Confronting and Addressing Historical Discriminations through KOS: A Case Study of Terminology in the Becker-Eisenmann Collection[†]

Brian Dobreski*, Jian Qin**, Melissa Resnick***

*School of Information Sciences, University of Tennessee, Knoxville, 449 Communications Bldg,
1345 Circle Park Dr., Knoxville, TN 37996 < bdobreski@utk.edu>
**School of Information Studies, Syracuse University, 311 Hinds Hall, Syracuse, NY 13244, <jqin@syr.edu>
***Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, 955 Main St., Buffalo, NY 14203,
<melissa_resnick@hotmail.com>

Brian Dobreski is an Assistant Professor in the School of Information Sciences at University of Tennessee-Knoxville. His research focuses on the social implications of metadata, resource description, and other knowledge organization practices, as well as the concepts of personhood and personal identity. Brian received his Ph.D. in information science from Syracuse University. His dissertation, *Values in Knowledge Organization Standards: A Value Analysis of RDA*, received an honorable mention for the ProQuest Best Dissertation Award, as well as Runner Up for the iSchools 2020 Doctoral Dissertation Award.

Jian Quin is Professor at the School of Information Studies, Syracuse University. Her research areas include metadata and knowledge modeling, knowledge networks in biomedical data, research data management, and scholarly communication. She has published widely in library and information science journals and given presentations at numerous national and international conferences and workshops. Her research has been funded by National Science Foundation and Institute for Museum and Library Services among others. She is the recipient of the 2020 Frederick G. Kilgour Award for Research in Library and Information Technology. Qin holds a PhD degree from University of Illinois at Urbana-Champaign.

Melissa Resnick is currently a Post-Doctoral Fellow at the University at Buffalo, Jacobs School of Medicine, Department of Biomedical Informatics. She holds a Ph.D. from the University of Texas Health Science Center at Houston, an M.L.S. from the City University of New York's Queens College, a M.S. from Rensselaer Polytechnic Institute and a B.A. in Biology from the University at Albany. Her Doctoral work was in classifying healthcare provider's information needs using Medical Subject Headings (MeSH). Her research interests include terminologies and their use in knowledge organization systems.

Abstract: While historical cultural materials inform users of the past, they may also contain language that perpetuates long-entrenched patterns of discrimination. In organizing and providing access to such materials, cultural heritage institutions must negotiate historical language and context with the comprehension and perspectives of modern audiences. Excerpted from a larger project exploring representation and access around historical terminology and personal identity, the present work offers insight into how knowledge organization systems may be used to help modern users confront and make sense of past, discriminatory language in the archive. Using keywords drawn from the titles of 19th and 20th sideshow performer photographs, this work details the construction of a mapping dictionary that brings together corresponding terminology from several vocabulary sources along with annotations designed to explain historical terms to modern audiences. The development of

208

Knowl. Org. 48(2021)No.3
B. Dobreski, J. Quin, M. Resnick. Confronting and Addressing Historical Discriminations through KOS

this dictionary revealed several major types of problematic and potentially discriminatory language including historical euphemisms, misnomers, outdated terms, and sensationalist monikers. The finished dictionary offers opportunities to address these through explanatory annotations and to provide a richer, multi-perspective approach to subject analysis for these and other historical materials.

1.0 Introduction

As part of their mission, cultural heritage institutions collect and provide access to historical materials. While they inform us about cultural history, these materials may also perpetuate long-entrenched patterns of racism, ableism, and other discriminations. Such materials force users to confront offensive and confusing terminologies, and pose challenges to representation, organization, and discovery (Wright 2019). In representing historical, cultural materials, knowledge organization practices and systems must constantly negotiate the terminologies of the past with those of the present. In doing so, they are faced with a dilemma: there is the danger of condoning and propagating discrimination if offensive terminologies are utilized, and there is the danger of censoring or obscuring historical context if they are not. Knowledge organization systems (KOS) have long struggled with discriminatory terminology, but may also provide an effective means for navigating it. Through their work, knowledge organizers have an opportunity to confront and address historical discriminations, helping contemporary user communities face the past ethically and responsibly while looking toward a more equitable future.

In this paper, we offer an example of how KOS can be used as a critical intervention into historical materials, using the Becker-Eisenmann Collection as a case study. This collection of 18th and 19th century photographs depicts American sideshow performers, many of whom were exhibited to the public due to their perceived physical or cultural differences and promoted with terms offensive to modern users (Kochanek 1997). Previous work with this and similar collections has shown how traditional approaches to subject analysis tend to result in reductive representations that struggle to navigate outdated terminology and depictions of disability, culture, and race (Rinn 2018). In the present work, excerpted from a larger study examining identity representation in the context of the Becker-Eisenmann Collection, we detail the creation of a terminological dictionary used to bridge historical perspectives with modern language. Starting with the original language used on the materials themselves, we collected terms and matched them to contemporary KOS while also annotating entries with information about the history and use of the terms. The annotation process revealed several types of problematic terminology usage within the collection. Ultimately, these annotations provide opportunities to intervene in discriminatory language and provide additional context necessary for users to confront and make sense of it. The resulting dictionary holds further potential for providing access for modern users to the Becker-Eisenmann and similar collections.

2.0 Literature review

Archives and other cultural heritage institutions offer modern users an opportunity to interface with the past, though in the process these users must confront outdated perspectives and terminologies. Western archival collections in particular often contain discriminatory language reflecting cultural legacies of oppression and colonialism, language that is inherited into resource descriptions and other knowledge organization practices (Chilcott 2019). While the offensiveness and potential harm of language is subjective and depends to some extent on the context (Jay 2009), there is the potential for discovery systems to institutionalize and heighten the hostility of any encounter with potentially offensive language. When navigating any terminological differences, KOS such as classifications, taxonomies, and thesauri may be used to control linguistic variation and offer wording more familiar to the modern user. These systems are not perfect solutions, however, and modern KOS have been criticized for continuing to propagate discriminatory language and perspectives (Knowlton 2005; Furner 2007). More broadly, KOS and other knowledge organization practices within the cultural heritage domain have been criticized for inadequately addressing race, ethnicity, culture, ability, and other aspects of personal identity (Clarke and Schoonmaker 2019).

Challenges around representing and labeling identity are particularly apparent in collections of historical materials associated with American sideshows. The sideshow was a popular form of entertainment in the nineteenth and early twentieth centuries in which persons deemed "unusual" were displayed to paying spectators. In most cases, the performers were found unusual due to mental or physical differences that would today be considered disabilities (Kochanek 1997). Shifts in twentieth century medicine led to a more scientific understanding of these individuals and

their conditions, which would in turn shift public perception of sideshows from spectacle to a source of exploitation. As a form of entertainment, sideshows have largely fallen out of favor, though the practice has left behind unique, historical collections of photographs and other documentation. The original language used to promote these performers is often antiquated, confusing, and offensive to modern users; providing access to these materials thus poses a challenge to cultural heritage institutions. Historical terminology and even some modern systems of subject representation have been criticized for continuing to discriminate against and marginalize people with disabilities, as well as those belonging to certain racial or ethnic groups (Adler et al. 2017). At the same time cultural heritage institutions may deprive modern users of important context if they obscure historical language (Wright 2019). As with many historical, cultural collections, sideshow materials thus challenge cultural heritage institutions to respectfully and accurately organize, represent, and explain traces of the past.

One such collection is the Becker-Eisenmann Collection, housed in the special collections of Syracuse University. This archival collection contains over 1,400 images of sideshow performers from the nineteenth and twentieth centuries. Many of these images are in the style of a *cartes de visite* or cabinet card, and were originally sold to generate interest and advertise for sideshow acts. The photographs themselves were often composed in a way to exoticize the depicted performer, a stylistic choice further emphasized by the use of evocative, spectacular titles (Bogdan 2012). The end result may have been curious and compelling to contemporaneous audiences, but modern users are more likely to find these materials discriminatory, offensive, and exclusionary (Altermark and Edenborg 2018). Indeed, in an analysis of the Becker-Eisenmann and several similar collections, Rinn (2018) found the representation of these materials to be problematic, particularly in navigating potentially offensive language. Prior work focused on enriching subject access through a faceted approach also noted the difficulties of addressing historical depictions and terminology around disability within this collection (Dobreski et al. 2019). The Becker-Eisenmann Collection thus offers an important set of materials with which to explore the difficulties of bridging historical language to the perspectives of modern users.

3.0 Goals and methodology

Work presented here is part of a larger, ongoing research project focusing the development of an automated, faceted subject indexing approach for the materials in Becker-Eisenmann Collection (see Dobreski et al. 2020). The goal of the present paper is to explore how KOS and annotations can be used to intervene in potentially discriminatory language associated with these materials. This was accomplished through the development of an annotated, terminological dictionary, an interim product that was utilized in subsequent work in the larger project.

The approach of the overall project is text-based; that is, analysis has relied on the textual data associated with the Becker-Eisenmann photographs (i.e., title or caption printed on the photograph) as opposed to visual inspection. For the present study, researchers limited their focus to photographs that depicted persons, excluding those featuring only animals, buildings, or objects. Title metadata was analyzed to make this determination. For those photographs with titles that signified the presence of a person, researchers extracted the phrases associated with each person. These phrases were broken up into terms and arranged into an initial keyword list.

Next, researchers reviewed each term and consulted three KOS as additional sources of terminology: LCSH, SNOMED CT, and Wikipedia. These sources were chosen due to their availability and ability to cover popular and medical perspectives. Researchers identified and recorded the closest match or matches to keywords from our initial list in each of these three sources; in Wikipedia, article titles were taken to represent the preferred term. If no close match was available, no terms were recorded. The keyword list was then enhanced with these equivalent terms from controlled vocabularies, forming a mapping dictionary.

Next, researchers supplemented the dictionary with notes clarifying the meaning, use, and misuse of the historical terms. During this phase, researchers attempted to focus on confusing or potentially offensive terms. However, understanding that this determination is subjective and may be viewed differently by different users, it was decided to implement a more comprehensive approach. Therefore, researchers chose to provide clarifying annotations for all terms in the dictionary.

4.0 Results

Out of 1,411 images in the Becker-Eisenmann Collection, 1,224 had titles that signified at least one distinct person being present. Within these titles, 801 phrases referring to persons were identified. Each of these phrases was further broken down into unique keywords and then de-duplicated, resulting in an initial list of 582 distinct keywords.

During the terminology matching phase, researchers found closest matches for the 582 keywords in each of the three KOS, yielding 387 LCSH terms, 296 SNOMED CT terms, and 502 Wikipedia terms. These were used to enhance the keyword list into a mapping dictionary. Table 1 shows an excerpt from the dictionary, giving the unique keyword found in item titles alongside closest matches from the three KOS.

Item Keyword	Wikipedia	LCSH	SNOMED CT TERM
Giants	Gigantism	Giants	Gigantism (disorder)
Girl	Girl	Girls	Female child (person)
Hairy Face	Hypertrichosis	Hypertrichosis	Hypertrichosis (disorder)
Hairy-Faced	Hypertrichosis	Hypertrichosis	Hypertrichosis (disorder)
Half Woman	Phocomelia	Phocomelus	Phocomelia (disorder)
Heavily Tattooed	Tattoo	Tattooed people	Tattoo of skin (finding)
High Diver	High diving	Divers	[no match]
High Wire Artist	Tightrope walking	Aerialists	[no match]
Human Art Gallery	Tattoo	Tattooed people	Tattoo of skin (finding)

Table 1. Excerpt from mapping dictionary.

Item Keyword	Example Use	Note	Wikipedia	LCSH	SNOMED CT
Baby	Baby Emma - Fat Girl	A euphemism used in stage names for obese performers	Obesity	Obesity	Obesity (disorder)
Bearded	Annie Jones, Bearded Lady	Used in stage names or descriptors for female performers with facial hair	Bearded lady	Hypertrichosis	Hypertrichosis (disorder)
Dog Faced	Jo Jo the Dog Faced Boy	A term used to describe performers with faces covered with hair	Hypertrichosis	Hypertrichosis	Hypertrichosis (disorder)
Midget	Italian midgets	A 19th century term for persons of short stature	Dwarfism	Dwarfs (persons)	Short stature disorder
Ubangi	Big Lipped Ubangi Woman	A misnomer used to promote African woman with lip plates; not a name of a specific culture or tribe	Lip plate	[no match]	[no match]

Table 2. Excerpts from dictionary showing example terms with annotation.

During the course of terminology matching, researchers recorded brief annotations explaining the meaning and usage each of the original item keywords, along with an example usage. This was used to further supplement the dictionary; Table 2 shows selections from the finalized mapping dictionary. Here, researchers have attempted to provide a selection of terms that illustrated a range of issues present in this collection's terminology.

5.0 Discussion

The mapping dictionary for the Becker-Eisenmann Collection reveals a number of terminological issues associated with these materials, leading to language in resource descriptions that may appear discriminatory, offensive, or confusing to users. One such issue is the use of historical euphemisms. These terms may appear innocuous to modern users, but are being used in unfamiliar and frequently marginalizing ways. For instance, in the context of sideshows the term "baby" referred not to a child performer, but to an obese adult performer. Such terms impede understanding for modern users and often harbor discriminatory perspectives. Historical misnomers are another potential source of confusion or offense. The term "Ubangi," for example, was frequently used to promote performers with lip plates, and was presented to sideshow audiences as a tribe name; this is now understood

Knowl. Org. 48(2021)No.3

211

B. Dobreski, J. Quin, M. Resnick. Confronting and Addressing Historical Discriminations through KOS

to be a fictitious usage (Lindfors 1983). Other terms such as "midget" may have been in common usage at the time, but are largely seen as marginalizing by contemporary communities. Finally, insensitive and sensational monikers such as "dog-faced" were used to promote performers whose presentations are now commonly understood as medical conditions. Each of these scenarios illustrates the difficulties that language presents in providing modern users access to historical materials. In such instances, however, we believe the annotated mapping dictionary provides opportunities for knowledge organizers and cultural heritage institutions to meaningfully intervene in confrontations between modern users and historical terminology.

In making sense of these images and the terms associated with them, users may need additional explanation and redirection to other sources for further context and perspective. The data assembled in the mapping dictionary described here can provide this support in two ways. First, subject analysis for these materials can be supplemented by assigning corresponding terms from multiple vocabularies to each image, along with noting the source of each term and linking back to the source vocabulary. This offers the user multiple perspectives on a concept as well as further transparency on where each label is coming from. Users may also wish to consult the vocabulary source as some KOS terminology may also prove unfamiliar or opaque to them. Second, the annotations can be provided alongside subject metadata in order to help users better understand historical usage of terms. Annotations offered in the current mapping dictionary are relatively brief, though these could be enhanced through additional citing literature and cooperation with subject experts. Overall, these strategies can offer support to users in confronting discriminatory terminology in historical archives without removing or obscuring contemporaneous language. These and other strategies for dealing with historical terminology are increasingly important given the continuing growth of digital humanities corpora and the need to better support access and understanding (Grabus et al. 2019).

6.0 Conclusion

Providing access to materials such as the Becker-Eisenmann Collection requires cultural heritage institutions to carefully negotiate historical language with the understandings of modern users. Vocabulary drawn from KOS can be used to intervene in these encounters and help users make sense of outdated, offensive, and discriminatory terminology. Work described in this paper demonstrates the development of an annotated mapping dictionary that can support this process. The dictionary is capable of providing labels from multiple vocabularies and transparency about their sources, as well as explanatory annotations to help users in confronting historical euphemisms, misnomers, and other potential problematic

terminology encountered when using these materials. Ongoing work in the larger study will continue to examine how to implement these features into subject analysis and user interface design. More broadly, this work answers calls for culturally sensitive management of archives as well as acknowledging historical language in a manner that explains and provides context to modern users in a responsible way (Wright 2019).

Further work is also planned to continue to develop the dictionary, including expansion with additional vocabulary sources and the enhancement of the explanatory annotations. The dictionary may also be of use in representing other historical collections of a similar nature, and automation of the matching and mapping process could provide efficient, low-cost means of enhancing subject access for such materials. Most importantly, future work must also focus on soliciting and responding to user feedback on this approach to handling historical materials and language. This feedback is vital in further refining understanding of how KOS can be used to address discriminatory terminology in a manner that educates without condoning, and creates a bridge between modern and past perspectives.

References

Adler, Melissa, Jeffrey T. Huber and A. Tyler Nix. 2017. "Stigmatizing Disability: Library Classifications and the Marking and Marginalization of Books about People with Disabilities." *Library Quarterly* 87: 117–35.

Altermark, Niklas and Emil Edenborg. 2018. "Visualizing the Included Subject: Photography, Progress Narratives and Intellectual Disability." *Subjectivity* 11, no. 4: 287–302.

Bogdan, Robert. 1988. *Freak Show: Presenting Human Oddities for Amusement and Profit*. Chicago, Ill.: University of Chicago Press.

Chilcott, Alicia. 2019. "Towards Protocols for Describing Racially Offensive Language in UK Public Archives. *Archival Science* 19, no. 4: 359–76.

Clarke, Rachel Ivy and Sayward Schoonmaker. 2019. "Metadata for Diversity." *Journal of Documentation* 76, no. 1: 173–96.

Dobreski, Brian, Qin, Jian and Melissa Resnick. 2019. "Side by Side: The Use of Multiple Subject Languages in Capturing Shifting Contexts Around Historical Collections." In *Proceedings from North American Symposium on Knowledge Organization, Vol.7,* 16–26. doi: 10.7152/nasko.v7i1.15615

Dobreski, Brian, Qin, Jian and Melissa Resnick. 2020. "Depicting Historical Persons and Identities: A Faceted Approach." *Knowledge Organization* 47, no. 8: 668–79.

Furner, Jonathan. 2007. "Dewey Deracialized: A Critical Race-Theoretic Perspective." *Knowledge Organization* 34, no. 3: 144–68.

Grabus, Sam, Greenberg, Jane, Logan, Peter and Jane Boone. 2019. "Representing Aboutness: Automatically Indexing 19th Century Encyclopedia Britannica Entries." In *Proceedings from North American Symposium on Knowledge Organization, Vol.7,* 138–48. doi: 10.7152/nasko.v7i1.15635

Jay, Timothy. 2009. "Do Offensive Words Harm People?" *Psychology, Public Policy, and Law* 15, no. 2: 81–101.

Knowlton, Steven A. 2005. "Three Decades since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 40, no. 2: 123–45.

Kochanek, Lisa A. 1997. "Reframing the Freak: From Sideshow to Science." *Victorian Periodicals Review* 30, no. 3: 227–43.

Lindfors, Bernth. 1983. "Circus Africans." *Journal of American Culture* 6, no. 2: 9–14.

Rinn, Meghan R. 2018. "Nineteenth-Century Depictions of Disabilities and Modern Metadata: A Consideration of Material in the PT Barnum Digital Collection." *Journal of Contemporary Archival Studies* 5, article 1.

Wright, Kristen. 2019. "Archival Interventions and the Language We Use." *Archival Science* 19, no. 4: 1–18.