidata (Tillay and Chapman 2019, 62). Finally, Sims et al. provide a methodology for assessing diversity in the history of the *Journal of Cinema and Media Studies*, rather than a library collection (Sims et al. 2021) This approach is notable as it is the first to combine assessing two potential categories of diverse representation together: creator identity and contents of the item. Though the adequacy of catalogue metadata being augmented by other sources of data was addressed in the literature, no evidence of this approach was discussed.

Working from these studies, libraries face tension between the rigour of the analysis and the size of the collection. More detail and nuance in defining and exploring EDI becomes overwhelming for large collections, even reduced to a single format such as film. Studies that do limit their scope to media do not compare

formats and the diversity of their respective content. These studies help inform criterion to assess diversity across a collection using the library's metadata, but the creation and collection of such data raises ethical questions. Tillay and Chapman do note that "catalogers and metadata librarians are exploring how to describe gender in sensitive and accurate ways" (Tillay and Chapman 2019, 59), but in a talk at Harvard University Libraries in November of 2021, Billey pushes back against this notion and explores the potential harm of any attempt at recording gender identity in library metadata, noting that "the degree to which individuals are able to and choose to disclose it, is complex, contextual, personal, and subject to change over time and in different environments and jurisdictions"(Billey 2022). Interestingly, Billey extends this argument to other markers of diversity in library materials and notes that the catalogue does not and similarly should not record creator race or ethnicity in a consistent way. Like Tillay and Chapman, Billey then suggests Wikidata might provide a more precise, flexible, and accessible source of bibliographic data.

Clarke and Schoonmaker investigate existing metadata elements that can indicate diverse content in library reading materials, considering elements that can serve as entry points to understand the diversity status of resource creators, such as gender and geographic region, as well as topical or thematic content (Clarke and Schoonmaker 2019). Like Billey, the authors conclude that elements representing racial, ethnic, national, or cultural identity do not exist for creators, and "there is no current integrated library system or catalogue that surfaces diverse resources" (Clarke and Schoonmaker 2019, 177). They note that while searching metadata describing diversity is a first step towards finding and accessing diverse materials, ultimately the "lack of distinct elements contributes to the implicit erasure of marginalized identities" (Clarke and Schoonmaker 2019, 173), failing to represent intersectionality.

Finally, Jahnke, Tanaka, and Palazzolo take this idea further in 2022 (Jahnke, Tanaka, and Palazzolo 2022), exploring how metadata practices can in fact obscure materials from discovery as representative of collection diversity. Erratic classification and indexing mean that even within a single library collection it is difficult to determine whether search results truly represent all relevant materials. The authors take a step back in the assessment process and note that "a major hindrance in assessing the diversity of collections has been a lack of shared understanding of what it means for a collection to be diverse, and therefore how this dimension should be assessed" (Jahnke, Tanaka, and Palazzolo 2022, 168). This reflects the two decades of studies that came before, that either fall short of an accepted criteria to assess diversity broadly, or stick to a single dimension such as gender or sexuality. While Jahnke focuses explicitly on written works, the complications that attend media as a format are amplified in the analysis of the time and cost of developing diverse collections; the requirement of additional resources

beyond budget, such as labour for the increased time it takes to select and vet content, master novel acquisitions workflows, and develop expertise. Particularly when the authors address that compared to “mainstream materials”, diverse collections require “a more proactive style of collection development in which librarians continually seek out materials on and from voices that are excluded by automated processes” (Jahnke, Tanaka, and Palazzolo 2022, 180), they speak to the assumptions within media collection work that pre-curated streaming platforms will exclude the marginalized perspectives offered in the physical collection.

Taken together, studies that address the inadequacy and potential ethical quandary of relying on catalogue metadata to assess the diversity of collections advocate for a cross-referenced approach with platforms such as Wikidata. Notably however, these studies do not attempt or document said approach. Jahnke, Tenaka, and Palazzolo’s point that developing more diverse collections requires more time and labour extends to the process of analyzing them as well. As a way forward, however, it is worth revisiting Proctor’s work which concludes that “when engaging in assessment efforts aimed at examining multidisciplinary subjects, one has to work with the only, albeit limited, tools available which include existing classification schemes and other methods and tools that are inherently flawed” (Proctor 2020, 226). How then to develop a means of assessing an interdisciplinary media collection in a large academic library, with existing flawed methods and tools?

Methodology

For the purposes of this paper, inspired in part by the UTL University’s Collection Diversity Plan, we align our definition of Equity, Diversity, and Inclusion with that document. In particular, equity recognizes that some groups were and are disadvantaged and therefore underrepresented or marginalized in many institutions. Diversity is defined as the sum of ways people are both alike and different, and recognizes and embraces difference as uniqueness. Inclusion values each individual’s unique contribution (UTL Collections Committee 2022). In the context of Hollywood, undeniably the hub of development and production for the films our libraries collect, we see EDI concerns centred around creator identity, clearly demonstrated in recent efforts to encourage awareness and rejection of the persistent celebration of white, male creators in the Oscar’s category of Best Director (Smith, USC Annenberg Inclusion Initiative, and Adobe Foundation 2023). To approach assessing EDI in our film collection materials, we are looking for a diversity of perspective from creator identities and uniquely diverse experiences represented in subject matter.

In order to assess the potential effects of the trend away from collecting physical media in academic libraries, we devised a methodology to analyze select physical and

digital media collections within UTL Library for diverse perspectives. We focused on two specific areas: demographic information amongst film creators in each collection, and keywords from social justice themes represented in the subject matter. As indicated in the literature, previous work in this area has also focused on analyzing representation in both creator identity and narrative content, and while the catalogue can be a fallible and incomplete source of information, it is also the most complete tool that all institutions use for articulating and assessing collection items. Looking at various collections as a whole, not just individual titles, aims to offer a view of the content of various collections compared to each other, normalizing the flaws in the data. Furthermore, we can test the viability of the catalogue metadata for this purpose.

To analyze demographic information amongst film creators, we collected data from the 700 (creator) MARC field, as well as the 508 (creation/production notes) field. Together, these fields provide as close to a comprehensive list of individuals involved in a given film as the metadata allows. For the purposes of our inquiry, we isolated names of persons involved in filmmaking roles that wield the most power and authority; namely directors, screenwriters and producers. We opted to exclude actors from our analysis due to the fact that the records do not necessarily indicate the lead actors in a given film, or the role the actor played.

Datasets of directors, screenwriters, and producers were compiled for numerous distinct media collections within the UTL catalogue. For the physical media piece, we analyzed a collection of approximately 24,000 DVD, Blu-ray and VHS holdings within University of Toronto's Media Commons library. Digital collections included Criterion on Demand, a popular subscription-based streaming media platform that contains thousands of films from several major Hollywood studios; Docuseek, a streaming media platform representing the works of numerous noteworthy North American documentary distributors; Kanopy, a popular streaming service that contains films from a wide breadth of film suppliers, ranging from educational films to classic Hollywood cinema; and finally the 'MyMedia' collection, a collection of individually licensed films hosted on the UTL's in-house media streaming platform.

To obtain demographic information on individuals within the creator datasets, we devised a Python script that queries Wikidata for the names of each identified creator; demographic information (gender, birth date, country of birth) is then extracted from the entries that match each creator's name. This is a two-part check: if the name matches, the script goes to occupation; if the occupation also matches then the data is collected, otherwise it is not included. In this way we collected less data but were more certain that the data collected is correct. We chose Wikidata as it is a large and ostensibly more complete source of demographic data that is freely available, as

compared to other sources such as WorldCat Entities or VIAF. As noted throughout the literature, the catalogue does not capture complete demographic data (Jahnke, Tanaka, and Palazzolo 2022; Clarke and Schoonmaker 2019; Billey 2022) and best practices on recording elements such as gender for creators are relatively recent, and remain contested. While basic demographic information still falls short of being able to track more nuanced aspects of identity, such as gender or race, Wikidata allowed us to collect more extensive information on creators and the ways in which they may bring more representation, and diversity of perspective to the film collection. While country of birth does not necessarily speak to identity, it does speak to a diversity of perspective and allows us to draw some conclusions, even if the available data is insufficient to conclusively say this perspective contributes to EDI. In addition, the size of the data sample we were analyzing, 56,421 names from our physical film collection alone, precluded the ability to research individual names, and necessitated an automated mass-data approach, which Wikidata allows.

For subject matter analysis, another Python script extracted all subject headings from the catalogue for every title in the physical collection and four digital collections. The Subject field was chosen over the description field as it offered a more concise indication of the content or focus of a film and uses single words or short phrases, chosen from a controlled vocabulary (Library of Congress Subject Headings), rather than needing to parse full sentences. These subject headings were normalized, sorted, and counted. We then isolated subject headings and assigned a descriptive label that allowed us to identify related content across the data. We categorized the subject data into neutral terms and terms relating to social justice. By neutral terms we are referring to terms related to EDI that offer no value judgment. These term categories include:

- Historically marginalized peoples or customs
- Physical or mental disabilities
- Sexual / gender minorities
- Women

For terms relating to social justice we used “Social Justice Definitions” (National Association of School Psychologists 2021), a list of terms sorted into categories including:

- Gender and Sex
- History and Theoretical Concepts
- Power, Prejudice, and Oppression
- Race and Racism

- Allyship and Advocacy
- Culture and Religion
- (Dis)ability and Size Discrimination
- Socioeconomic Status

This allowed us to sort subject terms into categories that reflect the EDI content represented in the various collections, and to compare them. Creating the two categories was necessary because the intentionality of the application of subject terminology differed, in the sense that some allowed for a greater understanding of how the content was being discussed. For example, the subject term “Woman” as in “Woman detectives” was more neutral as it does not speak to the intentionality of the subject’s use but does show that women as a distinct identity were being discussed. Whereas the term “feminism” also deals with women but addresses intentionality in a more meaningful way.

This proved to be a rather challenging process since certain subject headings could be related tangentially to various categories. In addition to discrepancies around intentionality, subject headings alone do not account for how the subject is treated in the narrative. For example, a film that deals with the Holocaust may include the subject heading “Jewish people,” but the subject matter may be unrelated to EDI concepts around religious tolerance. Or a subject term from our Social Justice list, “Black Lives Matter,” though clearly an issue of social justice, doesn’t articulate the context of the term in the narrative of the film. Using data from the 520 “Summary” field can help get at the treatment of the subject, but this field was used far less regularly and consistently, making data analysis far more difficult and less reliable. As Iezzi et al. point out, one of the challenges of large data analysis for a collection of documents written in natural language is that “[t]exts encode information in a form difficult to analyze from a quantitative viewpoint because their content does not follow a given data model” (Misuraca and Spano 2020, 18). Another related factor complicating our analysis was the inherent subjectivity of interpretation, and as such the potential for bias, be it personal, cultural, institutional or religious, that impact the relation of subject headings to EDI. For instance, we have a tendency to assume there is a fairly universal understanding of what constitutes EDI principals in North American academia, however an EDI Commitment Statement from evangelical-based institutions like Trinity Western University (Trinity Western University, n.d.) contain notably significant omissions compared to University of Toronto’s Statement (University of Toronto Libraries, n.d.), due to Trinity Western’s religious intolerance around sexual orientation. Using the categories and terms (including synonyms) established in the “Social Justice Terms” (National Association of School Psychologists

2021) acted as an authority for us which we feel was effective and demonstrated a perspective that focused on equality and justice.

Findings and Analysis

The analysis of creator demographics provided many interesting takeaways and observations. Unsurprisingly, the overall breakdown by gender across all collections skewed heavily toward American and male representation. The one exception to this was Docuseek, where 42% of creators on the platform identify as women, compared to 15% in the Media Commons physical collection, 21% in the collection of licensed Kanopy titles, and 16% in the Criterion on Demand collection (see appendix). Some of this difference can be attributed Docuseek’s core mandate to focus on “independent, social-issue and environmental films” (Docuseek, n.d.), whereas much of the content represented in the Media Commons physical collection, Kanopy and Criterion on Demand, consists of Hollywood productions, which underrepresent female voices.

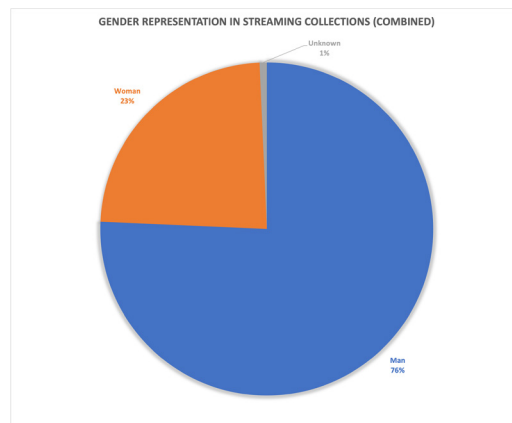


FIGURE 1 Percentage of creators by recorded gender across all streaming collections.

Canadian filmmakers are most well-represented in our MyMedia collection, with 20% of creators identified as Canadian, followed by the Media Commons physical collection, with 8.6% of individuals analyzed (notably absent from this analysis is the National Film Board’s ‘Campus Collection’, which would surely contain a high percentage of Canadian creators). This indicates that individual curation of titles (be it in the form of physical purchases or digital licenses) is immensely important in ensuring representation of Canadian creators, as all films added to MyMedia are by request rather than preselected subscription. The information we were able to extract from the metadata about creators allowed us to infer some demographic characteristics and draw some conclusions about various representation and diversity of voice, however the metadata lacked the degree of specificity we feel is important. For example, the overrepresentation of American male creators has no

indication of racial or cultural identity and how it may or may not influence their work.

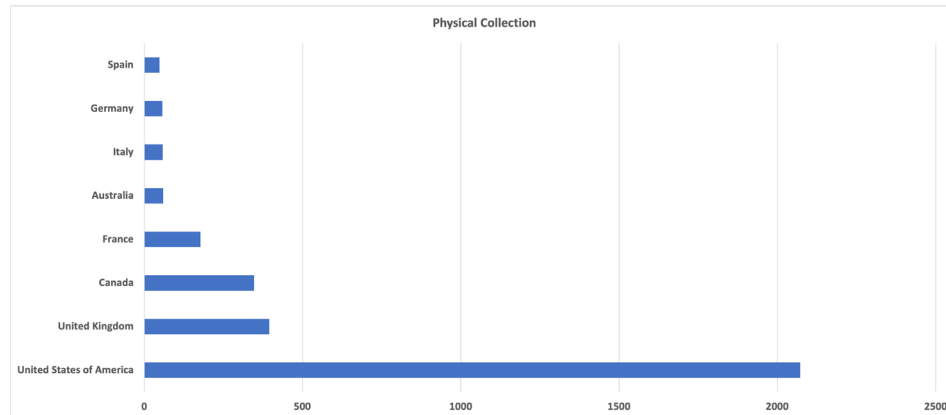


FIGURE 2 Country of origin for creators across physical collection.

Subject heading analysis similarly demonstrated the importance of title-by-title curation in ensuring adequate representation of EDI-related content in media collections. Starting with neutral terms, UTL's collection of licensed Kanopy titles contained the highest percentage of subject headings pertaining to each of the four categories. This is due to the curatorial nature of this collection; UTL licenses Kanopy titles only when they are required for course use. This has resulted in a relatively small collection of titles that are directly tied to current curricular trends and topics at UTL, particularly in the Arts, Humanities and Social Sciences areas. Similarly, UTL's MyMedia and physical collections (both of which consist at least partially of individually curated titles) contain a higher percentage of subject headings pertaining to EDI compared to an unmediated subscription-based collection like Criterion on Demand. Many of the films in the MyMedia Collection were added during the COVID-19 pandemic as the University of Toronto's classes moved online and remained at least partially so from March 2020 to January 2022. For faculty, alongside moving syllabus content online with streaming film, this time period coincided with the aforementioned movement to include more EDI-focused content to existing courses. 178 films have been added to the MyMedia Collection since early 2020.

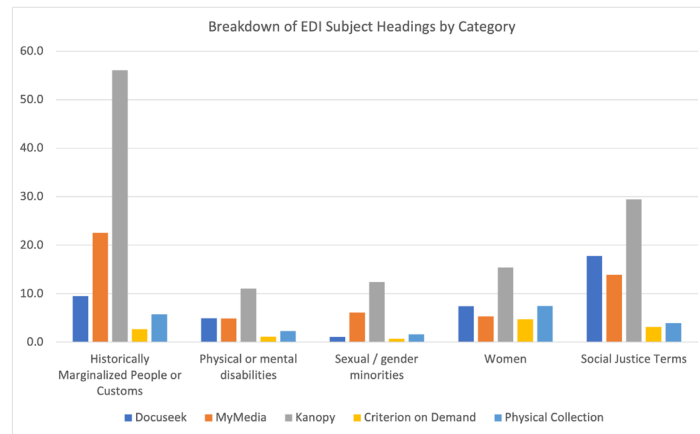


FIGURE 3 Breakdown of subject headings coded for neutral EDI terms.

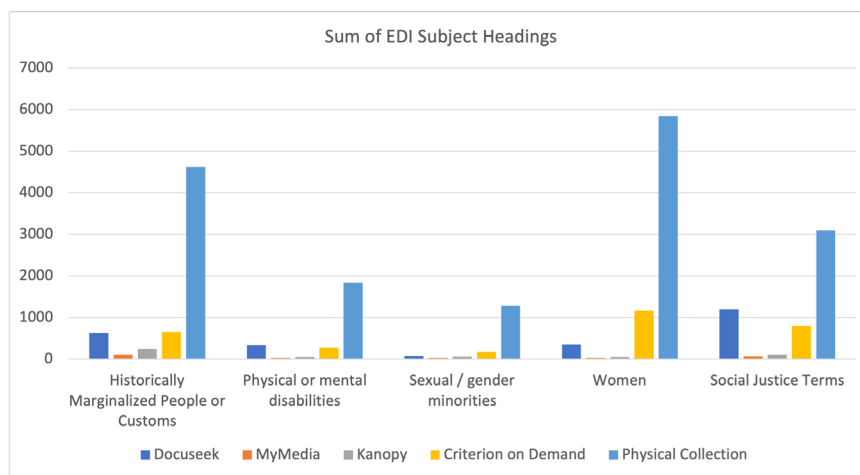


FIGURE 4 Sum of subject headings coded for neutral EDI terms by total number of subject headings analyzed.

In terms of general trends, among the social justice terms subject headings, content related to disability was low across all collections, both physical and streaming, while content related to gender and sex was highest across all collections. It is difficult to know whether there is less content in the collection or whether these headings were simply applied less frequently in cataloguing. The gender and sex category was more broad and could include, for example, “women detectives” or “male female relationships” as subject headings. Overall, the physical collection had the highest sum total number of EDI-related social justice subject headings, but when normalized by percentage it was Docuseek. To our original question of whether the physical collection brings more diverse content, it would seem that a more specialized streaming media platform with an explicit focus on “social issue” documentary film has the most diverse content for social justice terms (Docuseek bills itself as “the source for essential social issue and documentary film for education.”

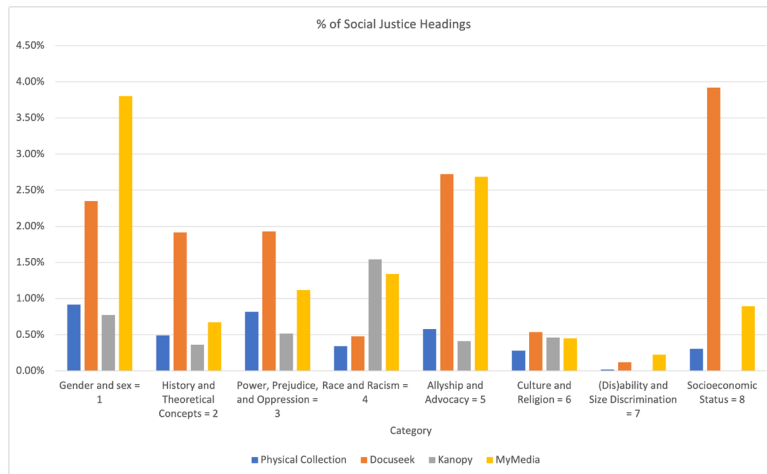


FIGURE 5 Social justice coded subject headings normalized by percentage.

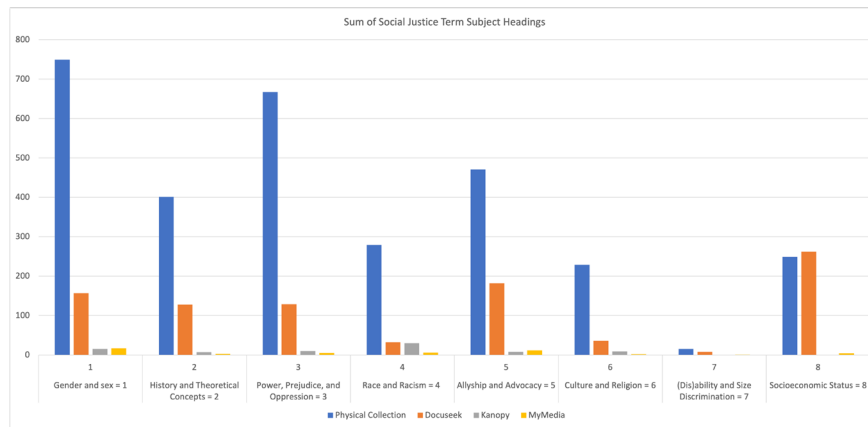


FIGURE 6 Social justice coded headings by total number of subject headings analyzed.

In evaluating this data, it is important to take sample sizes into account (our Kanopy collection, consisting of 619 titles, only contained 2077 subject headings, whereas our physical collection contained 28,411 titles and 81,625 subject headings). The physical collection is by far the largest and oldest one being examined, with content added over many decades. If we were to analyze MARC records for the entire Kanopy collection, numbers would likely be much closer to those for our physical collection. However, it is nonetheless noteworthy that UTL's highly curated Kanopy collection contained the highest percentage of EDI-related neutral subject headings out of all the collections analyzed. Similarly, UTL's 'MyMedia' collection contained the second or third highest percentage of EDI-related subject headings in the social justice terms. The importance of individualized title curation also tracks with University of Toronto's current teaching and research priorities around EDI-related topics.

Another important element to consider in our subject heading analysis, which may have affected the data, is how local cataloguing practices have changed over time. For example, it had been standard practice at UTL to limit the number of subject headings included in a record to a maximum of three. As a result, many of the older titles in our physical collection contain significantly fewer subject headings per title than newer titles in our physical collection and titles in our digital collections. This, in turn, limits the number of EDI-related subject headings within the metadata for these older titles. In addition, as film curricula have expanded to prioritize more diverse creators and content over time, the way in which subject headings get applied or prioritized to represent the aboutness of a film may change alongside the ability to include more subject headings.

One of the questions guiding this research was the anecdotal assumption the authors observed among librarians working with film collections, that a move away from collecting physical media and towards streaming platforms will lead to a loss of diversity in collections. While this paper attempted to address measurements of diversity across physical and streaming formats, it seems analysis missed another aspect of the initial assumption: the idea of choice, or curation. The perceived lack of diversity in streaming content comes from a perceived lack of titles to choose among, the idea that independent or older content may not appear on new streaming platforms, however the results of this paper seem to indicate that choice itself may be more important than format options in building a diverse collection.

In the analysis of subject headings, MyMedia, UTL's internal media hosting platform, showed more diversity in content than the physical collection. For example, MyMedia has social justice terms relating to gender and sex at close to 4 percent of total subject headings, whereas the physical collection only represents these subject terms at approximately 1 percent (see figure 5). This is significant as MyMedia and the physical collection allow the most curation by media librarians; the vast majority of films hosted on MyMedia are titles that have either been individually requested by course instructors for curricular use and approved for purchase, or selected by the media librarian. The physical collection also includes titles acquired on an approval plan, while MyMedia does not.

Among the streaming collections that offer set packages of titles, Kanopy also allows for the most curation, with librarians able to add or remove titles from the PDA pool or purchase titles outright at their discretion or faculty request. Kanopy's many diverse subject headings may, in part, come from a more recent prioritization of adding diversity to their offerings. As of 2020, the platform was "working with filmmakers and rights holders to expand the collection as the appetite for and interest in diversity, inclusion and social justice expand" (Louie 2020). While libraries

certainly benefit from this effort and as noted can choose which material to license, this trend does reflect that ultimately it is the platform that decides what content will be available to choose between, and when to prioritize diversity based on demand.

Discussion

This study's methodology demonstrated limitations in both the data collection and analysis stages. First and foremost, as explored in the literature review by Billey and Jahnke, Tenaka, and Palazzolo, the library catalogue is not designed to be an index of markers of diversity or intersectional subjectivity. We found a demonstrated lack in the ability to address the finite aspects of EDI and were only able to infer diversity at a higher level, for example creator gender identity is framed as binary and historically not always recorded, reflecting changes in cataloguing practices. The quality of the data collected for this paper is only as good as the data originally recorded for media items. As cataloguing personnel and practices change over time, the extent to which fields flagged as markers of diversity in this study were recorded for films over the years is likely inconsistent, and does not provide a complete picture. External data sources that are freely available can also be insufficient, with Wikidata lacking data for some creators we were able to identify, while some that were represented still lacked sufficient information. The sample sizes of our datasets were also a concern; our hope was that with larger datasets we were able to normalize discrepancies in the data, allowing us to derive some understanding. Our physical collection for example is much larger and older than our MyMedia collection. The amount of data in both of these collections however, when normalized by percentage, allowed us to understand the importance of responding to our community and current trends in collection needs.

As noted previously the catalogue is the most complete and accessible record of the content in any library collection, so by augmenting available information with a freely available external source like Wikidata, we attempted to develop a methodology that could be adapted by any library for its collection. Ultimately, this method proved insufficient and added a new dimension to our original research question: not only do we question whether ceasing physical media acquisition affects the diversity of our collection, but is this assumption measurable with the available tools?

Another key aspect of diversity for library collections is intersectionality, the interconnected nature of social categories such as race and gender that creates overlapping and interdependent systems of privilege or oppression. This becomes particularly relevant for the ways that subject matter and creator identity interact, and for the nuance subject headings fail to provide; if the subject heading "African

American Men” appears for a film, what treatment does this group receive in the film? Is it considered culturally accurate or stereotypically offensive? Is the creator an African American man? Does the creator’s identity change the way in which this film might be considered to add diversity and representation to the library collection beyond subject matter alone? Returning to the literature review, we face the same problem posed by Chu’s media analysis over 20 years ago, in which recording the nuance of diverse subjects requires reviewing each media individually, a task far too daunting for over 40,000 items.

Perhaps most pressing in terms of both data collection and analysis is the fallibility and bias of language. As previously noted, this begins with the catalogue but extends to the people doing the work of classifying film, and of interpreting this study. Even in the process of organizing the collected subject headings into categories, value, politicization, and bias entered into our description. Language is never neutral. Libraries and academia more broadly perpetuate harm and hegemony in their colonial structures. Cataloguing classification terms and the human labour that apply and interpret those terms are informed by these structures. The effort to analyze diversity in an academic library collection must acknowledge it was built by, for, and about a population that historically excluded Black, Indigenous, People of Colour (BIPOC and LGBTQ2S+ subjects and identities. Much like Hollywood filmmaking culture influencing the films and content available to libraries, academia as an institution influences the ways we think and speak about teaching materials, the language we use to describe and analyze.

Turning back to the question of format, this study cannot account for format availability in the past. For a film in our MyMedia collection, did we have the option to purchase a physical format at the time and choose streaming instead? Or was streaming the only available option? Streaming film subscriptions are still a relatively new access model and pose challenging questions about longevity and perpetuity for a library. Namely, are libraries still building long term collections, or are they simply renting content? The titles being analyzed for this study on educational platforms now are subject to the same licensing models as commercial platforms like Netflix, in which the user is responsible for tracking when films come and go from the platform. The idea of choice and curation is out of the hands of users, including librarians. The content is ephemeral as licenses expire, and subsequently as are the results of this study for platforms like Kanopy or Criterion on Demand; the number of films with EDI markers on each will change over time. Presently, educational streaming platforms have shared libraries’ focus on offering diverse and timely content, such as promoting collections and titles for Pride Month or Black History Month, but the long-term availability of these offerings is unknown.

As Jahnke et al. note for print acquisitions, cost-saving strategies such as leasing content rather than purchasing “have largely favoured vendor relationships and are part and parcel of a larger dynamic of business logic within higher education that privileges demand and immediate need over long-term support for scholarship and equity” (Jahnke, Tanaka, and Palazzolo 2022, 179). Similarly, the media collection development trend towards leasing is associated with streaming as a format, and “long-term support” with enduring physical collections. It is arguably too early in the history of streaming as a format to make definitive claims about collection content implications, but it is clear that long term access be it physical or digital ought to be a core consideration in building diverse collections.

Building on this research, librarians could take titles surfaced with the same subject term and creator identity data and compare with WorldCat holdings of physical media, to determine which physical copies might be rare or unavailable at other libraries, across formats. The holding library could also examine circulation data for the same titles, to determine how recently and often they have been used. While the presence of these items in the collection contributes to EDI mandates for collecting, it is their usage that influences positive changes in curriculum. Usage data also often guides the deselection process, but comparing this with rarity and availability can indicate items that could not be acquired again should interest return for study for a low-circulation item. Particularly with the question of format, it is worth noting which films with EDI content have no modern streaming options, to retain physical copies.

As suggested in the literature review, other related directions could include the use of bibliography, or rather filmography, to identify gaps in existing collection holdings. While such lists can be reductive, UTL as an institution has had success in creating bibliographies in collaboration with faculty or scholars who teach and specialize in EDI-related topics. This could be another opportunity to bring UTL scholar’s unique perspectives into collection development, rather than simply relying on existing checklists. Also mentioned in the literature is the fact that selection for both EDI supporting materials and media more generally is a time-consuming process that frequently involves working with independent creators or vendors and varied fee structures. One additional gap to explore from an academic personnel perspective could be tracking and reporting the time, labour, and monetary cost that goes into acquisitions for physical or streaming media that is considered to add an EDI perspective to the collection.

For UTL, we are considering applying this same study methodology to our sound collection, and broadening samples, such as including all available titles from Kanopy rather than only the titles we currently license.

Conclusion

This study made clear the difficulty of developing a methodology to analyze large sets of catalogue data that might indicate that the content or creator of a film contribute to EDI markers in library collections. The limitations of what information gets recorded in catalogue records, how this changes over time, and the difficulty of defining diversity with predetermined categories of metadata fields all contributed to data collection and analysis that, while not perfect, offered some insight into the original research question and format comparison. Ostensibly, the main concern for librarians in the shift from physical to streaming film collections is format, however based on this study it is important not to overlook curation, the ability to be selective and intentional about what content gets added to library collections. Physical collections at UTL and many academic libraries represent decades of collecting based on the selection of library staff and requests or recommendations of faculty and students, aligned with research and curriculum needs. These also reflect trends within academia, including ongoing efforts towards diversifying the content and voices in library collections through EDI mandates. UTL's MyMedia platform, while offering access to streaming film, is also a more curated collection as it hosts one-off purchase titles. Vendor streaming collections on the other hand are a newer and lesser-known form of acquiring and accessing content, most often offering a platform or package of preselected films. Like commercial streaming platforms, content may come and go, or be licensed for a single term rather than perpetual access. As Fossati notes in exploring the rapid and unceasing evolution of formats for film, "if we consider transition as an inherent property of media, technological hybridism is its characteristic" (Fossati 2021, 20). Format will always evolve, it is choice and variety and the opportunity for curation that matters most, the ability to continue to pick and choose the formats we collect to ensure the diversity of creators and ideas represented, alongside a diversity of access and acquisitions models.

ABOUT THE AUTHORS

Amelia Clarkson is the Media Collections Librarian at the University of Toronto Mississauga Library. Her research interests include collection access and preservation; digital platforms as pedagogical tools; information literacy and media literacy.

Trevor Deck, Music & Film Collections Management Librarian at U of T Libraries, holds an MLIS from UBC (2014). His research interests include collection assessment, collaborative collections and open access.

James Mason MMus, MLIS is a Metadata and Digital Initiatives Librarian at the University of Toronto Music Library. He conducts research in access and assessment related inquiries.

REFERENCES

Alimurung, Gendy. "Movie Studios Are Forcing Hollywood to Abandon 35mm Film. But the Consequences of Going Digital Are Vast, and Troubling." *LA Weekly*, April 12, 2012. <https://>

www.laweekly.com/movie-studios-are-forcing-hollywood-to-abandon-35mm-film-but-the-consequences-of-going-digital-are-vast-and-troubling/.

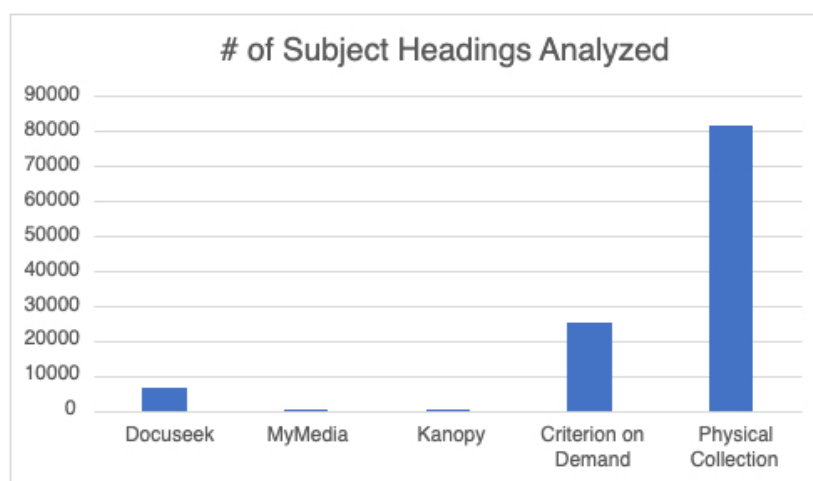
- American Library Association. 2020. "Libraries Respond: Black Lives Matter." Advocacy, Legislation & Issues, June 3, 2020. <https://www.ala.org/advocacy/diversity/librariesrespond/black-lives-matter>.
- Billey, Amber. 2022. "Recording Gender: An Ethical Cataloging Conundrum." *Medium* (blog), November 18, 2022. <https://medium.com/@amberbilley/recording-gender-an-ethical-cataloging-conundrum-1cee978c1186>.
- Buchanan, Kyle, and Reggie Ugwu. 2020. "Why Are There So Few Black Directors in the Criterion Collection?" *The New York Times*, August 20, 2020, sec. Movies. <https://www.nytimes.com/interactive/2020/08/20/movies/criterion-collection-african-americans.html>.
- Chu, Clara M. 2000. "See, Hear, and Speak No Evil: A Content Approach to Evaluating Multicultural Multimedia Materials." *Reference & User Services Quarterly* 39 (3): 255–64.
- Ciszek, Matthew P., and Courtney L. Young. 2010. "Diversity Collection Assessment in Large Academic Libraries." *Collection Building* 29 (4): 154–61. <https://doi.org/10.1108/OI604951011088899>.
- Clarke, Rachel Ivy, and Sayward Schoonmaker. 2019. "Metadata for Diversity: Identification and Implications of Potential Access Points for Diverse Library Resources." *Journal of Documentation* 76 (1): 173–96. <https://doi.org/10.1108/JD-01-2019-0003>.
- Docuseek. n.d. "Docuseek | About Docuseek 2." Accessed June 20, 2024. <https://docuseek2-com.myaccess.library.utoronto.ca/about>.
- Fossati, Giovanna. 2009. "Framing Film in Transition." In *From Grain to Pixel: The Archival Life of Film in Transition*, 13–14. Amsterdam University Press. <https://doi.org/10.2307/j.ctv8bt181>
- Gates, Racquel. 2020. "The Problem With 'Anti-Racist' Movie Lists." *New York Times (Online)*, July 17, 2020, sec. Opinion. <https://www.nytimes.com/2020/07/17/opinion/sunday/black-film-movies-racism.html>.
- Gray, Sharon, and Teresa Abaid. 2002. "Building a Multicultural Media Collection." *MC Journal: The Journal of Academic Media Librarianship* 8 (2): 9.
- Guidelines for Media Resources for Academic Libraries in Higher Education Task Force of the Association of College and Research Libraries. 2006. "Guidelines for Media Resources in Academic Libraries (2018 Revision)." Association of College & Research Libraries (ACRL), September 6, 2006. <https://www.ala.org/acrl/standards/mediareources>.
- Jahnke, Lori M., Kyle Tanaka, and Christopher A. Palazzolo. 2022. "Ideology, Policy, and Practice: Structural Barriers to Collections Diversity in Research and College Libraries." *College & Research Libraries* 83 (2): 166–83. <https://doi.org/10.5860/crl.83.2.166>.
- King, Rachel. 2014. "House of Cards: The Academic Library Media Center in the Era of Streaming Video." *The Serials Librarian* 67 (3): 289–306. <https://doi.org/10.1080/0361526X.2014.948699>.
- Louie, David. 2020. "San Francisco-Based Streaming Service Curates 'powerful' Social Justice Films Focused on Diversity, Inclusion." ABC7 San Francisco. November 26, 2020. <https://abc7news.com/kanopy-san-francisco-films-about-social-justice-movies/8284703/>.
- MacDougall, Ruby, and Dylan Ruediger. N.d. "Teaching with Streaming Video." Accessed May 11, 2023. *ITHAKA S&R*. <https://sr.ithaka.org/publications/teaching-with-streaming-video/>.
- Misuraca, Michelangelo and Maria Spano. 2010. "Unsupervised Analytic Strategies to Explore Large Data Collections." In *Text Analytics: Advances and Challenges*, edited by Domenica Fioredistella, Damon Mayaffre, and Michelangelo Misuraca, 17–28. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-52680-1_2.
- Proctor, Julia. 2020. "Representation in the Collection: Assessing Coverage of LGBTQ Content in an Academic Library Collection." *Collection Management* 45 (3): 223–34. <https://doi.org/10.1080/01462679.2019.1708835>.

- . 2018. "Testing Assumptions: Is Streaming Video Really Preferable to DVDs?" *Journal of Electronic Resources Librarianship* 30(2): 84–90. <https://doi.org/10.1080/1941126X.2018.1465516>.
- Sims, Yelana, Nina Lorenz, Nina Cartier, Melissa Phruksachart, Tony Tran, Aju Basil James, Crystal Camargo, et al. 2021. "Looking Back, Thinking Forward: A Digital Humanities Assessment of Equity, Diversity, and Representation in Twenty Years of Publishing in Cinema Journal and JCMS." *JCMS: Journal of Cinema and Media Studies* 61(5): 1–30. <https://doi.org/10.1353/cj.2021.0085>.
- Smith, Stacy L. 2023. "Inclusion at the Oscars," USC Annenberg Inclusion Initiative and Adobe Foundation, accessed August 2, 2023. <https://www.inclusionlist.org/oscars>.
- Tanasse, Gisèle. 2021. "Implementing and Managing Streaming Media Services in Academic Libraries: Choice White Paper No. 8." *ACRL/Choice*. <https://www.choice360.org/research/implementing-and-managing-streaming-media-services-in-academic-libraries/>
- The Science & Technology Council, The Academy of Motion Picture Arts and Sciences. 2007. "The Digital Dilemma 2: Perspectives from Independent Filmmakers, Documentarians, and Nonprofit Audiovisual Archives." <https://www.oscars.org/science-technology/sci-tech-projects/digital-dilemma-2>.
- . 2012. "The Digital Dilemma 2." <https://www.oscars.org/science-technology/sci-tech-projects/digital-dilemma>.
- Tillay, Rachel, and CC Chapman. 2019. "A New Digital Method for Assessing the Diversity of Creator Identities in the Howard-Tilton Memorial Library Media Services DVD Collection." *Music Reference Services Quarterly* 22 (1–2): 57–64. <https://doi.org/10.1080/10588167.2019.1601421>.
- Tran, Michael T. 2022. "The Racial Barriers in U.S. Independent Filmmaking." *Sociological Inquiry* 92, (2): 490–513. <https://doi.org/10.1111/soin.12469>.
- Trinity Western University. N.d. *CRC Public Accountability Page* | Trinity Western University. Accessed July 31, 2023. <https://www.twu.ca/research/canada-research-chairs/crc-public-accountability-page>.
- University of Toronto Libraries. N.d. *Inclusion, Diversity, and Equity Statement* | University of Toronto Libraries. Accessed July 31, 2023. <https://oneresearch.library.utoronto.ca/inclusion-diversity-and-equity-statement>.
- UTL Collections Committee. 2022. *UTL Central Library Collections Diversity Plan*. https://oneresearch.library.utoronto.ca/sites/default/public/collection_diversity_rev.pdf.

Appendix

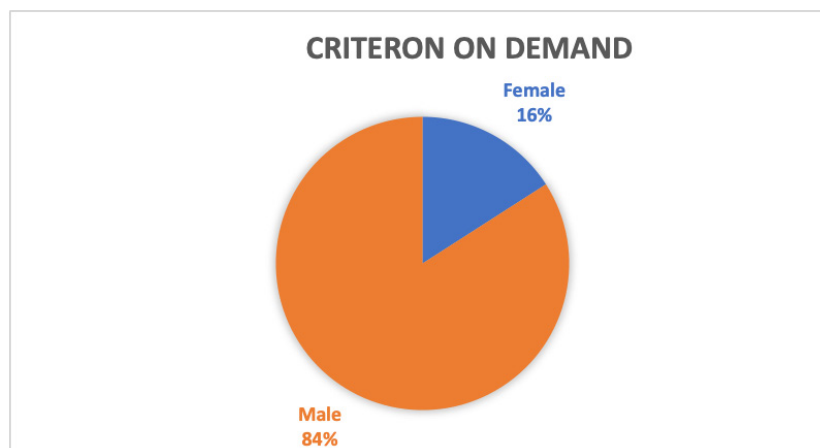
Counts of films and subject headings analyzed by collection:

Collection	Records	Total Subject Headings Count	Unique Subject Headings Count
Docuseek	1807	6687	3112
Kanopy	619	1944	573
Criterion on Demand	8433	25207	4217
Physical Collection (DVD/Blu-ray)	28,411	81,625	12,453
MyMedia	137	447	333

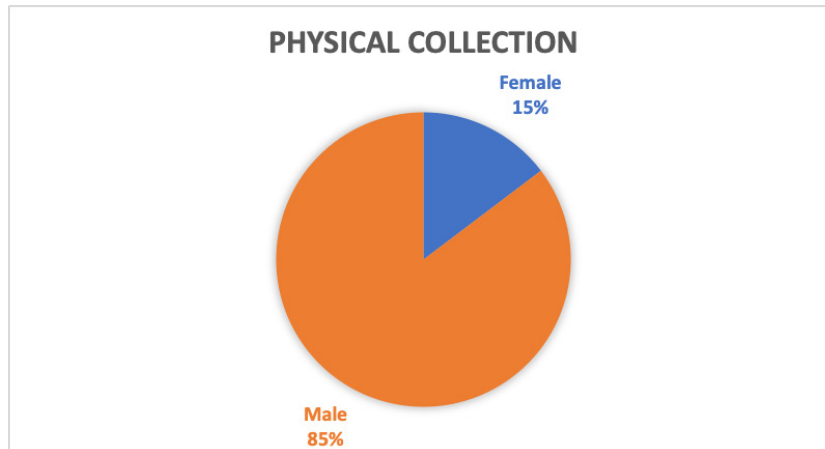


Creator Demographics by Gender*

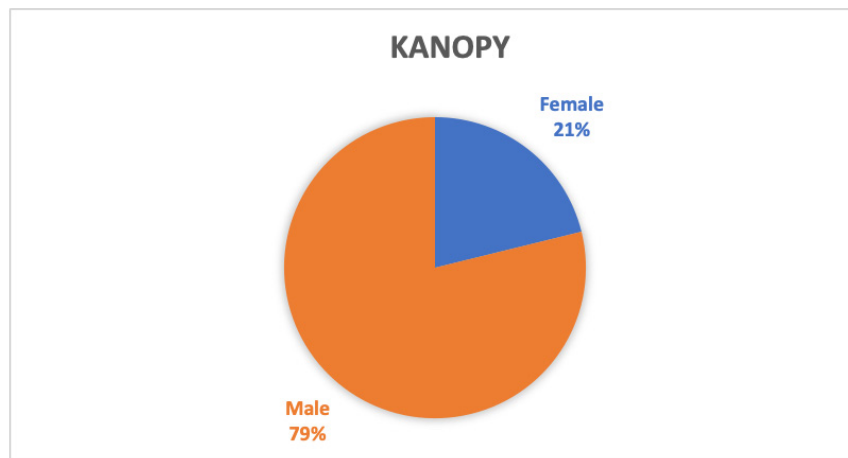
*Creators include individuals identified as directors, screenwriters, producers, editors or filmmakers



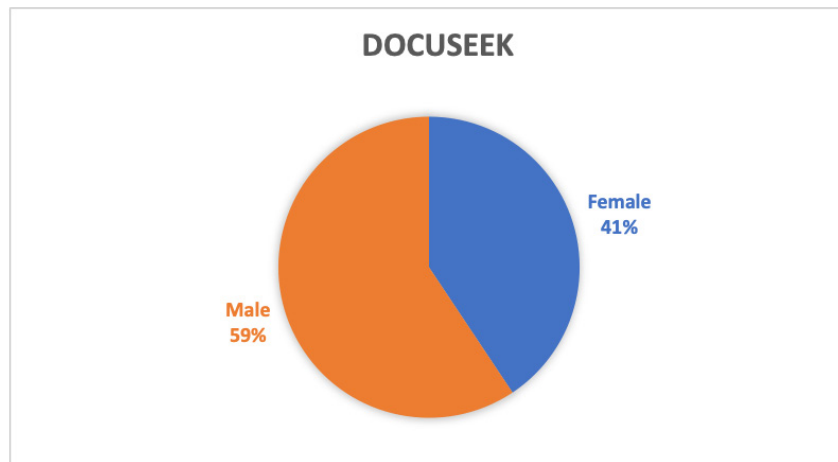
Criterion on Demand	Count
Female	254
Male	1341
Nonbinary	2
Transgender	2



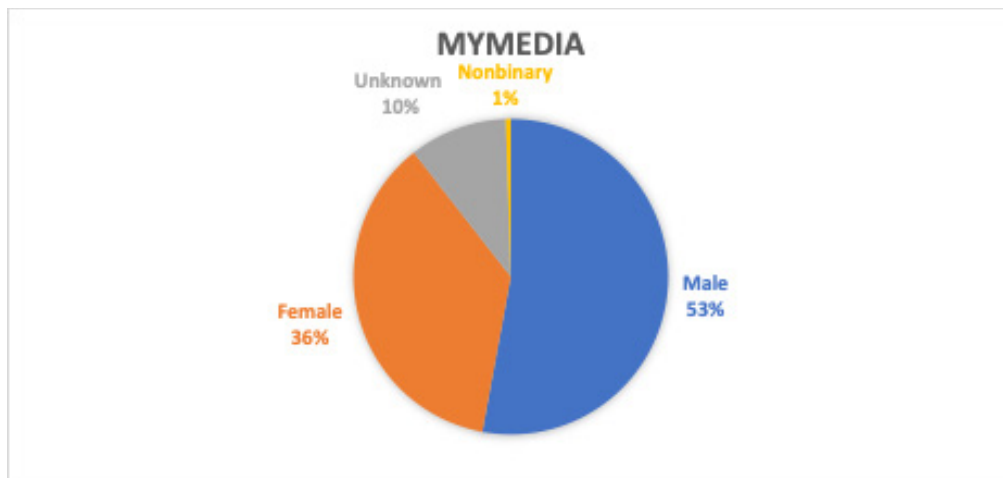
Physical Collection	Count
Female	1462
Male	8492
Intersex	1
Nonbinary	2
Transgender	2



Kanopy	Count
Female	66
Male	246

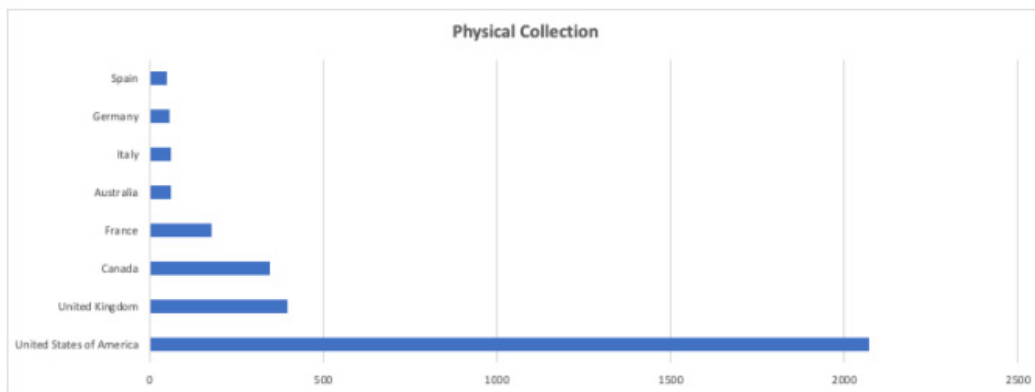


Docuseek	Count
Female	252
Male	368
Intersex	1

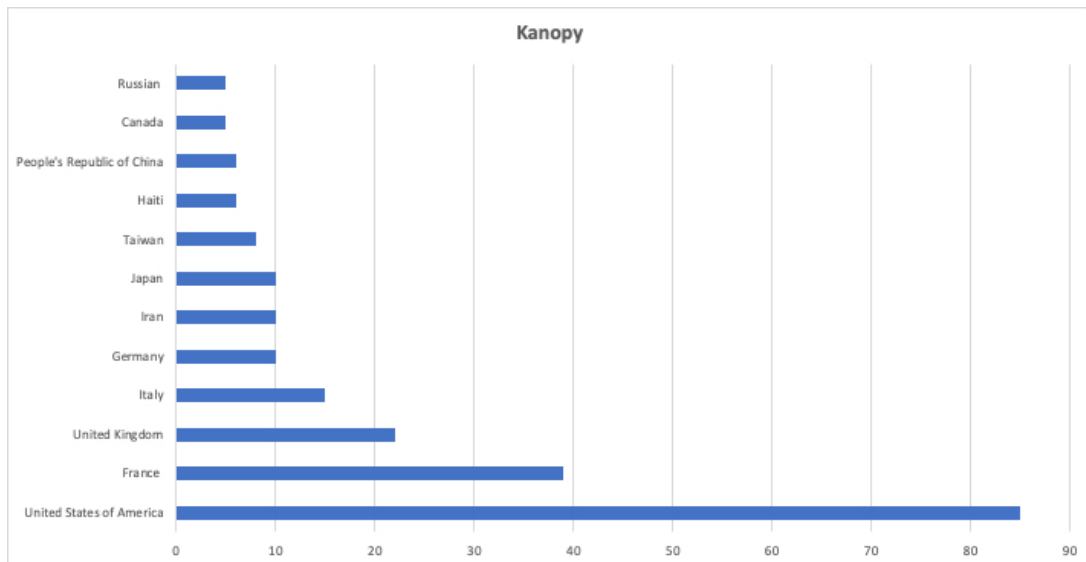


Sex	Count
Male	100
Female	69
Unknown	19
Nonbinary	1

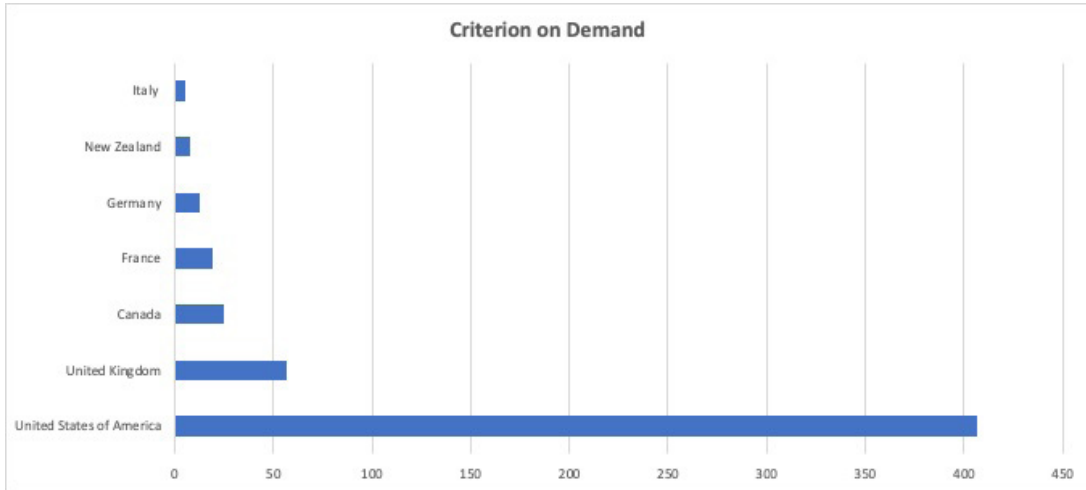
Creator* Demographics by Country of Birth



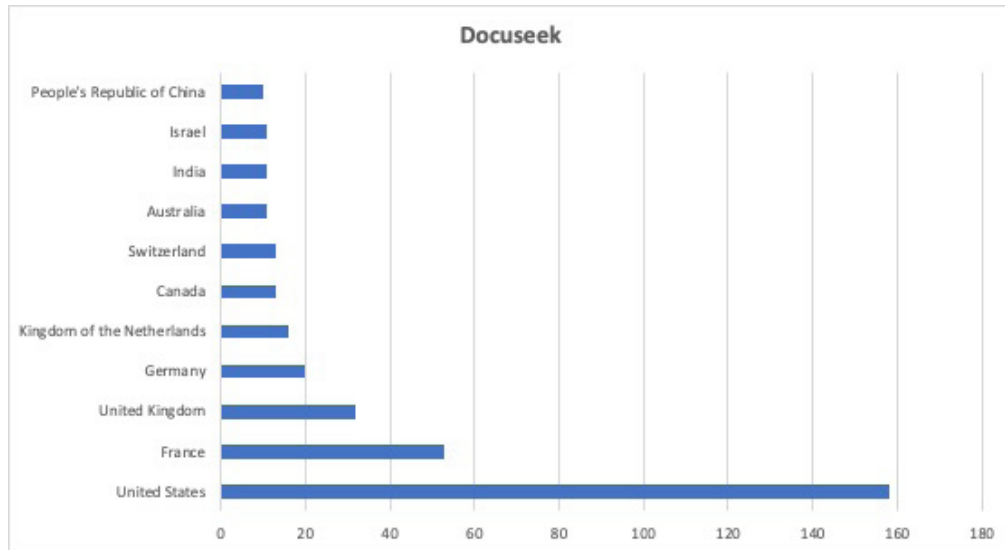
**46 other countries with <32 creators



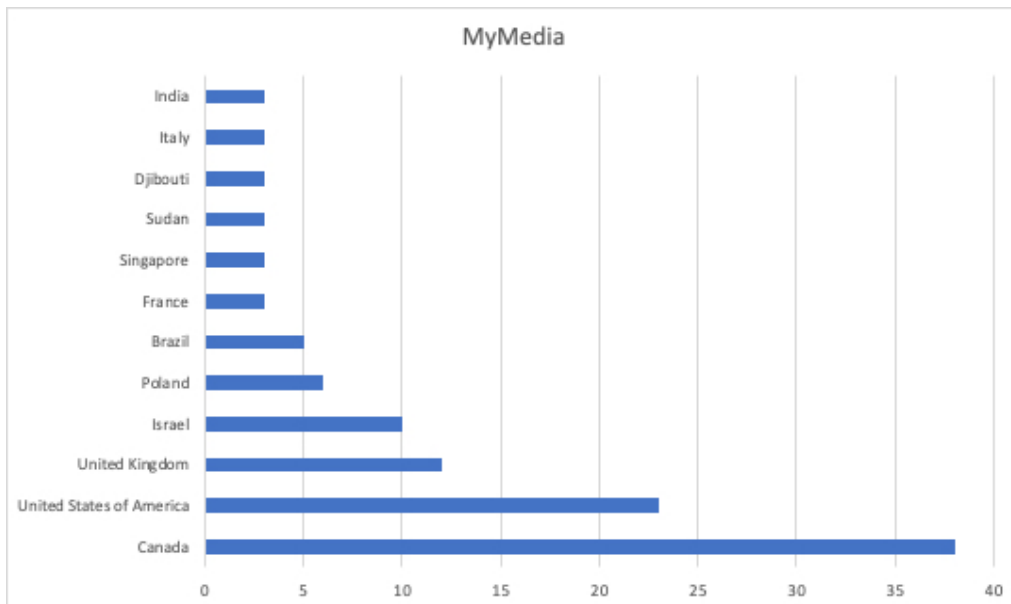
** 25 additional countries with <5 creators



* 17 additional countries with <5 creators



*38 additional countries with <10 creators



Subject Heading Analysis

Social Justice Terms

