

The Invisible Default

Examining Representation in Digital Collections

Kaylee P. Alexander, Dorothy Terry, Jasmine Kirby,
Rachel Jane Wittmann, and Anna Neatrour

ABSTRACT

This mixed-method study investigates the representation of race and ethnicity within the J. Willard Marriott Digital Library at the University of Utah. The digital collections analyzed in this study come from the Marriott Library's Special Collections, which represent only a fraction of the library's physical material (less than 1 percent), albeit those most public facing. Using a team-based approach with librarians from various disciplines and areas of expertise, this project yielded dynamic analysis and conversation combined with heavy contemplation. These investigations are informed by contemporary efforts in librarianship focused on inclusive cataloging, reparative metadata, and addressing archival silences. By employing a data-intensive approach, the authors sought methods of analyzing both the content and individuals represented in our collections. This article introduces a novel approach to metadata analysis—as well as a critique of the team's initial experiments—that may guide future digital collection initiatives toward enhanced diversity and inclusion.

INTRODUCTION

People often want to see themselves—their identities, interests, and communities—reflected in library and archival collections, be it for the purpose of conducting historical or genealogical research or for community discovery. When addressing tactics toward building more inclusive collections and increasing the discoverability of historically underrepresented histories, Dorothy Berry cites historian Edgar Porter Alexander: “Take me to the museum and show me myself, show me my people, show me soul America.”¹ But do library collections also reflect these values? And if not, how might they? Like museum collections seeking to do such work, library collections can become controversial, raising questions about whose stories can or should be preserved and from which perspectives the public should be able construct their sense of self.² Digitized library collections surface additional complications, as only a small fraction of physical material may be selected for the additional workflows and attention required by digitization, descriptive metadata, online access and display, and digital preservation. And yet, the digital collection is often also the most accessible, public-facing aspect of a library. As such, it is important that library collections—and especially digital ones—reflect the communities they serve and are embedded within. But this ideal is often complicated by historical biases that privilege some identities over others and collecting practices that rely on donations and bequests. As libraries and archives work toward

About the Authors

Kaylee P. Alexander (kaylee.alexander@utah.edu) (corresponding author) is Research Data Librarian, J. Willard Marriott Library, University of Utah. **Dorothy Terry** (dorothy.terry@utah.edu) is Librarian for Student Success, J. Willard Marriott Library, University of Utah. **Jasmine S. Kirby** (jskirby@odu.edu) is Online Learning Librarian, Old Dominion University. **Rachel Jane Wittmann** (rachel.wittmann@utah.edu) is Metadata Strategies Librarian, J. Willard Marriott Library, University of Utah. **Anna Neatrour** (anna.neatrour@utah.edu) is Head of Digital Library Services, J. Willard Marriott Library, University of Utah. © 2025.

Submitted: 3 December 2024. Accepted for Publication: 15 June 2025. Published: 15 September 2025.

more inclusive practices and collections development, gaining a critical, aggregate understanding of how their present collections are composed is a crucial first step.

The J. Willard Marriott Digital Library (hereafter, the Digital Library) was established at the University of Utah in the early 2000s, with a long-standing practice of digitizing materials from the Marriott Library's Special Collections, as well as providing repository services to additional partners from the University of Utah, state agencies, museums, historical societies, and public libraries throughout the state of Utah. The repository contains over one million items and has over four hundred digital collections with a focus on areas that reflect the University of Utah's role as a Research 1, predominantly white institution (PWI). Approximately 300,000 items from Special Collections are digitized and available in the Digital Library, constituting an exceedingly small fraction of items housed in Special Collections.

Digital Library Services at the J. Willard Marriott Library is a separate department that collaborates with Special Collections to build the Digital Library. Digitization priorities are often determined through various factors including the fragility of the item, copyright and deed of gift considerations, donor requests and grants, and research value. Recently prioritized items for digitization and inclusion in the Digital Library include materials that reflect historically underrepresented groups. These materials include a variety of oral histories that were completed from the mid-1980s to 1990s focused on specific groups in Utah, such as Hispanic Oral Histories and Interviews with African Americans in Utah.³ The Mitsugi M. Kasai Memorial Japanese American Archive within Special Collections includes manuscripts, photos, and oral histories that focus on the Japanese American experience in Utah, and was built over time due to an archivist's extensive ties and years-long continuous outreach to the community.⁴

Based on our awareness of current efforts in the field of librarianship centered on inclusive cataloging, reparative metadata, and archival silences, we wanted to assess our digital collections to see whether they adequately reflected regional demographics. In taking a data-intensive approach to understanding not only what but who is represented in our collections, we hope to demonstrate an original method of metadata analysis that can inform future digital collections endeavors geared towards diversity and inclusion.

POSITIONALITY

Before addressing our methods and preliminary findings, it is important to call attention to the specific characteristics of our institution and research team that affect the composition of our digital collection, the present-day relevance of investigating representation in our collections, and our methodological approach.

The University of Utah and Salt Lake City

Both the University of Utah and Salt Lake City are becoming more racially and ethnically diverse with nearly 73 percent of Utah's growth in 2023 being attributed to people identifying as something other than non-Hispanic white.⁵ Additionally, according to the University of Utah's Office of Budget and Institutional Analysis, 32 percent of the university's domestic incoming freshmen in 2022 were students of color; this is compared to 18 percent in 2009.⁶ Not only is the student population diversifying, but the university has also been engaged in revitalizing its Middle East Center and establishing a new Center for Pasifika Indigenous Knowledges. With this growth, it is vital that library users can see themselves reflected in our holdings, both physical and digital. While this project was underway, the Utah State Legislature passed House Bill 261 (HB 261). HB 261, or Equal Opportunity Initiatives, prohibits state-funded institutions of higher education from

using certain individual characteristics (e.g., race, ethnicity, gender, sexuality) in making employment or education decisions.⁷ Consequently, the University of Utah's cultural centers—some of which had served students for several decades—were closed.⁸ As legislators and parental rights groups in Utah and across the United States are increasingly pushing to remove materials from libraries that embrace our differences and limit the ways in which we can engage with diversity, equity, and inclusion, it is imperative that we continue to seek out and grow collections that adequately represent our population.⁹ The pushback against our diverse communities and history in Utah only makes our work more important.

Our Research Team

Our research team consisted of one Digital Library department manager, one metadata librarian, two public service librarians, and one research data librarian. Our group represented a mix of white and African American women of diverse sociocultural, religious, and educational backgrounds. The team members' backgrounds enabled robust and nuanced discussions about metadata terms that functioned as indicators of race and ethnicity as we encountered them throughout the project. Notably, however, none of the researchers on this project are originally from Utah, and many of the university's collections are related to local and regional histories. Consequently, greater attention may have been allocated to coding familiar material than to coding material that was less familiar to the authors.

ARCHIVAL SILENCE AND DIGITAL LIBRARIES

The Society of American Archivists (SAA) offers a succinct definition of archival silence as “a gap in the historical record resulting from the unintentional or purposeful absence or distortion of documentation.”¹⁰ Digitized material from special collections and archives can compound archival silences but offer opportunities to grapple with the consequences. Marlene Manoff, for example, states, “The digital archive includes content produced and selected by individuals in particular social and historical contexts, but it is also shaped by multiple additional factors, including the hardware and software that enable access to and manipulation of that content.”¹¹ Another aspect of digitization can be highlighted in the privilege associated with archives in the global north, as addressed by Michael Moss and David Thomas, who write in their book *Archival Silences* that “digitisation is silencing records because scholars at wealthy universities in the north have access to a diversity of material unavailable to scholars at poorer universities in the south. At the same time, digitisation programmes may mean that records that are non-western, non-canonical and non-quotidian are excluded, thus continuing to silence the poor and marginalised.”¹² Writing from a perspective of a records manager, Eira Tansey additionally points out that some of the causes of archival silences—such as lack of cooperation from records creators and insufficiently funded archives programs—can combine to create a digital dark age wherein “state archives have reported that there is a consistent gap between the authority to carry out state records policies, and the resources needed to actually perform or deliver duties and services. Archivists with institutional records mandates rarely have the authority or resources to go out and get all the electronic records on their own that are required to be transferred to the archives.”¹³

Interventions with the issue of archival silences, both physical and digital, can take a variety of forms. S. L. Ziegler points out how “For special collections and archives, self-censorship can mean failing to describe certain portions of a collection in a finding aid or restricting access to material.”¹⁴ The DC History Center has come to similar conclusions, having undertaken a gap analysis of their physical collections, ultimately cautioning that “this process also necessitates assumptions. Assumptions about race and sexual preference create additional potential for

erasure.”¹⁵ The potential for erasure is also prominent in digital collections metadata, as Lauren F. Klein has demonstrated in her analysis of the digital edition of the Papers of Thomas Jefferson. Klein draws attention to the fact that archival silence often manifests not in the absence of material related to a given person or community, but in the biases of the metadata record, which may exclude terms related to individuals or topics deemed, for one reason or another, irrelevant at the time of cataloguing. While Klein used text mining on digitized collections to identify and address gaps in metadata, our focus was on understanding the broader collections landscape our metadata reflected to users.¹⁶

DEFAULTING TO WHITENESS IN METADATA

Ensuring accuracy, discoverability, and access to resources are of the utmost importance in creating descriptive metadata for digital collections. When identifying race and ethnicity in digital collections’ metadata, the traditional practice has been to assign racial and ethnic identifiers only to nonwhite persons, especially in cases where most creators and subjects were white. Given the nature of most academic special collections consisting of historically affluent and privileged donors, these archival collections tend to be “shaped by white hegemony” and therefore dominated by white people.¹⁷ To decenter white dominance in archival metadata, several institutions have implemented conscientious and inclusive metadata practices that provide guidelines for metadata creators to identify white persons as they would identify persons of color.¹⁸ The University of North Carolina Libraries, Special Collections, and Technical Services Style Guide, for example, provides useful background information and examples of how to include white identification. They note that “Since August 2017, [they] have described the ethnic and racial identities of all creators and collectors so that whiteness is no longer the invisible norm.”¹⁹ Metadata indicating the whiteness of depicted subjects would have allowed for improved assessment and comparisons of ethnic representation in the Digital Library. Widespread best practices for approaching whiteness in metadata have yet to be adopted in the field, but these issues were kept at the forefront as the team began their coding exercises with the data from the Digital Library.

CODING TRIALS AND TRIBULATIONS: A DISCOURSE ON METHOD

Due to the more than 300,000 records in our digital collections, we faced several hurdles in identifying records related to historically underrepresented peoples. With this volume of material, we could not look at each object individually to determine representation, though approaching our metadata as a text corpus that could be mined and coded in the aggregate provided several avenues for experimentation and analysis.

We understood that our metadata would not be perfect but relied on the assumption that whiteness would be the “invisible” default.²⁰ Consequently, “white”—and, by proxy, all other “default” identities in the US context including Protestant, heterosexual, and cisgender—would not be mentioned in the metadata for the simple fact that it would not have been understood as necessary to call out. In the interest of concision, we use the term *nonwhiteness* somewhat metaphorically in the sections below as a catchall for a variety of “nondefault” categories in our metadata, including people of color, underrepresented religious groups, and other historically underrepresented individuals and communities. Whiteness and underrepresentation were not mutually exclusive classifications in our study as we recognize that one can be both white and nondefault, as in the case of religious minorities and some immigrant groups. Therefore, when the term *nonwhiteness* appears in our methods and analysis, it should not be taken literally as nonwhite-skinned.

There were, undoubtedly, objects in our collections that related to or represented nonwhite individuals but were not detectable in an analysis of the metadata alone, due to human error in metadata generation or past frameworks that deemed these subjects tangential or unremarkable. Still, we expected that, for most records featuring nonwhite subjects, the metadata would name them as such, especially in cases when they were a prominent feature of the item in question.

Bearing these assumptions in mind, we approached the collections metadata as a corpus in which each record's descriptive metadata represented a "text." We began by downloading the metadata for all records in the Digital Library ($n = 302,551$ as of December 2023) and cleaning the dataset to remove extra fields that did not function as core descriptive or identifying metadata; the columns we maintained for our analysis are summarized in table 1. In selecting the fields to analyze, we considered which fields might provide text-based indicators of nonwhiteness—such as terms relating to race, ethnicity, religion, or tribal affiliation in keywords, controlled subject vocabulary, and descriptions—as well as fields that would provide additional information for better understanding the records' context and the collecting patterns within our library (e.g., digitization date, format, genre). The fields represented in table 1 are a combination of technical fields, standardized Dublin Core fields, and local fields. Once we had narrowed the metadata fields we needed for our analysis, we began a process of collaborative coding that involved 1) identifying subjects and keywords that could be indicators of nonwhiteness and 2) coding each record with either a 1 (indicating the presence of a "nondefault" attribute) or 0 (indicating implied whiteness or other "invisible," default characteristics). Using OpenRefine, an open-source application used primarily for cleaning and wrangling "messy data,"²¹ we faceted descriptive metadata fields and filtered on terms commonly associated with nonwhiteness.²²

Table 1. Overview of Digital Library metadata fields maintained for analysis

Field	Description	Format	Role
Id	Unique internal identifier for the record	numeric	identifier
Title	Title of the record	text	context
Ark	Ark ID for the record	PID	identifier
collection_name	Name of larger collection to which item belongs	text	indicator
Creator	Name of the creator if known	text	context
date_digital	Date when item was added to digital collections	date	context
Date	Date associated with item's creation	date	context
description	Brief textual description of the item	text	indicator
Format	Item format: application/pdf, audio/mpeg, image/jpeg, text/html, video/mp4	categorical	context
Genre	Type of item (e.g., correspondence, photograph, book, manuscript, VHS, slide, etc.)	text	context
keywords	Set of keywords associated with the item	text	indicator
Language	Item language in ISO 639 Language code	categorical	context
outdated_terms	Offensive and/or outdated terms that were previously used in the descriptive metadata	text	indicator

Subject	LCSH terms associated with the item	text	indicator
Topic	Topics associated with the item	text	indicator
Tribe	Name of tribe associated with the item if related to an Indigenous group	text	indicator
Setname	Internal code for set to which item belongs	text	indicator

For our direct, batch coding process, we focused primarily on the two core metadata fields that contained LCSH terms and keywords: subject and keywords. We began with relatively obvious terms indicating nonwhiteness—*African American*, *Indigenous*, *Native American*, *Jewish*, etc.—that were unlikely to capture relevant words within irrelevant words (e.g., searching for the word stem *Jew* would have also matched on records containing the word *jewelry*) or terms that could be more context dependent (e.g., Black vs. black). The reverse also worked for batch coding records unrelated to those we were interested in measuring; filtering for terms related to geographical features and other materials that make up larger portions of our collections but do not relate directly to any specific identity, historically marginalized or otherwise, allowed us to pare down the number of records that needed coding. Still, we were left with tens of thousands of records that could not be coded in bulk. The necessity of applying a binary decision without accounting for the subtleties of the descriptive metadata and the close reading that would have been involved with studying the visual qualities of each digital object was often uncomfortable, inspiring reflection and discussion during the coding process. The faceting, clustering, and filtering features of OpenRefine, however, often helped us to mitigate issues by allowing us to code subsets of the data within their specific collections contexts or by helping us to identify related terms and phrasings that we had not anticipated.

As our study focus was on how metadata specifically spoke to issues of identity and representation, we avoided reviewing images to make determinations of whether a record counted. We did, however, use other metadata fields (i.e., *setname* and *collection_name*) to help limit the scope of our data into manageable subsets that provided an idea of which materials were likely to have relevant records. Reviewing the data at the collection level also provided an opportunity to perform quality control checks on our initial keyword-based coding and identify items that had potentially been miscoded. Coding at the collection level, however, was also an iterative process and different collections allowed for different levels of coding precision. One collection, for example, comprised individual personnel records that accounted for each employee's country of origin (see fig. 1).²³ This meant that we were able to isolate these records and comprehensively code them on an item level rather than assigning a homogenous 1 or 0 to the collection. This level of detail was not available for other collections, and we often resorted to batch coding items that had not been coded during the keyword search. Such asymmetrical coding practices undoubtedly introduced bias into the data but also provided opportunities for critical assessment of metadata generation methods and consistency standards.

historical context led to some interesting categorizations. Although we recognize an enormous population of Italians that migrated to Argentina, we count Italians as white and Argentinians as nonwhite. Adding to this, we classify Jewish people as nonwhite since they are a religious minority. However, we classify Albanians, who are often Muslim, as white, even though Muslims are also a religious minority. We classified Greeks as white and Turkish people as nonwhite, even though we are aware that the Ottoman Empire and its successor states were religiously, ethnically, and linguistically diverse, and that there was a population swap based on religion rather than ethnicity in 1923. There is no real way to tell whether someone would have continued to identify as Greek or Turkish had they not moved to Utah or which of these countries they would have ultimately been citizens of had they remained. Adding to this, Greek American history also served as a template for how ethnic history is collected and researched in Utah.²⁵

Finally, our definition of diversity in this initial analysis only encompasses racial, ethnic, and religious diversity. We did not include gender, disability, or sexuality as it was deemed outside of the scope of this project or too difficult to capture in an analysis of the metadata alone. There are methods, such as searching for traditional women's names, that can be used to try to identify gender in metadata. However, these methods are even more imperfect and complicated than what we have accomplished with this project. We leave this work, as well as work on disability and sexuality, open to future projects.

Bringing these issues to light can only mean making it easier for people to come up with better metadata solutions. Although our methods proved to be problematic and imperfect, the effect of engaging with this material led us to be more familiar with the variety of issues involved in this type of work. The coding process ensured that we spent hours grappling with our Digital Collections metadata and gave us a deeper appreciation for the time and care needed when engaging with our materials, especially those pertaining to human identity.

DISCUSSION

Despite the limitations and inconsistencies of our data, we were able to extract a few key insights about representation in relation to our Digital Collections. Using Tableau to visualize our data allowed us generate summary, descriptive statistics about our data as well as combine variables to see patterns and trends that were otherwise undetectable in our data tables. Notably, we discovered that less than 11 percent of our collections were coded as nonwhite per the definitions provided above (see fig. 2). Importantly, however, this did not mean that 89 percent of our collections constituted materials related to white people or whiteness; on the contrary, much more of our collections were related to nonhuman-related items, such as geographical features and wildlife. In fact, those Argentinian right whales alone constituted nearly a third of our collection (see fig. 3), attesting to the influence that individual projects and researchers have on digital collections development. Excluding just this one collection of nonhuman items increased our nonwhite ratio by 4 percent, which begs the question of how much more our percentage of nonwhite collections would increase if we were to exclude all items without a direct relation to humans or human activities from future analyses.

Our coding process was labor intensive due to our previous practices of not labeling people in metadata who are white. Keeping in mind emerging best practices from other institutions that utilize the term *white* for descriptive metadata if known, If we were to incorporate similar practices into our descriptive metadata workflows and engage in metadata remediation projects to include the term in the future, it would result in us being able to revisit our corpus and use it for

continual assessment and visualization with greater efficiencies in coding due to more precise descriptive metadata practices that do not assume whiteness without labeling it as such.

Figure 2. Donut chart showing the percentage of the Marriott Library's Digital Collections related to historically underrepresented populations and communities ($n = 302,551$).

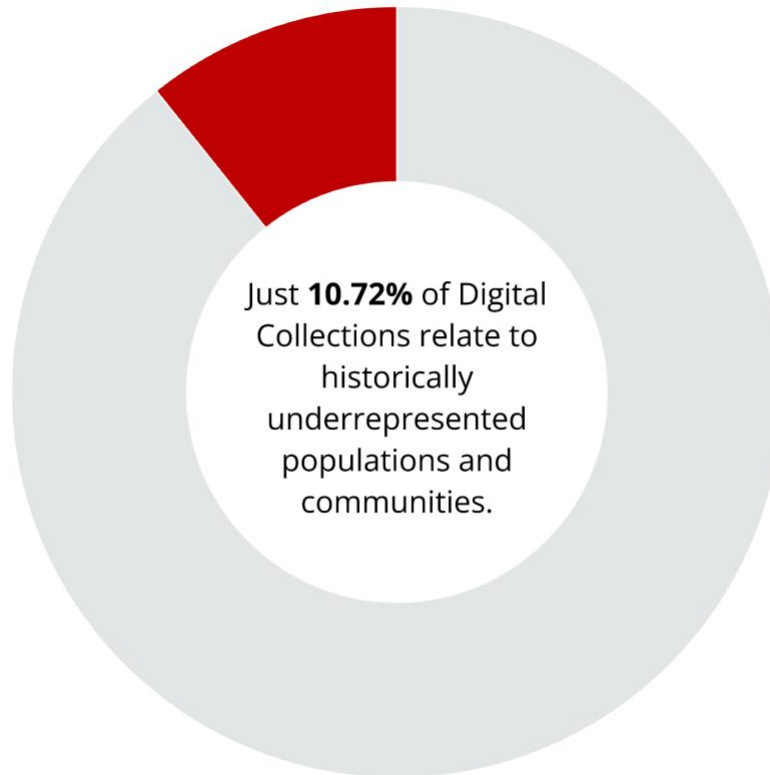
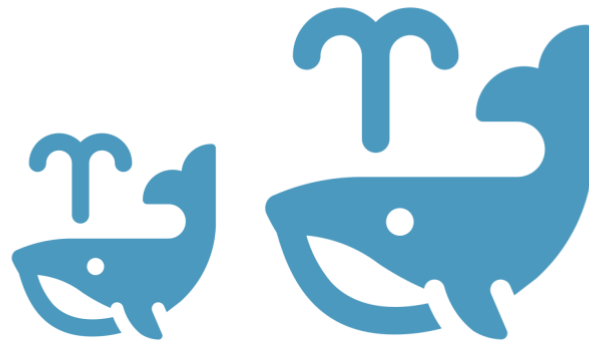


Figure 3. Infographic comparing the percentage of the Marriott Library's Digital Collections related to historically underrepresented populations and communities to the percentage of collections related to right whales ($n = 302,551$).



83,848 (27.71%) of the records in the Marriott's Digital Collections are related to Right Whales.

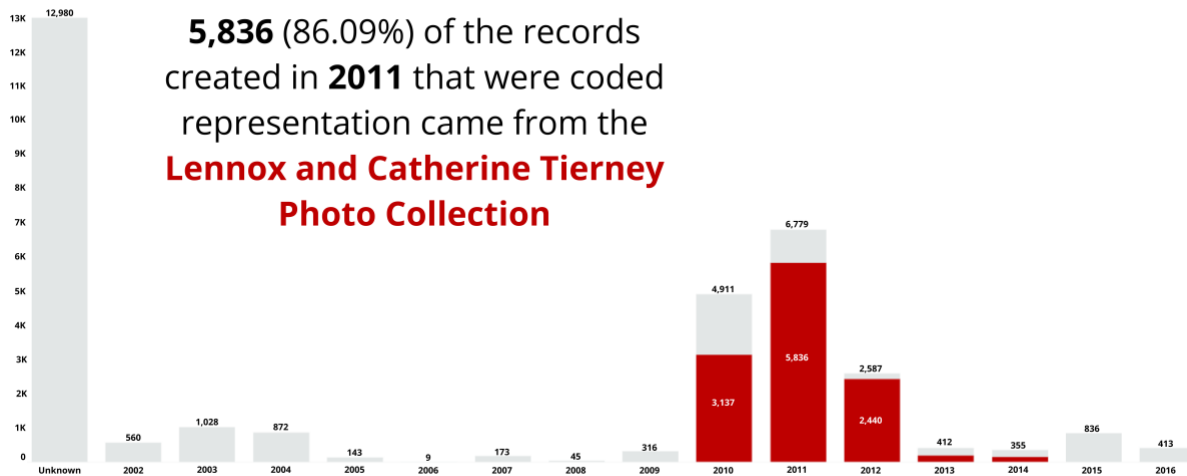
Just **32,423** (10.72%) relate to historically underrepresented populations and communities.

Perhaps even more usefully, we might have begun by examining the collections through the lens of regional representation before turning to questions of racial or ethnic diversity—first asking how well the collections reflect our local environment, rather than material from more geographically distant contexts. After all, collections flagged as diverse through keywords or subjects may not align with the kinds of diversity we're seeking. Especially given that 88 percent of our subset was visual material, we should be wary of collections that do more harm than good, such as those that reflect a white researcher's ethnographic gaze or, worse, offer damaging, performative representations of race. At the time of data collection, for instance, just over 2 percent of our subset was marked as sensitive or restricted content. Sensitive content refers to material that may include images of deceased persons, nudity, or otherwise graphic, violent, or sexual content. Restricted is a much broader category that includes items restricted due to copyright status, but notably also Indigenous American material restricted because of cultural sensitivity or a potential violation of the Native American Graves Protection and Repatriation Act (NAGPRA). Efforts to flag sensitive content and add appropriate warnings and restrictions are ongoing, but it would be worthwhile to look more closely at what is represented in each of these categories. While it remains to be seen whether content flagged in the context of racial and ethnic diversity differs meaningfully from such content elsewhere in the collections, a cursory look into some of the items bearing sensitive content warnings include blackface among other racist and racially violent imagery and some of those marked restricted include Indigenous archaeological material.

This project also prompted reflections on how collections are prioritized for inclusion in the Digital Library. The digitization of the right whales dataset was made possible by a CLIR Digitizing Hidden Collections grant, and represented an important partnership that led to the preservation

of a fifty-year legacy of research conducted by a professor of the university.²⁶ Another focused collection was developed in the mid-2000s, when the library digitized a subset of over eleven thousand slides documenting the art historical research of another professor, a specialist in Japanese art.²⁷ This collection greatly increased our counts of East Asian representation (see fig. 4); however, whether these photographs, taken by a white scholar, constitute representation is again somewhat questionable and likely requires a more nuanced categorization process than we undertook during this initial investigation. In fact, it is likely that much of our nonwhite material reflects a white gaze or is the product of armchair ethnography necessitating a more careful treatment in future iterations of this work.

Figure 4. Bar chart showing the number of new additions to the Marriott Library’s Digital Collections related to historically underrepresented populations and communities annually, 2002–2016, with items related to the Lennox and Catherine Tierney Photograph Collection Photograph Collection in red (n = 32,423).



Additionally, items are often selected for digitization based on the interests of patrons using our Special Collections. The Kennecott Miner Records, for example, were frequently requested, and their digitization was made possible through a reciprocal agreement with FamilySearch that resulted in the library receiving free digitization services for over forty thousand items. The COVID-19 pandemic necessitated remote projects, such as transcription services, to offer to student workers, so this prompted the entire collection to be transcribed at a speed typically not achievable during regular circumstances. Being able to visualize how the composition of our digital collections changed over time with such large-scale additions as the right whale slides, the Lennox and Catherine Tierney Photograph Collection, and the Kennecott Miner Records was useful in imagining future possibilities for targeted collection development. The team was also interested in producing additional visualizations of digital collections over time to better assess how the composition of the Digital Library may shift in five or ten years.

IMPLICATIONS AND FUTURE DIRECTIONS

Even though digital collections often represent a fraction of physical holdings, their reach through unimpeded, worldwide access is inherently powerful. It is imperative that organizations building the physical to digital collection pipeline understand and embrace their role as ethical curators.

Berry underscores this concept, commenting that “digital collections open the library door to everyone with internet access, requiring a reconceptualization of who makes up our patron base. This broadening of access makes it even more imperative to ensure we are creating digital collections that reflect the diverse interests and backgrounds that exist in our holding.”²⁸ The path towards creating digital collections that are more representative of ethnicities and other marginalized people from history and current populations at PWIs with predominantly white collections may seem challenging, but there are actionable steps toward creating more equitable digital collections and metadata. One step towards progress is developing a digital collection development policy with selection criteria that not only gives digitization priority to material that is at risk of loss or damage, of high research value, permissible to publish online, etc., but also prioritizes surfacing underrepresented voices in digital collections. Ziegler’s 2019 article, “Digitization Selection Criteria as Anti-Racist Action,” discusses the process of what is selected for digitization and how this process can be used toward creating more inclusive digital collections at Louisiana State University Libraries. Ziegler implores “those of us working in the field of cultural institution management, those of us who choose what to digitize and thus what narratives to promote, what history to highlight, and what legacies to further have the opportunity to enact anti-racist action through digitization prioritization policies that counteract our racist past.”²⁹

Digitizing more materials by and of marginalized and nonwhite subjects is a part of the strategy towards more equitable digital collections, but describing these collections conscientiously is another vital component. In “The House that Archives Built,” Berry ponders the pitfalls of mass digitization without detailed metadata to guide discovery: “digital collection development has been presented as a liberatory access provider, with the idea that reparative access is primarily a workflow adjustment.”³⁰ While this project focused on identifying any terminology in metadata denoting race, ethnicity, or religion, and not *what* terminology was used, it is important to acknowledge the need for reparative metadata assessment and inclusive descriptive practices as part of the ongoing strategy towards more equitable digital collections. The Inclusive Metadata Toolkit, published by the Digital Library Federation’s Cultural Assessment Working Group (CAWG), provides an array of strategies and resources for advancing inclusive metadata work that “seeks to employ a wider range of perspectives that have been traditionally excluded from and silenced in descriptive practices, in order to more accurately recognize, represent, and respect the breadth of human experiences.”³¹

Through this experiment in collections metadata analysis, it became clear that several critical changes would need to be made to our methodology moving forward. Although the aggregate metadata in their current form remain problematic in how they do or do not directly address racial and ethnic difference, our approach to identifying collections in which nonwhite populations can be improved and expanded to include terms relating to other historically underrepresented identities and populations, namely women and LGBTQIA+ individuals.

One promising plan for revisiting the initial question of how diverse the Digital Library is in its present form is the incorporation of a new metadata assessment tool that two of the authors on this article recently developed for reparative metadata assessment. The Marriott Reparative Metadata Assessment Tool (MaRMAT) is an open-source, Python-based application that uses specialized lexicons to identify potentially harmful, outdated, or otherwise problematic language and content within metadata records. By leveraging these lexicons, MaRMAT empowers metadata practitioners to conduct thorough and thoughtful reviews of their collections, supporting more inclusive and accurate descriptions.³² While this tool was initially developed in conjunction with the Inclusive Metadata Toolkit to facilitate the identification of potentially harmful and outdated

language in library metadata, the tool is highly customizable, allowing users to build and incorporate custom term lists tailored to their specific assessment goals.³³ In building a customized lexicon with categories for various “nondefault” identities (i.e., nonwhite) and “default identities” (i.e., white), it would be possible to efficiently—and with reduced selection bias—query and categorize our digital collections metadata, a task we look forward to undertaking in the near future. Since the beta launch in October 2024, MaRMAT has undergone several improvements that will significantly improve our approach and allow us to look more closely at *how* certain groups are represented in the metadata and the extent to which they demonstrate historical biases rather than present-day paradigms of representation. A key example here would be in looking for instances of “unnamed women” or women only identified as wives or daughters of named men.³⁴

In addition to examining our digitization priorities, the library is actively engaged in developing digital exhibits that spotlight materials from the Digital Library, prioritizing materials for additional digital curation and scholarship. Projects such as “Century of Black Mormons,” “Racial Lynching in Utah,” “‘Resignation Is an Unacceptable Solution’: Black Faculty and Staff at the U during the Black Campus Movement,” and “Utah SOUL: A History of Black Students at the University of Utah” have provided opportunities for university researchers to highlight unexpected materials from our collections, but also contribute meaningfully to local history, telling a more diverse story about Utah.³⁵

CONCLUSION

This research was undertaken in large part to understand the current composition of the Digital Library and assess the extent to which it did or, more likely, did not reflect demographic realities. Ultimately, however, this experiment yielded more questions than results, such as how large-scale collections have the potential to shift the composition of the Digital Library and how to better represent collections that may feature diverse peoples and places but are developed by external observers. People often assume that technology and data-intensive methodologies make research easier or even less biased, though this is rarely a reality. For this project, the labor involved in coding our collections metadata through an iterative process helped illuminate just how important a humanistic approach to digital collections assessment is, even when aided by data cleaning and visualization tools like OpenRefine and Tableau. While we initially hoped this exercise would provide a roadmap for future collections development, what resulted was a self-reflective analysis of how our metadata may be analyzed to ask new questions about diversity and inclusion in the digital sphere. As work progresses on MaRMAT, we will advance with an iterative process of continual metadata assessment that will look at the specific communities that the Digital Library represents and serves, both historically and at the present time. This will provide a more nuanced view into the way our digital collections reflect our institutional and community demographics.

Going into this project, we knew that, given our institution’s status as a PWI in a state with a reputation—regardless of its demographic realities—for being ethnically and religiously homogeneous, we were not going to encounter the most diverse of collections. Still, the process of grappling with our metadata at both the item and collection level allowed us to get a better sense of issues in our digital collections. Treating digital collections metadata as a corpus for coding and visualization allowed us to interact with the Digital Library in a new way. Having a team approach this project from a digital library, public services, and data visualization/digital humanities perspective created an environment that produced generative discussion and a more holistic perspective on digitization and descriptive metadata work. Seeing in stark terms that our Digital

Library composition was around one-tenth nonwhite and one-third slides of Argentinian right whales is also prompting us to be more aware of what is prioritized for routine digitization. Our collections are not representative of Utah's demographics—past or present—though they certainly reflect a history of religious and political whitewashing in Utah that has been inadvertently subsumed into present-day patterns of digitization.

The Digital Library team typically publishes older materials due to copyright limitations. Again, older materials, even those addressing minority populations, tend to reflect the power and privilege of those who were able to obtain literacy, cameras, and other tools of story-making. We are drafting a digital collection development policy that is informed by this experience, but this process will take time. A potential avenue for improvement will be in prioritizing collection development outreach to faculty working on material that is notably absent and yet critical to the history of the university.

Engaging with and analyzing our digital collections metadata allowed us to directly consider the labor, priorities, and evolving practices that have contributed to the development of the Digital Library over the past twenty-five years. Achieving our goal of greater representation in the Digital Library will not be quick or easy to achieve, but by developing tools for easier analysis and articulating our collection development priorities, we hope to be able to assess the composition of the Digital Library periodically in the future to see if we are able to develop a digital library that reflects the community it serves.

ENDNOTES

- ¹ Quoted in Dorothy Berry, "Take Me into the Library and Show Me Myself: Toward Authentic Accessibility in Digital Libraries," *Transactions of the American Philosophical Society* 110, no. 3 (2022): 111, <https://www.jstor.org/stable/45420503>.
- ² On libraries and acculturation, see, for example, Fiona Blackburn, "The Intersection between Cultural Competence and Whiteness in Libraries," *In the Library with the Lead Pipe*, December 1, 2015, <https://www.inthelibrarywiththeleadpipe.org/2015/culturalcompetence>.
- ³ See "Hispanic Oral Histories," J. Willard Marriott Digital Library, accessed August 29, 2024, https://collections.lib.utah.edu/search?facet_setname_s=uum_hoh and "Interviews with African Americans in Utah," J. Willard Marriott Digital Library, accessed August 29, 2024, https://collections.lib.utah.edu/search?facet_setname_s=uum_iaau.
- ⁴ See "Mitsugi M. Kasai Memorial Japanese American Archive," J. Willard Marriott Library, Digital Exhibitions, accessed August 29, 2024, <https://exhibits.lib.utah.edu/s/japanese-american-archive>.
- ⁵ Megan Banta, "Utah Is Getting More Diverse, Except for These Places," *The Salt Lake Tribune*, July 22, 2024, <https://www.sltrib.com/news/2024/07/22/utah-is-getting-more-diverse>.
- ⁶ "Fast Facts," University of Utah Office of Budget and Institutional Analysis, 2023, <https://www.obia.utah.edu/wp-content/uploads/sites/10/2023/01/Fast-Facts-2023-Final.pdf>.
- ⁷ See Equal Opportunity Initiatives, H.B. 261, State of Utah General Session (2024) (enacted), <https://le.utah.gov/~2024/bills/static/HB0261.html>.

- ⁸ See Courtney Tanner, “Tears and Fears: University of Utah Students Say Goodbye to Cultural Centers Closed Under Anti-DEI Law,” *The Salt Lake Tribune*, July 2, 2024, <https://www.sltrib.com/news/education/2024/07/02/what-university-utah-students-had>.
- ⁹ Emma Green, “What Comes After D.E.I.?” *The New Yorker*, April 14, 2025, <https://www.newyorker.com/magazine/2025/04/21/what-comes-after-dei>; Madison Markham and Samantha LaFrance, “The State of Book Bans: Utah’s ‘No-Read List,’” *PEN America*, August 22, 2024, <https://pen.org/the-state-of-book-bans-utahs-no-read-list>.
- ¹⁰ “Archival Silence,” Society of American Archivists, accessed August 29, 2024, <https://dictionary.archivists.org/entry/archival-silence.html>.
- ¹¹ Marlene Manoff, “Human and Machine Entanglement in the Digital Archive: Academic Libraries and Socio-Technical Change,” *portal: Libraries and the Academy* 15, no. 3 (2015): 517, <https://doi.org/10.1353/pla.2015.0033>.
- ¹² Michael Moss and David Thomas, eds., *Archival Silences: Missing, Lost and, Uncreated Archives* (Routledge, 2021), 11.
- ¹³ Eira Tansey, “Institutional Silences and the Digital Dark Age,” *The Schedule: A Blog for the Society of American Archivists’ Records Management Section*, May 23, 2016, <https://saarmrt.wordpress.com/2016/05/23/institutional-silences-and-the-digital-dark-age>.
- ¹⁴ S. L. Ziegler, “Digitization Selection Criteria as Anti-Racist Action,” *Code4Lib Journal* 45 (2019), <https://journal.code4lib.org/articles/14667>.
- ¹⁵ Autumn Kalikin, “Committing to Repair—The Gaps Analysis Part II,” DC History Center, July 2, 2024, <https://dchistory.org/committing-to-repair-the-gaps-analysis-part-ii>.
- ¹⁶ Lauren F. Klein, “The Image of Absence: Archival Silence, Data Visualization, and James Hemings,” *American Literature* 85, no. 4 (December 2013): 683, <https://doi.org/10.1215/00029831-2367310>.
- ¹⁷ Charlotte Lellman et al., “Guidelines for Inclusive and Conscientious Metadata,” in *Center for the History of Medicine: Policies & Procedures Manual*, accessed March 2024, <https://harvardwiki.atlassian.net/wiki/spaces/hmschommanual/pages/49446971/Guidelines+for+Inclusive+and+Conscientious+Description>.
- ¹⁸ See Meg Hixon, “De-Centering Whiteness in the Archives at the 2018 Midwest Archives Conference,” *Society of American Archivists Human Rights Archives Section* (blog), April 5, 2018, <https://hrarchives.wordpress.com/2018/04/05/de-centering-whiteness-in-the-archives-at-the-2018-midwest-archives-conference>.
- ¹⁹ “SCTS-Documentation / Style Guide.md,” UNC-Libraries, GitHub, last updated October 2020, <https://github.com/UNC-Libraries/SCTS-Documentation/blob/main/Style%20Guide.md>.
- ²⁰ On whiteness as an invisible default, see, e.g., Derald Wing Sue, “The Invisible Whiteness of Being,” in *Addressing Racism: Facilitating Cultural Competence in Mental Health and Educational Settings*, ed. Madonna G. Constantine (Wiley, 2006), 15–30.

- ²¹ See OpenRefine, accessed July 8, 2025, <https://openrefine.org>.
- ²² Here, again, we use the term *nonwhiteness* to refer to a variety of invisible nondefault traits within a contemporary US context.
- ²³ “Kennecott Miner Records,” J. Willard Marriott Digital Library, accessed August 29, 2024, https://collections.lib.utah.edu/search?facet_setname_s=uum_kmr.
- ²⁴ The digitized historical slides of Argentinian right whales currently account for 28 percent of the Marriott Library’s digital collections. “Right Whale Collection,” J. Willard Marriott Digital Library, accessed August 29, 2024, https://collections.lib.utah.edu/search?facet_setname_s=uum_rwc.
- ²⁵ See, e.g., Helen Z. Papanikolas, “The Exiled Greeks,” in *The Peoples of Utah*, ed. Helen Z. Papanikolas (Utah State Historical Society, 1976), 409–35, <https://historytogo.utah.gov/exiled-greeks/>.
- ²⁶ Matilyn Mortensen, “Whale of a Project: Library Digitizes 50 Years of Patagonia Research,” @TheU (blog), September 13, 2023, <https://attheu.utah.edu/facultystaff/whale-of-a-project-library-digitizes-50-years-of-patagonia-research>.
- ²⁷ “Lennox and Catherine Tierney Photograph Collection,” J. Willard Marriott Digital Library, accessed August 29, 2024, https://collections.lib.utah.edu/search?facet_setname_s=uum_lctpc.
- ²⁸ Dorothy Berry, “Centering the Margins in Digital Project Planning,” *Journal of Critical Digital Librarianship* 1, no. 1 (2021): 15–22, <https://repository.lsu.edu/jcdl/vol1/iss1/3>.
- ²⁹ Ziegler, “Digitization Selection Criteria.”
- ³⁰ Dorothy Berry, “The House Archives Built,” *up//root*, June 22, 2021, <https://www.uproot.space/features/the-house-archives-built>.
- ³¹ “Inclusive Metadata Toolkit,” Cultural Assessment Working Group (CAWG), Digital Library Federation, October 21, 2024, <https://doi.org/10.17605/OSF.IO/2NMPC>.
- ³² Kaylee P. Alexander, Rachel Wittmann, Aiden deBoer, and Anna Neatrou, Marriott Reparative Metadata Assessment Tool (MaRMAT), last updated June 18, 2025, <https://doi.org/10.17605/OSF.IO/ED6AU>.
- ³³ See “Inclusive Metadata Toolkit.”
- ³⁴ See, for example, Elspeth A. Olson, “Mrs. His Name: Reparative Description as a Tool for Cultural Sensitivity and Discoverability,” *Journal of Western Archives* 14, no. 1 (2023): article 10, <https://digitalcommons.usu.edu/westernarchives/vol14/iss1/10>. Another promising tool for performing this type of analysis is a Python code developed by Noah Geraci that uses Spacy and gender-guesser to identify personal names with the format “Mrs. [male first name] [last name].” See ngeraci, “mrs,” GitHub, accessed November 25, 2024, <https://github.com/ngeraci/mrs>.
- ³⁵ See “Century of Black Mormons,” J. Willard Marriott Library, Digital Exhibitions, accessed November 26, 2024, <https://exhibits.lib.utah.edu/s/century-of-black->

[mormons/page/welcome](#); “Racial Lynching in Utah,” J. Willard Marriott Library, Digital Exhibitions, accessed November 26, 2024, <https://exhibits.lib.utah.edu/s/utah-lynching/page/welcome>; “‘Resignation Is an Unacceptable Solution’: Black Faculty and Staff at the U during the Black Campus Movement,” J. Willard Marriott Library, Digital Exhibitions, accessed November 26, 2024, <https://exhibits.lib.utah.edu/s/BlackFacultyandStaff/page/introduction>; and “Utah SOUL: A History of Black Students at the University of Utah,” J. Willard Marriott Library, Digital Exhibitions, accessed August 29, 2025, <https://exhibits.lib.utah.edu/s/utahsoul/page/introduction>.