

**Brian Dobreski** — University of Tennessee  
**Jian Qin** — Syracuse University  
**Melissa Resnick** — Independent Consultant

## **Side by Side: The Use of Multiple Subject Languages in Capturing Shifting Contexts around Historical Collections**

### **Abstract**

Archival representations strive to convey the original context around historical collections, but perspectives on certain topics and materials may change over time. Such shifting perspectives pose problems for providing appropriate subject access without obscuring contexts or misrepresenting resources in their broader cultural settings. This study seeks to examine these problems through an investigation of historical materials associated with American sideshows, public opinions on which changed drastically from the 19th to the 20th centuries. Using as a case the Ronald G. Becker Collection of Charles Eisenmann Photographs, an archival collection of digitized images of sideshow performers, we explored a variety of systems of subject representation for major concepts and terms relevant to sideshow performers and activities. Combining keyword extraction and descriptive analysis of current collection metadata (e.g., LCSH) with comparisons to other popular systems of subject representation, we found complex and conflicting perspectives on concepts such as dwarfism, hypertrichosis, and obesity. Results revealed that current LCSH-based subject metadata for these materials tends to reduce historical and demographic context around performers and highlight contemporary medical perspectives. At the same time, original language used in these images is now often seen as demeaning, exploitative, and offensive. Furthermore, neither of these perspectives may match with self-identifying language used in contemporary society. Taken together, however, these sets of terminologies and their relationships may provide a more robust representation of changing perceptions and terminologies over time. Findings from this study reveal the potentials of utilizing a faceted approach and multiple subject languages together to further clarify and contextualize archival collections.

### **Introduction & Background**

Archives organize, maintain, describe, and provide access to collections of historical documents and other materials. The ways in which these artifacts are represented are closely tied to the contexts of their creation and use; such contextual information is fundamental to fostering the discovery and understanding of these objects by potential end users (Society of American Archivists, 2013). A critical task of the archive is then to convey the socio-cultural contexts surrounding archival resources and the persons, places, and groups associated with them. An archival collection's description may strive to capture its original context, but contemporary perspectives on historical materials may change over time. This presents particular challenges to subject analysis and representation for these materials. How can subject access be maintained for concepts that have seen shifts or conflicts in terms of popular terminology? As subject languages update and change over time, do archives run the risk of obscuring certain contexts and misrepresenting resources in their broader cultural settings? How can subject representation address archival materials that have been seen through changing contexts and perspectives over time?

The Ronald G. Becker Collection of Charles Eisenmann Photographs (Becker- Eisenmann Collection) at Syracuse University represents an archival collection with a particularly

complex and shifting context. This collection of photographs by Charles Eisenmann and Frank Wendt contains over 1,400 images of human and animal sideshow performers from the 19th century. Eisenmann photographed these performers as a means of helping them market and promote themselves; their original purpose was thus to entertain and incite curiosity. Many of the images are composed in such a way as to emphasize the abnormal or curious nature of the subjects, which ranged from genuine physical abnormalities to manufactured abnormalities (e.g., tattoos), to spectacular activities (e.g., fire eating). Once intended for popular entertainment, these images now serve an array of purposes, including as reference for artists and photographers, as well as primary sources for researchers in the areas of disability, culture, and performance.

These images have been and continue to be viewed from a number of differing perspectives, many of which have changed over time. While sideshows or freak shows were widely accepted forms of entertainment in America from mid-19th to mid-20th centuries, public opinion soon shifted from spectacle and curiosity to sympathy and embarrassment (Bogdan, 1988). This is in part due to the growing medical and scientific perspective concerning many of the abnormalities displayed in such shows. As Victorian Era medical science “discovered” and explained deformities in scientific terms, curiosities were recast as disabilities and performers as clinical specimens, fundamentally altering the surrounding public discourse (Kochanek, 1997; Blase, 2017). Accompanying such shifting perspectives has been a shifting set of terminologies used to refer to sideshow performers and the conditions that they embody. For instance, original images in the Becker-Eisenmann Collection carry monikers with spectacular language (e.g., “dog-faced boy”), offensive in contemporary culture though representative of the ways in which these performers were promoted at the time. Over time, popular terminology for the conditions these performers embody has grown more reflective of the medical and scientific community’s pathologizing perspectives; Adler, Huber, and Nix (2017) found evidence of this in their historical examination of terminology associated with disability and deformity within several prominent subject languages, including MeSH and LCSH.

At the same time, both popular and medical terminologies may conflict with how contemporary persons with disabilities choose to self-identify (Bogdan, 1988). This creates discrepancies between how these groups label themselves now and how materials about relevant historical figures may be labeled. Beyond this, reducing historical performers to their embodied conditions strips away other important context concerning them and their lives. In a review of subject representation approaches, Rinn (2018) argued that when representing historic individuals with disabilities, a range of diverse terminology should be used to capture the whole of that person’s identity and experience, rather than reducing them to a diagnosis; she found, however, that accomplishing this purely with LCSH proved problematic. Though uncontrolled keywords may offer more terminological flexibility here, Robeiro (1996) found that this approach posed serious problems for information retrieval in archives when compared to controlled vocabularies. Given the various challenges concerning labelling and language when referring to historical sideshows and their performers, how can subject

representation in archival practice negotiate historical and contemporary perspectives to fully and effectively depict such materials?

This study seeks to examine these issues through an exploration of subject language in relation to materials in the Becker-Eisenmann Collection. Specifically, it explores the use of multiple subject languages and approaches to capture and communicate the shifting and conflicting perspectives surrounding these materials and the individuals they depict. Using this photographic collection as a case, this work addresses the following questions: What major concepts are represented in current collection metadata, and what labels are used? How are these concepts represented and labelled in other popular systems of knowledge representation? What concepts and contexts are obscured in current subject representations? How can multiple subject representation systems be leveraged together to capture and reflect the various contexts and perspectives associated with these materials? Findings suggest potential strategies for a more robust, holistic approach to representing archival materials with complex contexts, and hold implications for archives, knowledge organization, and the digital humanities.

## **Methods**

To address the major research questions, researchers devised and implemented a multi-part research protocol. To serve as the main data set for this study, researchers downloaded a copy of the Becker-Eisenman Collection metadata. This data set contained 1,411 records representing unique images held in the collection. These metadata records utilized a modified version of Dublin Core, and though a number of metadata elements were present, only title and subject elements were analyzed as part of this study, as described further below. In addition to the main data set, this study employed multiple terminological data sources. Analysis proceeded in three separate phases; specific data sources and methods for each of these phases are next presented.

In the first phase, descriptive analysis was conducted on the title and subject metadata for the total set of 1,411 records. Title metadata for the images reflects the monikers printed on them at the time of their publication; in the absence of a printed title, archivists devised and assigned a brief title using language from similar objects in the collection (e.g., “Unidentified Bearded Lady”). Keyword analysis and text extraction was performed on this title data to isolate descriptive terms and phrases associated with each of the images (e.g., “Chinese midget,” “Bearded lady”). Current subject metadata for the images largely comprises LCSH terms, and each image in the collection has been tagged with one subject heading each. Keyword analysis was conducted on this metadata in order to determine the total number of unique subject headings assigned and the total occurrence counts for each heading.

In the second phase, a comparative analysis was performed among several systems for knowledge representation for key, selected concepts from the Becker-Eisenmann Collection. Three concepts were chosen due to their prominence in first round analysis, and were represented in current collection metadata with the LCSH “Dwarfism”, “Morbid obesity”, and “Hypertrichosis.” Records using the LCSH “Hair” and “Abnormalities, Human hair” were included in the analysis for “Hypertrichosis.” A fourth concept, “Knife throwing,” was

chosen in order to include an activity-focused topic as well. For each of these four concepts, researchers used the results of the first phase analysis to record distinctive title keywords and phrases commonly appearing in records with these LCSH. These title keywords represent 19th century language associated with the four concepts. Next, researchers manually consulted other terminological sources in order to examine how these concepts were represented or labelled elsewhere. These terminological sources included SNOMED CT, MeSH, Wikipedia, and Reddit. The controlled vocabularies SNOMED CT and the 2018 version of MeSH were consulted in order to capture additional medical and scientific perspectives; the UMLS was used to aid in searching these systems. Best terminological matches within these systems were recorded for each of the four concepts where possible. The online communities Wikipedia and Reddit were then searched in order to determine contemporary terminologies and patterns of usage. For Wikipedia, the names of articles referring to these concepts were recorded, while for Reddit the names of sub-Reddits representing any communities self-identifying around these concepts were recorded.

In the third phase, a more in-depth keyword analysis was conducted on the title metadata for records tagged with the most common LCSH in this collection, "Dwarfism." Previously identified title keywords and phrases for all 320 records from this subset were manually reviewed. Researchers attempted to identify recurring facets referred to by these keywords, including age, gender, country/ethnicity, and family relationship; keywords and phrases were grouped into one or more of these facets where possible. Family relationships were represented by keywords such as "wife," "sister/sisters," and "mother." Keywords describing gender included "boy," "girl," "woman/lady," and "man/men." Since exact ages were not given, keywords such as "woman/lady," "man/men," and "mother" were classified as adult, while those such as "child," "boy," and "girl" were classified as child. Keywords representing country/ethnicity encompassed such terms as "Borneo," "Chinese," and "German." Researchers then determined the prevalence of each of these four facets of context among the records in the "Dwarfism" set.

## **Results**

Below, results are presented for the three phases of analysis: analysis of current title and subject metadata, comparative analysis of key concepts in other subject systems, and determining additional context missing from current subject representation.

### *Current Title and Subject Data*

The first phase of analysis examined terminology used in existing metadata for the collection, focusing on the title and subject elements. Keyword extraction of the title metadata resulted in 536 distinct terms and phrases appearing across the collection of 1,411 records; the most commonly appearing were "midget" and "midgets," occurring in 136 and 61 records respectively, as well as "Fat Lady," which was in the title of 57 records. Table 1 presents the ten most commonly occurring title keywords.

**Table 1.** Ten most commonly occurring title keywords.

Title Keyword	Record Count
midget	136
midgets	61
Fat Lady	57
giant	37
Tattooed Woman	25
Bearded Lady	25
Fat Boy	23
Albino Girl	23
midget woman	22
Siamese Twins	20

Current subject analysis for the collection consists of one subject heading per record. Descriptive, quantitative analysis of this subject metadata revealed 52 unique subject headings used across the 1,411 records. Of these 52 headings, “Dwarfism” was the most prevalent and was associated with 320 records, followed by “Morbid obesity” at 164 records. Table 2 presents the ten most commonly occurring subject headings.

**Table 2.** Ten most commonly occurring subject headings.

Subject Heading	Record Count
Dwarfism	320
Morbid obesity	164
Circus performers	140
Abnormalities, Human	122
Giants	79
Tattooed people	79
Albinos and albinism	66
Hair	55
Conjoined twins	55
Circassians -- In popular culture	52

Of the 52 subject headings occurring in the collection, all but 2 of them were found to be compatible with LCSH, the only exceptions being “Ectrodactyly” and “Abnormalities, Human hair.” Overall, subject headings tended to represent the physical conditions embodied by the depicted performers, though activities (“Fire eating”) and objects (“Circus wagons”) were represented as well. Though title keywords conveyed greater conceptual diversity, the

most frequently occurring keywords also reflected the physical conditions and statures of the individuals being depicted.

*Comparative Analysis with Other Systems*

In the second phase of analysis, title keywords and current subject headings from the collection’s metadata were compared to terminology in other popular systems for subject representation. The comparison presented here is limited to four major concepts that were determined during first phase analysis, labeled here by their LCSH terminology: Dwarfism, Obesity, Hypertrichosis, and Knife throwing. Table 3 shows the subject and title terms used in current collection metadata, along with matches determined from consultation of MeSH, SNOMED CT, Wikipedia, and Reddit.

**Table 3.** Closest matches for major concepts in other subject systems.

Subject Terms (LCSH)	Title Keywords	MeSH	SNOMED CT	Wikipedia	Reddit
Dwarfism	midget, dwarf	Dwarfism	Short stature disorder	Dwarfism	Dwarf, Little people
Morbid obesity	fat	Obesity, morbid	Obesity	Obesity	Morbidly obese, Fat
Hypertrichosis, Hair	hair(y), dog faced	Hypertrichosis	Disorder of hair growth	Hypertrichosis, Hirsutism	Hirsutism
Knife throwing	knife throwing	N/A	N/A	Knife throwing	Knife throwing

Each of the four concepts displays different patterns of usage across the selected terminological sources. Records in the collection that were assigned the “Dwarfism” subject heading utilized the terms “midget” or “dwarf” in their title metadata. Though “dwarf” and “dwarfism” were common in other systems of representation, some contemporary communities in Reddit preferred the “little people” label instead. The 19th century term “midget” was not a preferred label in any systems consulted. In contrast, records with the “Morbid obesity” subject heading typically used the term “fat” in their titles, a term which some current Reddit communities used as well. Otherwise, systems referred to this concept as “obesity,” with or without the “morbid” qualifier.

A set of records using terms such as “hairy” or “bearded” in their title metadata were all assigned the “Hypertrichosis” subject heading. In truth, the excessive or unusual hair growth of the individuals depicted is not the result of a singular condition. This is recognized in SNOMED CT, which groups a number of related disorders under the term “Disorder of hair growth.” Other systems distinguished between two prominent conditions; for example, Wikipedia has separate pages for “Hypertrichosis” and “Hirsutism.” Finally, the “Knife

throwing” concept was not present in the medically focused systems of MeSH or SNOMED CT, and showed no variation in terminology in the other systems consulted.

Overall, these four examples demonstrate different patterns of usage across time and communities for major concepts associated with the Becker-Eisenmann Collection. Concepts referring to embodied conditions saw more variations in usage. In contrast, the activity focused “Knife throwing” concept showed no variation in terminological usage.

*Dwarfism: Additional Contextual Facets*

The third part of our analysis compared title metadata to current subject metadata for materials in the collection in order to understand additional contextual information that may not be represented in existing subject analysis. For this examination, we chose as a subset of materials all records that had been assigned the subject of “Dwarfism.” For these 320 records, we reviewed the previously extracted keywords from the title field metadata. These extracted keywords were then coded according to types of context represented. These contexts, example key words, and occurrence data are presented in Table 4.

**Table 4.** Facets of additional context present in title keyword data.

Additional Contexts	Example Key Words	Number of Records	Percentage (n=320)
Gender	Lady, Sisters, Baron	242	75.63
Age	Child, Men, Mother	97	30.31
Country/Ethnicity	Borneo, Chinese	26	8.13
Family Relationships	Family, Mr. & Mrs., Twins	59	18.44

Coding of title keywords for this set of materials revealed four major facets of additional context that was present but not represented by the current subject analysis (i.e., the “Dwarfism” heading). Occurring most frequently were terms signifying the gender of the depicted person, which were present in over 75% of the titles. Also present were terms that signified age, ethnicity, and family relationships of the persons depicted. Together, these four dimensions represent additional facets of context that are readily available could be used to enhance current subject analysis for a sizable percentage of records. Due to the scope of the collection itself, we were able to supply two more facets of context as well: time period and occupation. Though neither of these facets was represented in the previously analyzed title metadata, both are implicit in the materials themselves: the collection is composed of 19th and 20th century images of sideshow performers. All images can thus be tagged with time period (19th or 20th century) and occupation (performer, entertainer) facets.

Despite their common subject heading of “Dwarfism”, the records analyzed in this phase reflected a diversity of characteristics that represent important context in understanding the individuals being depicted. Analysis of title metadata and collection characteristics yielded a total of six facets of context missing from current subject representation: gender, age, ethnicity, family relationship, time period, occupation.

## **Discussion**

Analysis of current title and subject metadata for the Becker-Eisenmann Collection materials shows the emphasis this collection places on variations of the human body. These images were created to generate interest for sideshows and their performers, and did so largely by depicting spectacular aspects of the subjects' height, weight, hair, and other aspects of their appearance. Title metadata employs spectacular, 19th century language, much of which may now be seen as demeaning, exploitative, or offensive. In translating these concepts into a contemporary subject language, the current LCSH- based subject analysis of these materials tends to reposition the spectacular as the clinical and highlight a modern medical perspective. Relabeling hairy or "dog-faced" performers as "Hypertrichosis" or "Hirsutism" avoids the perpetuation of derogatory language, but is at the same time akin to an act of retrospective diagnosis that reflects non-contemporaneous knowledge and reduces historical context.

In equating people with the conditions they embody, this approach to subject representation also reduces demographic context that aids in more fully understanding these materials. Our review of all 320 records that shared the single subject heading of "Dwarfism" showed important demographic diversity among these individuals that lay obscured. Specifically, we found six facets of additional context that could be used to enhance current subject representation: age, gender, ethnicity, relations, occupation, and time period. Rinn (2018) similarly argued for a more holistic approach to representing historic individuals with disabilities, but found that this was difficult to accomplish solely using LCSH. While Ribeiro (1996) advocated for controlled vocabularies in archival settings, she postulated that some combinations of controlled and uncontrolled terminology may be effective. Figure 1 shows a hypothetical reimagining of subject metadata for one item from the Becker-Eisenmann collection, using a combination of the various controlled and uncontrolled terminologies examined in this study to place multiple perspectives on this image side by side.

**Figure 1.** Enhanced subject representation for The Wild Men of Borneo, Midgets.

	<b>Individual depicted (LCNAF)</b>	Wano, 1825-1905
	<b>Individual depicted (LCNAF)</b>	Plutano, 1827-1912
	<b>Groups depicted (LCSH)</b>	Circus performers
	<b>Groups depicted (LCSH)</b>	Men
	<b>Groups depicted (21<sup>st</sup> century terminology)</b>	Little people
	<b>Groups depicted (21<sup>st</sup> century terminology)</b>	Dwarfs
	<b>Groups depicted (19<sup>th</sup> century terminology)</b>	Midgets
	<b>Conditions depicted (MeSH)</b>	Dwarfism
	<b>Conditions depicted (SNOMED CT)</b>	Short stature disorder
	<b>Relations depicted (LCSH)</b>	Brothers
	<b>Nationalities depicted (LCSH)</b>	Americans
	<b>Places referenced (19<sup>th</sup> century terminology)</b>	Borneo
	<b>Time period depicted (LCSH)</b>	Nineteenth century

This enhanced subject representation leverages historical language derived from the original materials, as well as contemporary, colloquial terminology. Here, modern terms gleaned from group self-identification in online communities such as Wikipedia and Reddit during our second phase analysis have been used to supply additional perspectives on the persons depicted. While the combination of these facets and subject languages provides a more holistic representation, it is important to note that it also depicts disagreements on labelling that may confuse or challenge users. In this example, the terminology sources have been indicated to provide further context. In such a discovery system, additional information could be provided to users on the nature and origin of these subject languages to aid in their understanding of why certain labeling choices were made. The subject metadata framework for depicting visual representations of persons presented here may also aid in automatic indexing of images. Metadata creation tools could incorporate these facets into templates utilizing pre-determined sets of subject terms to increase the consistency, completeness, and efficiency of subject representation. Overall, the example presented here opens a number of new discovery and access potentials, and represents just one possibility for how variations in perspective could be conveyed through multiple subject languages.

As demonstrated in our analysis, language used to identify sideshow performers and the conditions they embody varies not only across communities but across time as well. Terms used to refer to persons with dwarfism, for instance, have shifted over the years and are likely to continue to shift. Capturing and conveying terminological shifts among “midget,” “dwarf,” “little person,” and emerging new terms are challenging, and current models of subject

representation lack the expressiveness necessary to convey such complex terminological interrelationships. A more feasible solution may entail relating terminology within and among different knowledge organization systems themselves through the use of SKOS and other tools, so that these terminological relationships would not need to be reconstructed within the metadata for each individual resource. Such an approach could facilitate users' understanding and their leveraging of different perspectives in discovery and use of historical materials.

### **Conclusion**

Archival representations are designed to help users understand the context surrounding historical resources. Capturing and conveying the varying and conflicting perspectives on these materials and the topics they represent can provide a broader sense of context for modern users, but also poses terminological challenges. As shown in our case study of subject representation in the Becker-Eisenmann Collection, labelling historical depictions of physical abnormalities and persons with disabilities is particularly problematic. Though the collection's LCSH-based subject metadata tends to view the depicted individuals through a solely medical perspective, other terminological sources, both controlled and uncontrolled, can be used to highlight self-identifying and popular perspectives. Beyond varying perspectives on the physical conditions embodied in these materials, our analysis also revealed relevant contextual information omitted from the collection's current subject analysis. A more holistic approach to representing archival materials with multi-faceted contexts could open up new discovery and understanding of such collections of resources, facilitating new approaches in research in the digital humanities and other areas.

Taken together, multiple subject languages may provide a more robust representation of changing perceptions and language over time, though care must be given to how terminologies are labeled and presented to end users. Further work is needed to better understand how to design complex subject metadata and effectively utilize it in discovery and retrieval in archival settings. Ultimately, work done to relate terms at the vocabulary level, rather than the individual resource level, may offer more sustainable approaches to leveraging multiple subject languages in the representation and discovery of historical materials.

### **References**

- Adler, Melissa, Huber, Jeffrey T., and A. Tyler Nix. 2017. Stigmatizing Disability: Library Classifications and the Marking and Marginalization of Books about People with Disabilities. *Library Quarterly* 87: 117–135.
- Blase, Rachel. 2017. *The Changing Societal View of Freaks: Popular Culture, Medical Discourse, and Physical Differences in 19th and 20th Century* (Doctoral dissertation).
- Bogdan, Robert. 1988. *Freak Show: Presenting Human Oddities for Amusement and Profit*. Chicago, Ill.: University of Chicago Press.
- Kochanek, Lisa A. 1997. Reframing the Freak: From Sideshow to Science. *Victorian Periodicals Review Fall* 227–243.

Brian Dobreski, Jian Qin, & Melissa Resnick. 2019. Side by Side: The Use of Multiple Subject Languages in Capturing Shifting Contexts Around Historical Collections. *NASKO*, Vol. 7. pp. 16-26.

Ribeiro, Fernanda. 1996. "Subject Indexing and Authority Control in Archives: The Need for Subject Indexing in Archives and for an Indexing Policy Using Controlled Language." *Journal of the Society of Archivists* 17: 27–54.

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: 1.

Ronald G. Becker Collection of Charles Eisenmann photographs. Special Collections Research Center, Syracuse University Libraries, Syracuse, N.Y.

Society of American Archivists. 2013. *Describing Archives : A Content Standard (2nd ed.)*. Chicago, Ill.: Society of American Archivists.